Multivariate image processing: methods and applications

Ch. COLLET - J. CHANUSSOT - K. CHEHDI

version 2.0 May 26th 2009

Table des matières

E. APTOULA, S. LEFÈVRE

Chapitre 13. Parallelizing Image Analysis Applications for Spectral Microscopy   381
E. GABRIEL, S. SHAH

Chapitre 14. Hypercomplex models and processing for vector images . . .   413
Stephen J. SANGWINE, Todd A. ELL, Nicolas LE BIHAN

Chapitre 15. Panoramic integral-field spectrograph : huge ultraspectral data to understand Universe history   443
Ch. COLLET

Chapitre 1

Multivariate image processing : methods and applications

New generations of sensors coupled with powerful capacity of computer always increasing, allow to answer effectively to the needs and new requirements of various applicative domains. Indeed, multivariate imaging appears nowadays to be a great medium to propagate complex, massive and subtil information in order to feed data analysis systems. The generic name "multivariate imaging" recovers very diverse information media (color imaging, multimodal data, multi-spectral and hyper-spectral data, multi-dates images, heterogeneous, massive data or multi-sources observations). Various methodological objectives are challenging, requiring pre-treatments (reduction of dimensionnality, research of adapted representation (i.e., parcimonious spaces), registration etc., up to advanced modellings allowing accurate analysis and fusion process, the segmentation or the classification in various contexts such as microscopy imaging of vegetable biology or remote sensing in planetology. The relevance and the exploitation of the rich, subtil and complex information of the multivariate imaging system thus encounter more and more interest in a lot of domains, as in astronomy, remote sensing, medicine, agronomy, chemistry, pharmacology, etc. The exploitation of such wealth, requires the analysis of an important (sometime huge) quantity of information brought by this type of data and thus the design of new methods and adapted tools in relation with the physics and the mathematics knowledge. The availability, the abundance and the specific structure of such multicomponent information, face with new challenges that the scientific community try now to overcome. The contents of this book arose from the research action "multivariate imaging" led within the framework of the Research Group Information, Signal, Images and Vision (GDR ISIS)

supported by the CNRS (http : // gdr-isis.org) clustering research center, academics and industrial skills. This action which we animated from 2005 till 2008, aimed at the analysis and the manipulation of the multivariate imaging, driven from theoritical or applicative motivations. The multivariate imaging exhibits several interests, among them :

– the wealth and the complexity of the data (multispectral microscopy, color images, heterogeneous data (e.g., photo and geographic maps, RADAR and optics images, atlas and multimodal data in MRI, etc.), chemical spectroscopy, active and passive multispectral sensors, satellite remote detection in planetology (Earth, Mars) ;

– the diversity of the goals concerning all the image processing schemes (reduction of dimension, registration, filtering, classification, segmentation, spectral analysis, fusion, etc.) ;

– the variety of the methods (linear or non-linear, mathematics morphology, statistical approaches, supervised versus non-supervised methods, etc.) the set of the developed approaches being extremely broad.

The data can be of variable nature : multiband, multimodal, multi-view, heterogeneous.

– The multiband data : they go from the simplest case of the color image (coded for example in RGB) toward the more complex cases of multispectral images (less than ten spectral bands), hyperspectral (less than hundred spectral bands) or superspectral (more than thousand spectral bands (where each pixel - also named spexel for spectral elements has a high resolution in both spaces (spatial and spectral)).

– The multimodal data : here, the notion of modality recovers the acquisition of data observed from the same scene but by means of various protocols of acquisition. For example in MRI, we distinguish the modalities T1, T2 or the Flair which bring additional information. So in imaging MRI of the brain, in the case of the detection of multiple sclerosis, the interest of a multimodal observation will consist in detecting lesions according to Flair acquisition (where lesions appear brighter Ð e.g., with high luminance) and T2 (where the lesion is bright as the cerebrospinal liquid - LCR) and to use the third modality (T1 who allows to differentiate well LCR, grey and white matter) to confirm the detection.

– The multiview data : the same scene is observed according to different positions with the same sensor or various sensors using the same modality. This process, useful to realize the reconstruction in 3D of the scene for example, can become complex when the multisview data are realized according to various modalities : a registration is then required to allow to exploit the wealth of the information carried by each of the modalities. The multidates data : they correspond for example to data supplied by the same airborne sensor flying over a zone at different periods : the problem of detection of change and its quantification can then be a very interesting problem.

– The heterogeneous data : we can have access to matrices of heterogeneous data (e.g., matrix of Mueller in polarimetric imaging) for which the understanding of the physical models associated with these parameters is indispensable : such multivariate images require a bidisciplinaire collaboration between physicists and signal processing experts.

These various examples illustrate well the extreme variety of the data which recovers the naming " multivariate imaging ", which are also named multicomponent imaging. The recent development of sensors delivering a large amount of images, the costs always cheaper of such sensors generate considerable volume of data for which new methods of analysis must be developed. Let us quote for example the domain of the astronomy where the notion of virtual observatory took place at the beginning of the two thousand years, with the capacity offers henceforth by the internet, to connect, thanks to new interoperability tools, data of observations of the Universe taken in various periods and by various telescopes, according to various protocols (varying time of exposure, with adaptative optics or not, etc.). An astronomer has now the possibility, with a few number of mouse clicks, to manipulate from his office equipped with a computer and a fast internet connexion, the data of a virtual observatory clustering all the observations stored on several decades, so as long as they are available in a digital format. He will then try to obtain denoised images, and possibly super-resolved images by appealing to the specialists of signal processing. Within the framework of the virtual Observatory, we shall for example wish to merge heterogeneous multispectral information of galaxies. In planetary imaging, to combine several satellite images, it will be necessary to take into account the topography of the ground and the parameters of observation. The domains of applications covered by the title "multivariate imaging" are thus very vast, rich and challenging.

The main topics presented in this book were grouped together in four parts :

1) the registration and the fusion of images, where the developed methodologies deal with satellite imaging (chap. 2 - 4) and concern the problem of fusion of optical and radar heterogeneous data (interferomety / radargrammetry) for the 3D reconstruction of urban zones (chap. 3) ; the fusion of data of various resolutions (multispectral and panchromatique) to increase the quality of the multispectral images with low spatial resolution (chap. 4) ; the registration (required step to the fusion of images) is also applied to airborne data acquired by spectrometer imager (chap. 2) ;

2) the detection of changes in remote sensing applications with a synthetic display by satellite imaging for the observation of the Earth in the context of the temporal analysis of changes in multi-or hyperspectral images of type is described without the chapter 5. This part also presents the bayesien framework for the analysis mixed (spatial/spectral) of multispectral cubes for the chemical identification of compounds in spectroscopy by analysis in independent components and separation of non-negative sources with applications concerning the data analysis of the probe Express Mars and on the analysis of spectroscopic signals. (Chap. 6) present the detection of emission

lines allowing to find the kinematic properties of clouds of dust in the heart of the galaxies in radioastronomy (chap. 7) ;

3) denoising and segmentation technics are approached on this part. The chap. 8 presents various methods of restoration of multichannels noised by a correlated signal, using a decomposition on wavelet basis whereas in the chap. 10 are studied the blind characterization of the types of noise in remote detection. The segmentation of multivariate images by clustering approaches requiring the search of an adapted metrics and the space dimension reduction is presented to the chapter 9. This part ends in a presentation of the tools of mathematical morphology spread to the multivariate case and applied to the classification of textures on color images (chap. 11) ;

4) the analysis of massive multicomponent images is the object of the last chapter and illustrates some of the new challenges : in astronomy with the arrival of ultraspectral instruments of observation of the Universe generating massive data for which the fusion of the various views requires the complete modelling of the instrument to allow an optimal fusion (optimal Under bayesian criterion), chap. 12 ; in multispectral molecular imaging where the parallel algorithms and the development of new architectures of treatment of large-sized images becomes a key element of the analysis (chap. 13) ; or still the modelling by hypercomplex numbers which generalizes the notion of complex variable and which can find applications in the manipulation of multivariate data in a mathematically elegant way (chap. 14).

The emergent challenges brought by multivariate imaging, the wealth of the exchanges and the large dynamism of the community in this domain, the mutual enrichment emerging from the discussions between the various scientists, strongly motivated us to realize this book. It addresses the researchers, the PhD students, the students of Masters and the engineers interested in the new methodologies of analysis for complex and heterogeneous multidimensional data, for which the physics plays even more an essential role in the understanding and the comprehension of multivariate data.

<div align="center">
May 2009
Christophe Collet and Jocelyn Chanussot and Kacem Chehdi
</div>

Registration and Fusion

Chapitre 2

Registration of multimodal/multitemporal images

2.1. Introduction

Multivariate, multimodal and multitemporal image registration has become a very active research field during the past two decades. Image registration now encompasses a broad range of applications including remote sensing, medical imaging, computer vision and astronomy. Many concepts, data and algorithms are available in the literature and on the Internet, which can be used to align two or more images acquired by different sensors and fuse them into a single, enriched one.

Although the tasks of image registration for multimodal and multitemporal images are performed quite easily in some specific cases, they can sometimes become very complex, depending on either the sensors used (i.e. the range of pixel values and distributions, the internal parameters), the imaging conditions (i.e. the position and look direction of the imaging sensor), and the imaged subject (i.e. flat area, 3-D scene, multi-temporal acquisition). Therefore, a sound knowledge of the imaging system and the imaging conditions is often required to achieve good registration performances.

In 1998, Maintz and Viergever have suggested a classification of the registration methods covering the range of medical imaging [MAI 98]. This classification accounts for the following items :

––––––––––––––

1) The dimensionality of the data, including 2-D, 3-D data and times series (4-D, i.e. 3-D plus time);

2) the nature of the registration basis, i.e. extrinsic (requiring external markers) or intrinsic (based on the image content only, relying on either salient landmarks, region-based structures or gray level values);

3) the nature (i.e. rigid, affine, projective, elastic) of the transformation between images;

4) the domain (i.e. local or global) of the transformation;

5) the interaction with the user (i.e. automatic, semi-automatic, or interactive);

6) the type of optimization procedure (i.e. the parameters are either computed directly from the data or searched for);

7) the modalities involved (i.e. monomodal or multimodal images);

8) the subject (i.e. intrasubject-based, intersubject-based or atlas-based registration);

9) the object (i.e. head, thorax, abdomen, etc.)

While the last two items of this classification are specific to medical image registration, the first seven items can apply to several fields such as remote sensing, or computer vision to cite a few. It is worth noting that nearly three hundred reference papers were pointed out in their survey at that time, and that more than 2200 scientific papers related to image registration were published during the last ten years.

Besides the international widespread interest – either acamedic and industrial – for more and more efficient approaches to image registration, the influence of the French image and signal processing laboratories in this domain is patent and well recognized. Indeed, several French research groups have brought to the international community their experience in formalizing many complex registration problems, especially in medical imaging and computer vision (see for instance [THI 98], [**?**], [HER 02], [**?**]).

Several reference books on image registration are available to the reader interested in general and applied image registration problems [HAJ 01] [GOS 05]. These books cover a wide range of many different aspects of image registration, from the preprocessing of image data to the fusion and mosaicking of registered images, through the alignment criteria and transformation functions of selected features such as points, lines, regions or gray-level templates.

In this chapter, we do not claim to present a complete state of the art of image registration in general. Excellent surveys are available to the reader, such as [BRO 92], [MAI 98] and [ZIT 03]. Nevertheless, we will recall the fundamentals of image registration, rapidly focusing our discussion on pixel-based image registration, which is recognized as one major advance for several application-oriented tasks such as data

fusion or image mosaicking. The main tools which are used to process images acquired either by different imaging sensors (*multimodal registration*) or at different dates (*multitemporal registration*) will be described. We will first give an overview of the different approaches used to perform automatic image registration. Next, we will consider the main processing steps of image registration and how these steps should be combined in pixel-based registration schemes. Throughout the chapter, we attempt to illustrate some aspects of multimodal/multitemporal image registration with examples taken from the remote sensing application field.

2.2. Overview of image registration approaches

2.2.1. *Definitions*

Image registration is the process of aligning two or more images of the same scene taken at different times, from different viewpoints, and/or by different sensors. Typical applications of image registration lie in the fields of :

– medical imaging : monitoring tumor growth, aligning several imaging modalities like magnetic resonance imagery (MRI), computed tomography (CT), ultrasound imaging, etc.

– computer vision : automatic quality control, robotics, computer-assisted surgery, non destructive testing, etc.

– remote sensing : airborne/spaceborne Earth Observation (EO), change detection, integration and merging of information in geographic information systems (GIS), etc.

Basically, image registration aims at establishing a correspondence between two images on a point-to-point basis. Generally one image is referred to as the *target* (or *reference*) image I_t, and the other one is referred to as the *source* (or *sensed*) image I_s. Therefore, in the following, we will consider the registration problem as trying to register the source image onto the target image as best as possible. The registration is then the search for an optimal mapping $T^\star$ which relates the source image to the target image. Such a mapping is sometimes assumed to follow a *geometric transformation model* T_Θ governed by a parameter vector Θ. The registration problem is then reduced to the search of the optimal parameter $\Theta^\star$.

Let $I : \Omega_I \subset \mathbb{R}^n \to \mathbb{R}; M \mapsto I(M)$ be a n-dimensional image. From a very general viewpoint, the problem of aligning a source image I_s with a target image I_t, can be stated as follows [BRO 92, MAL 06] :

$$T^\star = \arg\max_{T \in \mathcal{T}} S(F[I_t(\mathbf{t})], F[I_s \circ T(\mathbf{t})]) \quad , \tag{2.1}$$

where $F : \mathbb{R} \to A \subset \mathbb{N}$ is a function aiming to extract spatial features on the image supports ; $T : \mathbb{R}^n \to \mathbb{R}^n$ is the mapping function to be found, acting on the coordinates

t of the target image, and $S : A \times A \rightarrow \mathbb{R}$ is a similarity measure between sets of spatial features.

As a consequence, three sub-problems arise from this definition, and have different impact respectively for the user and for the designer of registration techniques.

The first sub-problem relies on the nature of the mapping and its modelling. Actually, the choice of a particular mapping model, be it parameterized or not, is strongly dependent on the nature of the image acquisition process, and, if not application dependent and thus strictly imposed, is likely to be inferred by the user from his own experience. Thus a careful choice of the mapping function is expected from the end-user point of view.

The second sub-problem arises from the discrete spatial nature of digital images to register. As a matter of fact, the above definition of an image hides some difficulties because the image support is actually not continuous ($\Omega_I \subset \mathbb{Z}^n$), whereas its mapping via a transformation function T generally gives points in $\mathbb{R}^n$. This question will be addressed in section 2.3.4.

The third sub-problem, and the most important one for the algorithm designer, is the nature of the similarity function and the extraction of spatial features, which conditions the approaches to be used often much more than the sole nature of the mapping function. This problem will provide us a starting point of a classification of image registration techniques.

Two main classes of image registration approaches are generally considered : *feature-based methods* and *pixel-based* methods. These classes are also known as respectively *geometric registration* and *iconic registration* in the medical imaging community [CAC 02, MAL 06]. The first class is based on the correspondence between features such as points, lines, edges, or shapes of regions, which are first extracted from the target and source images. The second class looks for the direct correspondence between pixel values of both images, via Fourier-based methods, sum of squared or absolute differences, or more general information theoretic criteria.

2.2.2. *Feature-based image registration*

In feature-based or geometric image registration, the problem in Eq. (2.1) is generally simplified to :

$$T^{\star} = \arg \max_{T \in \mathcal{T}} S(T(\mathcal{A}_{I_t}), \mathcal{A}_{I_s}) \quad , \tag{2.2}$$

where $\mathcal{A}_I = \{M \in \Omega_I | F(I(M)) = 1\}$ is a set of detected interest features locations (points, lines, regions, or surfaces in the 3-D case) in image I.

Feature-based image registration usually consists of the following four stages :
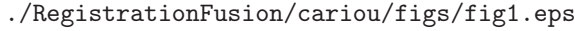
– *Feature extraction* : Salient and distinctive objects (closed-boundary regions, edges, contours, line intersections, corners, etc.) are manually or, preferably, automatically detected. For further processing, these features can be represented by their point representatives (centers of gravity, line endings, distinctive points), which are called control points in the literature.

– *Feature matching* : In this stage, the correspondence between the features detected in the source image and those detected in the target image is established. Various feature descriptors and similarity measures along with spatial relationships among the features are used for that purpose.

– *Mapping model estimation* : The nature and parameter vector $\Theta^\star$ of the transformation function are estimated. The mapping model must align the target image features with the source image features. The parameters of the mapping function are computed by means of the established feature correspondence.

– *Image transformation/re-sampling* : The source image is transformed (or *warped*) by means of the previously obtained mapping function. Image values at non-integer coordinates are computed by an appropriate interpolation technique.

A typical block diagram of feature-based image registration is shown in Fig. 2.1. One can notice that this approach is direct, i.e. the registration is performed in a one-pass fashion. The objective of the feature extraction and matching stages is to feed the mapping estimation stage with a set of co-located point set pairs to be further aligned. Thus, as such, the original pixel content is 'forgotten' during the computation of the mapping. This makes the feature-based approach insensitive to illumination variation and noise. Moreover, if a high number of features are detected in both images, this approach can allow the use of statistical techniques, such as the least squares (LS) approach, to find the best mapping parameters. The reader interested will find an excellent overview of these techniques in the reference books [GOS 05] and [HAJ 01].

The use of feature-based methods is recommended if the images contain enough distinctive and easily detectable objects. This is often the case in applications such as remote sensing or computer vision where images are supposed to contain enough details to allow the extraction of salient features. However, it may happen that images (or parts of images) do not exhibit such fine details, and rather show blurred features or very smooth gray level (or color) variations.

2.2.3. *Pixel-based image registration*

In applications where simple and salient spatial features are cannot be easily extracted from both source and target images, one may find more interesting to use a pixel-based or iconic approach. In this case, one tries to derive some matching criterion which accounts for pixels values and variations in the images, either at a global

Figure 2.1. *Typical feature-based direct image registration system.*

level or at a local level. Most pixel-based approaches use either statistical criteria like the sum of squared difference or the mutual information, or structural criteria based on invariant descriptors such as given by Fourier-based methods.

In this case, the general registration problem is reduced to :
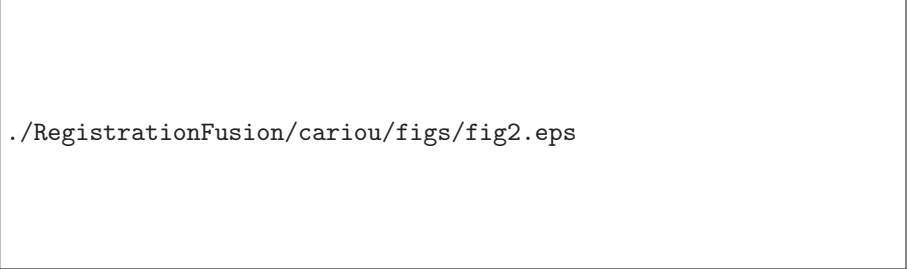
$$T^\star = \arg \max_{T \in \mathcal{T}} S(I_t(\mathbf{t}), I_s \circ T(\mathbf{t})) \quad . \tag{2.3}$$

This means that the whole (or compact sets of) the image content is used to compute the optimal mapping $T^\star$, and not only the features locations. This approach is also referred to as *dense registration* [du 05].

Whereas Fourier-based methods are well suited for matching images distorted by a global rigid transformation in a single pass, much attention has been paid in the past decade to approaches which try to estimate the correct mapping function by iterative optimization of the similarity measure S. These approaches start with an initial estimate of the mapping function $T^{(0)}$, apply it to the source image and from the image comparison stage derive an updated mapping function $T^{(1)}$, and so forth until a stopping criterion applies. A block diagram of such an iterative approach is depicted in Fig. 2.2. It is worth noting that the development of this class of registration methods was mostly impulsed by the medical imaging application field for the reasons mentioned above [MAL 06].

2.2.4. *Discussion*

Compared to feature-based approaches, pixel-based (or iconic-based) registration approaches are very interesting for many reasons. One (obvious) reason is that the registration is performed thanks to large subsets of the image (if not the whole image), and not only sparse spatial features extracted from the source and target images. Another reason is that the concept is well fitted to the image fusion process which often

./RegistrationFusion/cariou/figs/fig2.eps

Figure 2.2. *Typical pixel-based iterative image registration system.*

takes place in a further step : indeed, potentially *every pixel* in the source image is compared to a corresponding pixel in the target image, and is expected to bring its contribution to the registration process. Then, to some extent, the search for quantitative or qualitative correspondences between groups of registered pixels is likely to be more accurate with iconic approaches than feature-based approaches. Also, at last, but not at least, pixel-based approaches are rather easily extensible to multivariate, multimodal and multitemporal data, which is an important characteristic for the processing of most image types.

In the following, we will briefly present the main processing steps of image registration. We will first describe the different kinds of geometric transformations which can model the deformation between the source and target images, and how they can be set up through resampling. Since most of the present research focuses pixel-based or iconic registration, we will then deliberately restrict this section to these approaches, and describe the main similarity criteria and registration schemes which can be used.

2.3. Mapping models and image resampling

Several image mapping functions are well described in the literature and available to the user. The choice of the mapping model is essential to obtain reliable results ; this choice must be considered seriously and assessed before any processing. As mentioned above, the transformation functions depend on the nature of the scene and on the acquisition system used. However, under some assumptions, and before applying some sophisticated model, it may be interesting to approach the optimal mapping sought by successive less and less 'rigid' mapping models (or equivalently more and more parameter constrained models). We present below the most classical 2-D transformation functions by order of increasing complexity.

2.3.1. *Linear mapping models*
– Rotation, scale and translation model

The simplest mapping model relates to the so-called *rigid transformation*. More precisely, the transformation from the target to the source image coordinates is written as a combination of coordinates rotation, scaling and translation (hence the RST acronym) :

$$T : \mathbb{R}^2 \quad \rightarrow \quad \mathbb{R}^2$$

$$\mathbf{t} = [t_x, t_y]^T \quad \mapsto \quad \mathbf{s} = [s_x, s_y]^T$$

with

$$\begin{cases} s_x & = & \alpha(t_x \cos\beta + t_y \sin\beta) & + & u_x \\ s_y & = & \alpha(-t_x \sin\beta + t_y \cos\beta) & + & u_y \end{cases} , \qquad (2.4)$$

where α, β and (u_x, u_y) are respectively the scaling, rotation angle and translation vector parameters of the mapping. Equivalently, rewriting Eq. (2.4) in matrix form gives :

$$\mathbf{s} \quad = \quad \alpha\mathbf{R}\mathbf{t} + \mathbf{u} \qquad (2.5)$$

Since this mapping involves four parameters, only two pairs of corresponding points in the source and target images would be theoretically needed to compute the model (assuming no noise and perfect RST model fitting). However, it may appear some co-location errors between images, conditionally upon an RST modeling. In this case, a LS approach may be easily implemented to find the optimal parameter set.

– Affine model

A more general mapping model, called the *affine model*, is described by the following six parameters system (in matrix form) :

$$\mathbf{s} \quad = \quad \mathbf{A}\mathbf{t} + \mathbf{u} \quad , \qquad (2.6)$$

where $\mathbf{A}$ is a non singular matrix with real entries, and $\mathbf{u}$ a translation vector. In this case at least three pairs of non-aligned corresponding points in the target and source images are required to estimate the set of six parameters of the affine model. A LS approach can also be used to estimate these parameters in the case where a larger set of imperfectly aligned points is available.

The important properties of this transformation lie in its geometric invariants. Indeed, the affine transformation preserves the parallelism of lines, the ratio of areas, and the ratio of lengths on collinear or parallel segments.

– Homographic/perspective projection model

Ideally, the RST and affine mappings should be used in applications where the imaging process induces no distortion (e.g. as produced by optical lenses), and where it can be assumed that each point of an object is imaged as if it was located at the central optical axis of the sensor. This is for instance the case with (almost satellite) remote sensing of flat and moderate size areas, because of the long distance between the imaging system and the imaged scene. However, these assumptions do not longer hold when dealing with more common imaging conditions where the size of the imaged objects

and their distance to the camera are of the same order. In this case, the *homographic* (or *perspective projection*) model should be used.

The homography transformation uses the following relationships between the source and target coordinates :

$$
\begin{cases}
s_x & = & \dfrac{h_{11}\, t_x + h_{12}\, t_y + h_{13}}{h_{31}\, t_x + h_{32}\, t_y + h_{33}} \\[3mm]
s_y & = & \dfrac{h_{21}\, t_x + h_{22}\, t_y + h_{23}}{h_{31}\, t_x + h_{32}\, t_y + h_{33}}
\end{cases}
\quad ,
\tag{2.7}
$$

which can be rewritten in homogeneous coordinates by :

$$
\begin{bmatrix} s'_x \\ s'_y \\ h \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} t_x \\ t_y \\ 1 \end{bmatrix} \quad ,
\tag{2.8}
$$

and $\mathbf{s} = [s_x\ s_y]^T = [s'_x\ s'_y]^T/h$.

Since Eq. (2.8) is unique up to a scale factor, the parameter estimation problem is reduced to eight unknowns and therefore the minimum number of (non-aligned) points pairs necessary to solve the system of equations is four.

Homography has more degrees of freedom than affine transformation, and thus, less invariants. Nevertheless, some interesting properties are still preserved, such as concurrency (three lines in a projective plane are concurrent if all three of them intersect at one point), co-linearity (the dual of concurrency), tangency and inflection.

2.3.2. *Non linear mapping models*

The above models are often used as 'first order' models which can roughly approximate the deformation map with a limited computational load. However, they are inefficient when trying to estimate even smooth mapping functions, such as a quadratic deformation due to optical lenses in short focal systems. This is why more sophisticated mapping functions should be used. Non linear mappings are essential to consider problems like optical lens correction or the geocorrection of remote sensing images. They are also very useful in difficult problems involving image registration such as change detection or seismology, or the alignment of 3-D structures acquired by medical imaging systems. Several approaches can be used, and include :

– *polynomial mapping*

The polynomial model considers an extension of the above global approaches for which the target coordinates write :

$$\begin{cases} s_x & = & \displaystyle\sum_{\substack{i=0 \\ i+j\leq n}}^{n}\sum_{j=0}^{n} a_{ij}\, t_x^i\, t_y^j \\[2em] s_y & = & \displaystyle\sum_{\substack{i=0 \\ i+j\leq n}}^{n}\sum_{j=0}^{n} b_{ij}\, t_x^i\, t_y^j \end{cases}. \qquad (2.9)$$

The total number of coefficients $\{a_{ij}, b_{ij}\}$ for this model is $N = (n+1)(n+2)$. Consequently, at least $N/2$ pairs of corresponding points in the source and target images are required to estimated these parameters. The LS approach also applies in the noisy case for the polynomial model. This model is generally used with a low n value (typically less than 5) and for small and rapidly decreasing coefficients of $(i + j)$th order terms, with $(i + j) \geq 2$.

– *piecewise linear mapping* : in this case, one chooses to map some triangular mesh obtained from a set of control points in the source image, onto the corresponding mesh in the target image. These meshes can be given for instance by a Delaunay triangulation algorithm [de 08]. A linear (affine) transformation function within each corresponding triangle is then applied. This model, while simple in essence, produces continuous mappings which are not differentiable. It is often used in computer graphics but is not adapted for accurately registering remote sensing images unless acquired over flat areas with spaceborne systems.

– *radial basis function (RBF) mapping* : one important class of non linear mapping functions makes use of radial basis functions [POW 87], which are able to capture deformations at a local level. A RBF mapping is defined as follows in the 2-D case :

$$\begin{cases} s_x & = & a_{00} + a_{10}\, t_x + a_{01}\, t_y + \displaystyle\sum_{i=1}^{k} A_i\, g(\mathbf{t}, \mathbf{t}_i) \\[2em] s_y & = & b_{00} + b_{10}\, t_x + b_{01}\, t_y + \displaystyle\sum_{i=1}^{k} B_i\, g(\mathbf{t}, \mathbf{t}_i) \end{cases}, \qquad (2.10)$$

in which the first three terms relate to affine mapping whereas the last one incorporates a radially symmetric single-valued function $g(.,.)$ accounting for local deformations. This model requires a set of k control points $\{\mathbf{t}_i\}$ and associated real-valued weights $\{A_i\}$ and $\{B_i\}$. Several functions g have been proposed and studied in the literature, such as *B-splines*, *thin-plate splines (TPS)*, or *multiquadrics*. For instance, TPS require the RBF to follow $g(\mathbf{t}, \mathbf{t}_i) = d_i^2 \ln d_i$, with $d_i^2 = \|\mathbf{t} - \mathbf{t}_i\|^2 + \sigma^2$, where σ^2 controls the stiffness of the mapping surfaces $s_x(\mathbf{t}), s_y(\mathbf{t})$. Moreover, they require that three other stationarity constraints be satisfied to find the solutions $\{a_{00}; a_{10}; a_{01}; \{A_i\}_{i=1...k}\}$ and $\{b_{00}; b_{10}; b_{01}; \{B_i\}_{i=1...k}\}$.

– *Elastic mapping* : TPS mapping can be seen as an example of the wider class of elastic mapping, where no a priori parametric modeling of geometric deformations

is involved, and the set of corresponding source and target points may be extended to the whole set of pixels in both images. First proposed in [BAJ 89], elastic registration methods generally try to iteratively minimize the energy required for bending or stretching some thin elastic rubber sheet representing the displacement field $\mathbf{s}(\mathbf{t})$. Some regularity constraints are usually applied to insure the convexity of the minimization problem : they can take the form of a linear elastic modeling [HER 02], of a diffusion tensor modeling [THI 98] used in optical flow registration [NAG 86], or model the displacement field by selected Fourier coefficients [CHR 99].

2.3.3. *Hybrid mapping*

Hybrid mapping functions may be of great use in some particular cases. For instance, the imaging system may consist of a single line sensing one spatial dimension (across track) and the second spatial dimension is acquired (along track) by moving this sensor such as to scan the scene under study. This is what happens with airborne or spaceborne imagers. Assuming no distortion due to local elevation (i.e. a flat area scene), the mapping from the deformed source image to the target scene image can be modeled as global and linear in the across track spatial dimension (e.g. homographic) and locally deformed and non linear in the along track dimension. Such a transformation was referred to as 'semi-rigid' in [CAR 06].

2.3.4. *Image resampling methods*

Once obtained from the optimization of the similarity measure between the target and the source image contents, the mapping is used to warp the source image in a way that it aligns with the target image, and shares the same output reference grid. This important step in registration is called *image resampling*. The resampling can be done in a *forward* scheme, i.e. each pixel of the source image is moved to its corresponding location in the reference output coordinate system. However, this scheme is often avoided because (*i*) the alignment of the warped image coordinates on the output image reference grid is highly unlikely and can produce missing values in the resulting registration result, and (*ii*) it requires that the mapping $\mathbf{s} = T(\mathbf{t})$ be invertible, which may not always be the case. For instance in airborne spectrographic scanning systems, the pitch motion of the aircraft can produce multiple imaging of the same ground area, in which case the mapping is no longer bijective. This is why the *backward* scheme is to be preferred. In this approach, one seeks to fill the output grid with the pixel values of the source image, the coordinates of which are the images of the output grid coordinates by the direct mapping T. Though the direct mapping is in turn very unlikely to correspond to a grid point of the source image, 2-D interpolation can be used, consisting of replacing the registered output pixel value with an approximate value using a set of neighboring pixels in the source image.

2-D interpolation is generally performed by rather classical techniques such as the nearest neighbor, bilinear interpolation, bicubic interpolation, or other more or less sophisticated splines-based techniques. The reader interested in interpolation techniques will find an excellent survey in [LEH 99]. In most applications, a low level complexity interpolation approach is used, especially in iterative registration methods in which it is computed at each iteration. Interpolation with linear, cubic B-splines or sinc interpolators, followed by resampling (in $\mathbb{R}^2$) with respect to the grid of the target image can be used to produce registered images with sub-pixel accuracy, which are required to perform fusion tasks in medical imaging [**?**] or remote sensing [**?**]. In the latter paper, the authors investigated the origin of artifacts resulting in abrupt changes in the image comparison criteria and showed that these are produced by the blurring effects of the interpolators. However, in some cases such as hyperspectral airborne imaging, the sensed radiance at one pixel location should remain unaltered by any combination of the radiances of neighboring pixels ; the nearest neighbor approach must then be used, despite its low registration accuracy [LAN 02].

Figure 2.3 illustrates the application to a reference image of several mapping models T with increasing complexity.

2.4. Similarity criteria for pixel-based registration

As stated above, pixel-based registration requires the choice of a criterion for the comparison of images. Since images naturally follow some variable content, the choice of statistical tools to describe the joint dependence of pixel sets in the source and target images is clearly justified. Therefore, pixel values are treated as random variables (for gray-level images) or random vectors (for multivariate images). We give in the following the main criteria which are widely used in many applications. Only mono-component images are considered in the following, though the definitions below apply to some extent for multivariate images.

2.4.1. *Sum of squared differences*

The simplest comparison function between two images is the well-known sum of squared differences (SSD) criterion. It is defined as follows :

$$SSD(T) \;\; = \;\; \frac{1}{2} \int_{\Omega} \left[I_t(\mathbf{t}) - I_s \circ T(\mathbf{t}) \right]^2 \, d\mathbf{t} \quad , \tag{2.11}$$

This criterion is well suited to the comparison of images acquired with the same modality, i.e. with the same sensor. It has been proved that this criterion is optimal in

./RegistrationFusion/cariou/figs/mappings.eps
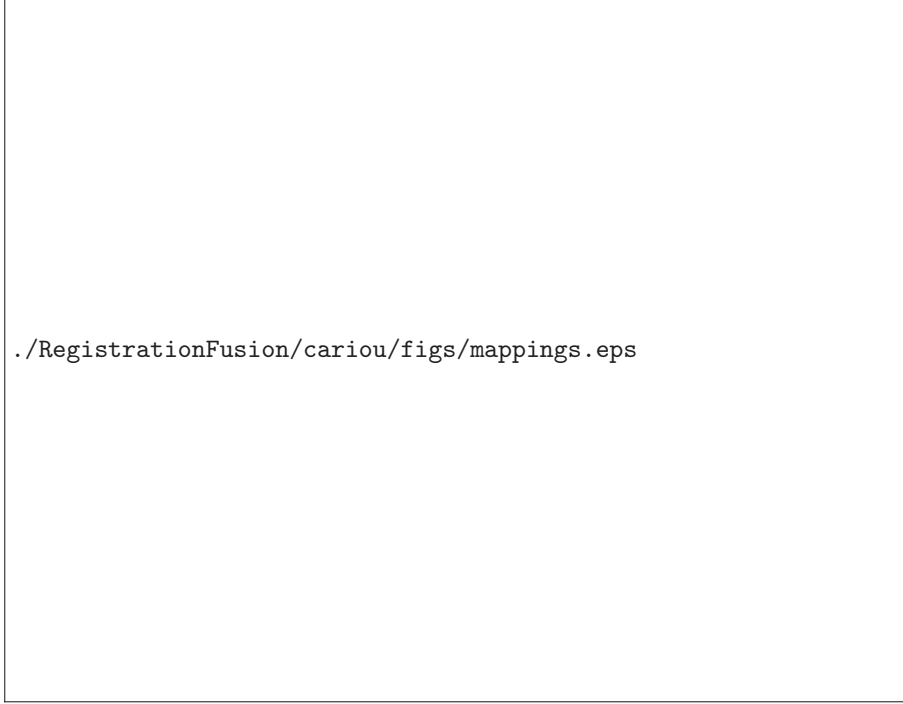
Figure 2.3. *Application of some geometric transformations to a reference image.*

the case of images disturbed only by Gaussian noise [VIO 95]. However, this method suffers from the fact that few locally misregistered pixels may lead to a high value of the criterion because of squared terms. This occurs for instance for pixels located at each side of the boundary between two uniform regions. To alleviate this difficulty, one can use the sum of absolute difference (SAD) criterion :

$$SAD(T) \quad = \quad \int_{\Omega} |I_t(\mathbf{t}) - I_s \circ T(\mathbf{t})| \ d\mathbf{t} \quad , \tag{2.12}$$

or the Huber criterion [HUB 81], a mixture of the SSD and SAD criteria which is more robust to outliers [ARY 07].

Note that, in comparison with the other criteria presented below, SSD and SAD are basically dissimilarity criteria, hence their opposite is taken as the similarity function S in Eq. 2.3.

2.4.2. *Correlation coefficient*

The correlation coefficient (also known as the normalized cross-covariance) is defined as follows :

$$CC(T) = \frac{\mathbf{E}[(I_t(\mathbf{t}) - m_{I_t})\,(I_s \circ T(\mathbf{t})) - m_{I_s})]}{\sqrt{\mathbf{V}[I_t]\,\mathbf{V}[I_s]}} \qquad , \tag{2.13}$$

where $\mathbf{E}[.]$ is the expectation operator, $\mathbf{V}[.]$ is the variance, m_{I_t} (resp. m_{I_s}) is the mean value of the target image (resp. the source image). The correlation coefficient is less restrictive than the SSD or the SAD criteria since it allows source and target images to follow a general linear relationship. It is linked to the autocovariance function used for direct pixel-based registration, as shown in Section 2.5.1.

2.4.3. *Correlation ratio*

First introduced in [ROC 98] for image registration, the correlation ratio (also called explained variance) aims at measuring the functional dependence between two random variables, and therefore is more general than the correlation coefficient. It is defined by :

$$CR(T) = \frac{\mathbf{V}[\mathbf{E}[(I_s \circ T(\mathbf{t})|I_t(\mathbf{t})]]}{\mathbf{V}[I_s \circ T(\mathbf{t})]} \qquad . \tag{2.14}$$

The correlation ratio can be seen as the proportion of variance of pixel values in the registered source image which is explained by the pixel values of the target image. Care must be taken about the relationship between the random variables : in the case of non monotonic functional dependence, a correct ordering of the random variables must be used.

2.4.4. *Mutual Information*

In many multisensor or multitemporal applications, the user is very likely to face the problem of aligning images which do not follow a linear relationship, and perhaps not even a functional relationship. For instance, in multitemporal applications, objects in the same imaged scene may have appeared or disappeared or simply moved. In such situations, the comparison criterion must be chosen to cope with large discrepancies between pixel intensities due to either the different sensors which are used for the target and source images, and also to the variability of the scene content between different acquisition dates.

Derived from Shannon's information theory, the mutual information (MI) shared by two random variables quantifies the amount of information gained between their marginal distributions and their joint distribution. It can be written as the Kullback-Leibler divergence between their joint probability density function (pdf) and the product of their marginal pdfs [COV 91]; it can also be written in terms of entropies as :

$$MI(X,Y) = H(X) - H(X|Y) \quad , \tag{2.15}$$

where $H(X) = -\int f_X(u) \log f_X(u) du$ is the entropy (or uncertainty) of X and $H(X|Y) = -\int \int f_{X,Y}(u,v) \log f_{X|Y}(u) du dv$ is the conditional entropy of X given Y. The mutual information is maximum when the entropy of one random variable conditionally on the other variable vanishes. The MI criterion is more general than the correlation ratio since it does not require any functional relationship between two random variables.

The MI criterion provides a robust and efficient tool for statistical image comparison [VIO 97] [MAE 97], and is often considered as one of the best criteria for image registration, despite its highly non-linear setting. When applied to the registration problem, the MI criterion is given by :

$$MI(T) \quad = \quad \int_{\mathbb{R}^2} f_{I_t, I_s \circ T}(u_1, u_2) \log \frac{f_{I_t, I_s \circ T}(u_1, u_2)}{f_{I_t}(u_1) . f_{I_s \circ T}(u_2)} du_1 du_2, \tag{2.16}$$

where f_{I_t} (resp. $f_{I_s \circ T}$) is the pdf of the target image (resp. of the registered source image), and $f_{I_t, I_s \circ T}$ is the joint pdf of both images.

Other similarity functions based on the mutual information have been proposed in the literature, mostly to face the problem of excessive noise in parts of the image intensity dynamics. These are called the *normalized mutual information* [HAJ 01], the most robust of which is probably the following [STU 99] [SKE 07] :

$$NMI(X,Y) = \frac{H(X) + H(Y)}{H(X,Y)} \quad , \tag{2.17}$$

where $H(X,Y)$ is the joint entropy of the random variables.

The mutual information criterion has been intensively used during the past decade in the field of medical imaging for purposes of 2-D and 3-D multimodal registration [CHR 99, MAE 97, MIL 01, HER 02, GUO 06]. It has also been applied

./RegistrationFusion/cariou/figs/fig4.eps

Figure 2.4. *Examples of joint probability density functions of two random variables showing the kind of statistical relationships compatible with the statistical pixel-based similarity criteria. (a) : SSD and SAD criteria (identity relationship) ; (b) : CC criterion (affine relationship) ; (c) : CR criterion (functional relationship) ; (d) : MI criterion (non functional relationship).*

later with success to the registration of satellite and aerial remote sensing images [KER 07, FAN 05, ING 04, CHE 03, COL 03, CAR 08a] and also to change detection [CHA 07, **?**] for multitemporal image data.

Figure 2.4 summarizes the different functional relationships which correspond to the four statistical pixel-based similarity criteria presented above.

2.5. Pixel-based registration schemes

Pixel-based registration approaches can be roughly divided into two major families. The first family comprises methods which allow to recover the optimal parameters of a given global mapping transformation in a *direct* fashion. These methods generally involve the use of signal processing techniques, and require the global mapping to follow a rather simple function such as a rigid transformation.

The second family refers to methods which seek to optimize in an iterative fashion the chosen similarity criterion, involving potentially every pixel of the source and target images. These methods can handle a broader set of geometric transformations (typically non linear ones), but their main feature lies in their ability to work without using any control point.

2.5.1. *Direct approaches*

Direct approaches can be applied in cases where the deformation between the source and target images is restricted to rigid transformation.

Most direct image registration approaches are based on the 2-D Fourier transform properties, and more precisely upon the Fourier shift theorem first suggested for the registration of images up to a translation of vector $\mathbf{u}$, $T_{\mathbf{u}} = \mathbf{Id} + \mathbf{u}$ [BRA 65] :

$$\mathcal{F}_{I \circ T_{\mathbf{u}}}(\nu) = \mathcal{F}_I(\nu) \exp(i \, 2\pi \mathbf{u}^T \nu) \ , \tag{2.18}$$

where $\mathcal{F}_I(\nu)$ is the 2-D discrete Fourier transform of image I, and $\nu \in [-0.5, 0.5] \times [-0.5, 0.5]$ is the 2-D normalized frequency [**?**].

In order to remove a translation between the source and target images, respectively I_s and I_t, one can use the two-dimensional *cross-correlation* scheme. Cross-correlation is a powerful tool for the estimation of the translation vector between two images acquired by similar sensors, i.e. when an affine relationship between the random fields can be inferred. It can easily be set up in the Fourier transform domain, leading to a fast algorithm thanks to the two-dimensional Fast Fourier Transform (2-D FFT).

More precisely, the (normalized) cross-correlation function is defined as :

$$CC(\mathbf{v}) = \frac{\mathrm{E}[I_t(\mathbf{t}) \, . \, I_s(\mathbf{t} + \mathbf{v})]}{\sqrt{\mathbf{V}[I_t] \, \mathbf{V}[I_s]}} \quad . \tag{2.19}$$

In order to fasten its calculation for large images, the sample cross-correlation function is computed as the inverse discrete Fourier transform of the cross-power spectrum :

$$CC(\mathbf{v}) = \mathcal{F}^{-1} \left[\frac{\mathcal{F}_{I_t}(\nu) \, . \, \mathcal{F}_{I_s}^{\star}(\nu)}{|\mathcal{F}_{I_t}(\nu) \, . \, \mathcal{F}_{I_s}^{\star}(\nu)|} \right] (\mathbf{v}) \quad , \tag{2.20}$$

where $\mathcal{F}^{-1}(.)$ is the 2-D (complex) inverse Fourier transform, and $\star$ stands for complex conjugation. Assuming $I_t = I_s \circ T_{\mathbf{u}}$ and plugging Eq. (2.18) in Eq. (2.20) gives $CC(\mathbf{v}) = \delta(\mathbf{v} - \mathbf{u})$, hence a Kronecker function in the spatial domain, and the location of the cross-correlation peak provides an estimate of the translation vector $\mathbf{u}$ between the source and target images.

This basic technique for translation removal has some important practical issues :

– Removing the positive mean of images is often necessary before computing the cross-correlation function : otherwise, the cross-correlation function may exhibit peak values at undesired locations. Actually, this corresponds to the computation of the 2-D cross-covariance function.

– This technique can be applied to some extent to image pairs which do not exhibit a strict linear dependence of pixel values, assuming the deviation from linearity is fair. In the case of multimodality, it is possible to first apply a high-pass filter or an edge extraction procedure on both the source and target images in a way to keep the most prominent global features of the image content [RED 96]. It can also be applied to rotated and scaled image pairs, provided the rotation angle is small and the scale factor is close to one.

– In the case of images degraded by independent additive noise, a small error in the peak location can have a large influence on the joint distribution of the target and source registered images, as can be seen in Figure 2.5.

– Since images are defined over the 2-D lattice, the discrete Fourier transform (DFT) applies and the FFT algorithm is generally used to accelerate the registration.

The location of the detected correlation peak can be inaccurate since, using the 2-D FFT, it is constrained to belong to the $\mathbb{Z}^2$ lattice. Also, it might be useful to recover the translation vector $\mathbf{u}$ with better accuracy. This can be done by *phase correlation* which considers the linearity of the phase shift $\exp(i\,2\pi\mathbf{u}^T\nu)$ with the spatial frequency : indeed, the translation vector $\mathbf{u}$ is the slope of the phase difference $\angle\{\mathcal{F}_{I_t}(\nu)/\mathcal{F}_{I_s}(\nu)\} \in [0, 2\pi[$. This approach requires phase unwrapping, which is a classical tool in radar interferometry [ZEB 98] and can be performed independently along each dimension. With this approach sub-pixel translation vector estimation is achieved [FOR 02].

We illustrate below the use of a method based on 2-D cross-correlation in the aerial remote sensing context. Fig. 2.6 shows an overlay of two contiguous overlapping flight tracks acquired by a CASI airborne multispectral/hyperspectral pushbroom imager, which are to be registered. The images are issued from a post-acquisition processing stage provided by the sensor manufacturer, and are corrected to account for the roll and pitch motions of the aircraft. One can notice the misregistration between the two images, which is due to either inaccuracy of the positional data, and variations in terrain elevation. Figure 2.7 shows the result obtained using the above Fourier-based
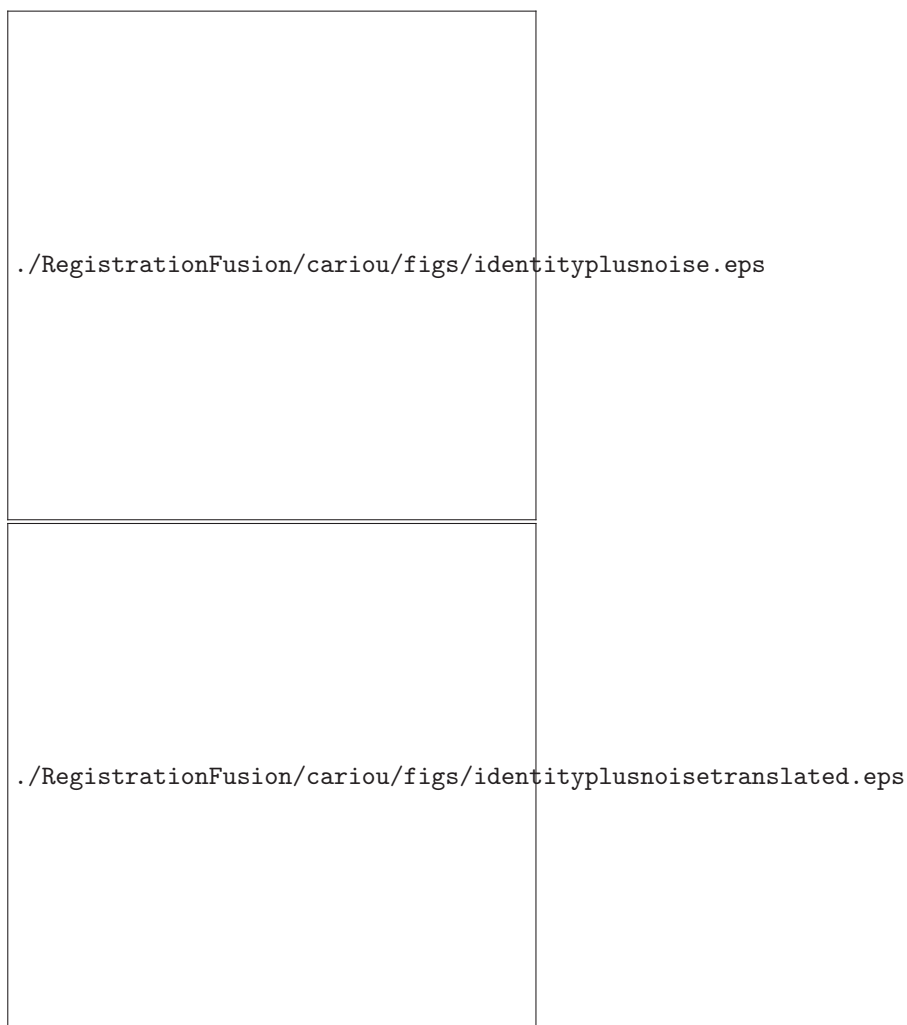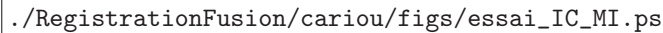
Figure 2.5. *Effect of a small misregistration on the joint distribution of source and target images. The target image is obtained from the source image by adding white Gaussian noise with variance, SNR = 33dB. Top : joint distribution for a perfect registration ; bottom : joint distribution for a translation vector* $\mathbf{u} = [0\ 1]^T$.

method on this image pair. More precisely, several translation vectors are estimated over 128×128 regularly spaced overlapping image tiles. A robust estimate of the

```
./RegistrationFusion/cariou/figs/essai_IC_MI.ps
```

Figure 2.6. *Overlay of original contiguous flight tracks. Red layer : target
image ; cyan layer : source image.*

global translation vector **u** is then computed as the mean vector of the translation vector field obtained for the tiles set, which is first median filtered. This estimate is then applied to the source image to compensate for the global translation. It is interesting to note that the translation vector estimate corresponding to each tile could be used to specify one pair of control points per tile ; a feature-based registration method could then be applied.

In the case of more general geometric transformations such as RST, one can make use of image-based invariant features. RST invariants are referred in the literature to as either log-polar features or Fourier-Mellin features [RED 96] and can be used to match images up to a similarity transformation [GHO 94] and to provide estimates of the four parameters model [ADA 01].

2.5.2. *Iterative approaches*

Iterative pixel-based approaches are mainly based on integro-differential or partial derivative equations (PDEs). They and can handle several similarity criteria such as SSD, CR and MI, and allow to tackle problems such as rigid (e.g. RST) matching [MAI 98], semi-rigid matching [CAR 06], but also matching involving continuous and differentiable 2-D displacement vector fields, and taking into account exact geometrical constraints [CAR 08a]. PDE-based registration methods are well known in medical imaging ([CHR 99], [MIL 01], [HER 02]), but have also rapidly gained popularity in remote sensing applications. [COL 03]. Their importance in multimodal/multitemporal image matching has become evidence in the last decade.

```
./RegistrationFusion/cariou/figs/resuICpc.ps
```
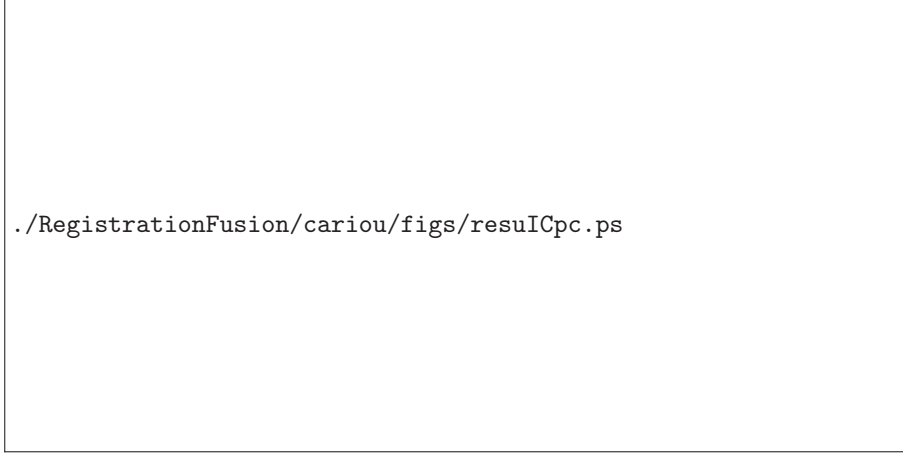
Figure 2.7. *Overlay of cross-correlation-based corrected flight bands with a pixel-based direct approach.*

Consider the pixel-based registration problem in Eq. (2.3), and assume T depends on a parameter vector Θ, hence T_{Θ}. Without loss of generality, it is often useful to transform the maximization problem into a minimization one and to superimpose on the data fitting problem some regularization constraint on the solution to yield :

$$\Theta^{\star} = \arg \min_{\Theta} \mathcal{J}(T_{\Theta}, I_s, I_t) \quad , \tag{2.21}$$

with

$$\mathcal{J}(T_{\Theta}, I_s, I_t) = -S(I_t, I_s \circ T_{\Theta}) + \lambda R(T_{\Theta}) \quad , \tag{2.22}$$

The minimization problem leads to the well know Euler-Lagrange equation :

$$\nabla_{\Theta} \mathcal{J}(T_{\Theta}, I_s, I_t) = 0 \quad , \tag{2.23}$$

which generally does not provide a direct solution $\Theta^{\star}$ due to the complexity of the objective function. The optimization is however more tractable using a *gradient descent* :

$$\begin{cases} \Theta_t & = -\nabla_{\Theta} \mathcal{J}(T_{\Theta}, I_s, I_t) \\ \Theta|_{t=0} & = \Theta_0 \end{cases} , \tag{2.24}$$

with the notation $\mathbf{\Theta}_t = \partial\mathbf{\Theta}/\partial t$. The implementation of the gradient descent consists in successive applications of the current mapping $T_{\mathbf{\Theta}}$ to the source image, and infinitesimal modification of the current mapping parameter $\mathbf{\Theta}$ based on the chosen comparison function between I_t and $I_s \circ T_{\mathbf{\Theta}}$, and so forth until convergence to a stationary point of $\mathcal{J}(T_{\mathbf{\Theta}}, I_s, I_t)$, expected to be the minimum of the objective function provided it is convex.

To illustrate this, let us consider the case of affine registration under the SSD (dis)similarity criterion. One searches for the optimal parameter set $\mathbf{\Theta} = \{\mathbf{A}^\star, \mathbf{u}^\star\}$. Assuming no regularization constraint ($\lambda = 0$) :

$$\{\mathbf{A}^\star, \mathbf{u}^\star\} = \arg \min_{\{\mathbf{A}, \mathbf{u}\}} \frac{1}{2} \int_{\Omega} \left[I_t(\mathbf{t}) - I_s(\mathbf{A}\mathbf{t} + \mathbf{u}) \right]^2 \, d\mathbf{t}. \tag{2.25}$$

It can be easily shown that the gradient descent in Eq. (2.24) writes :

$$\begin{cases} \mathbf{A}_t &= \int_{\Omega} \left[I_t(\mathbf{t}) - I_s(\mathbf{A}\mathbf{t} + \mathbf{u}) \right] \ \nabla_{\mathbf{A}} I_s(\mathbf{A}\mathbf{t} + \mathbf{u}) \, d\mathbf{t} \\ \mathbf{u}_t &= \int_{\Omega} \left[I_t(\mathbf{t}) - I_s(\mathbf{A}\mathbf{t} + \mathbf{u}) \right] \ \nabla_{\mathbf{u}} I_s(\mathbf{A}\mathbf{t} + \mathbf{u}) \, d\mathbf{t} \\ \mathbf{A}|_{t=0} &= \mathbf{I} \\ \mathbf{u}|_{t=0} &= \mathbf{0} \end{cases} \tag{2.26}$$

Therefore, the infinitesimal modification of the affine transformation model is brought by the integration of the product in the spatial domain between, on the one hand, the gradient of the currently registered source image with respect to the current model parameters and, on the other hand, the local difference between the gray level values of the target and currently registered source. Of course, if the local differences between pixels were to vanish (as might be expected when the images are perfectly registered), the variations of the parameters would also vanish and the optimal parameter $\mathbf{\Theta}^\star$ would then be found. A theoretical study of gradient-based registration performance, though restricted to translation, is available in [ROB 04].

A more complicated situation occurs when dealing with the optimization of a registration criterion with respect to a non parametric displacement vector field $\mathbf{u}(\mathbf{t})$.

In this case, a regularization term should be used to insure the convergence of the criterion to a minimum, and the criterion to optimize becomes [HER 02] :

$$\mathcal{J}(\mathbf{u}, I_s, I_t) = -S(I_t, I_s \circ (\mathbf{Id} + \mathbf{u})) + \lambda \int_\Omega \phi(D\mathbf{u})\, d\mathbf{t} \quad , \tag{2.27}$$

where $D\mathbf{u}$ is the Jacobian matrix of $\mathbf{u}$; ϕ is a real, positive-valued regularization function acting on $D\mathbf{u}$ matrices, such as the Tikhonov functional [TIK 77], the linear elasticity functional [MIL 93], or diffusion tensors [NAG 86].

Then the optimization can be performed in a variational setting through the following generic *gradient flow* :

$$\begin{cases} \mathbf{u}_t &= W(\mathbf{u}, I_s, I_t)\nabla I_s \circ (\mathbf{Id} + \mathbf{u}) + \lambda\, \mathbf{div}(\phi_{D\mathbf{u}}(D\mathbf{u})) \\ \mathbf{u}|_{t=0} &= \mathbf{u}_0 \end{cases} \tag{2.28}$$

where $W(\mathbf{u}, I_s, I_t)$ is a weighting function obtained from the computation of the first variation of the image dissimilarity criterion $-S(I_t, I_s \circ (\mathbf{Id} + \mathbf{u}))$. Weighting functions derived from the SSD, CR and MI criteria are given in [HER 02].

In the above formulation, the regularization of the displacement field is performed via an additive energy-like constraint. However, it can also be performed by smoothly filtering the variations of $\mathbf{u}$ in the following manner, inspired from [CHR 99] :

$$\begin{cases} \mathbf{u}_t &= [W(\mathbf{u}, I_s, I_t)\nabla I_s \circ (\mathbf{Id} + \mathbf{u})] * \mathbf{g} \\ \mathbf{u}|_{t=0} &= \mathbf{u}_0 \end{cases} \tag{2.29}$$

where $\mathbf{g}$ represents the 2-D impulse response of some low-pass filter.

Finally, it is important to notice that, due to their high computational load, iterative pixel-based registration approaches are very often performed in a *coarse to fine* way : by using a pyramidal or wavelet image decomposition, one estimates the transformation parameters or displacement field at the lowest decomposition level and report the result to the upper level before computing a finer estimation [ZAV 05]. Also, other criterion optimization approaches are available, such as the Levenberg-Marquardt algorithm [THE 98], the stochastic gradient [COL 03] and even evolutionary algorithms [De 07].

Figure 2.8 shows the overlay of the target and the corrected source images using a variational technique involving the gradient flow of Eq. (2.29). The initial conditions are given by the result of the cross-correlation method shown in Fig. 2.7. The registration result is far more accurate than the previous one, though some parts of the overlay do not match perfectly, due to the limitation of the filtering procedure which cannot account for high frequency content in the displacement map.

```
./RegistrationFusion/cariou/figs/resu_MI_pc.ps
```

Figure 2.8. *Overlay of non-rigid deformation corrected flight tracks using the MI criterion and a non parametric displacement vector field in a pixel-based iterative approach.*

2.5.3. *Case study : registration and mosaicking of aerial remote sensing images*

The mosaicking of Earth Observation (EO) images is the task of combining several images into an image mosaic which covers a large area and therefore can be used for mapping, land use classification, and change detection [WEH 98] [LEP 07] [TUO 04]. Although image mosaicking is a well known task for satellite imagery, where subpixel registration accuracy can be easily achieved, the case of airborne hyper-spectral pushbroom imaging systems is still very challenging due to (*i*) atmospheric turbulence with effects on the external orientation of the camera ; (*ii*) variations of the terrain height ; (*iii*) variations in the spectral radiance for same targets under different viewing conditions [SCH 02] [RIC 02] [PAL 03]. Traditionally, the mosaicking of raw images acquired by airborne pushbroom imagers (CASI, AISA, or AVIRIS for instance) is performed only once the georeferencing step is done for every image to be included in the mosaic. Georeferencing is the way of providing accurate coordinates in a given geodetic or map coordinate system to pixels of remotely sensed images. In the case of airborne pushbroom imagers there are two ways of performing the georeferencing : the first one is based on the use of a set of co-located control points in the source and the target images, from which an adequate interpolation procedure is used to georectify the raw distorted image [MCG 98]. The second one which is now commonly used is called *direct georeferencing* [HAA 00]. It requires a set of ancillary data related to the exterior orientation of the pushbroom imager and collected jointly with the raw image. These data are then mixed up in a post-acquisition processing to produce the georectified/georeferenced image. The set of ancillary data is acquired by

a GPS/INS (inertial navigation system) and provides the position (typically in Latitude/Longitude), the height above the reference ellipsoid (typically WGS84) and the attitude parameters (in roll, pitch and yaw) of the aircraft for each scan line of the raw image. However, this dataset is generally not sufficient to provide accurate positioning in the geocorrected image, and it is necessary to use a digital elevation model (DEM) in order to achieve sufficient accuracy (typically sub-pixel to few pixels RMS error), especially for wide angle cameras.

The automated registration of a raw image acquired by an airborne spectrographic imager onto a reference image requires accurate modeling of the deformation processes which are involved. Let us define $\Omega = \Omega_i \times \Omega_j$ as the set of coordinates of a raw image. $\Omega_i = \{1, \ldots, N\}$ is the set of across-track coordinates, and $\Omega_j = \{1, \ldots, M\}$ is the set of along-track coordinates. From classical photogrammetry [WOL 74], the ground coordinates (x_{ij}, y_{ij}) of a pixel with coordinates (i, j) in the raw image are given by :

$$\left[\begin{array}{c} x_{ij} \\ y_{ij} \end{array} \right] = \left[\begin{array}{cc} \cos \theta_j & -\sin \theta_j \\ \sin \theta_j & \cos \theta_j \end{array} \right] \left[\begin{array}{c} X'_{ij} \\ Y'_{ij} \end{array} \right] + \left[\begin{array}{c} X_j \\ Y_j \end{array} \right] , \tag{2.30}$$

with θ_j the instantaneous aircraft heading and (X'_{ij}, Y'_{ij}) the relative unreferenced and unoriented ground pixel coordinates

$$\left[\begin{array}{c} X'_{ij} \\ Y'_{ij} \end{array} \right] = \left[\begin{array}{c} (z_{ij} - Z_j) \dfrac{r_{12,j} c_i - r_{13,j}}{r_{32,j} c_i - r_{33,j}} \\[3mm] (z_{ij} - Z_j) \dfrac{r_{22,j} c_i - r_{23,j}}{r_{32,j} c_i - r_{33,j}} \end{array} \right] , \tag{2.31}$$

with $\mathbf{z} = [\{z_{ij}\}]_{i \in \Omega_i, j \in \Omega_j}$ the topographic height of the imaged scene ; $[\mathbf{XYZ}] = [\{X_j\}; \{Y_j\}; \{Z_j\}]^T_{j \in \Omega_j}$ are the locations of the sensor in an arbitrary coordinate system (for instance, X is the position in Easting, Y is the position in Northing, and Z is the height above the ground level) ; c_i is the ratio of the length between the ith pixel and the principal point in the image plane, and the focal length of the camera ; $c_i = \tan(\frac{FOV}{2}) \frac{2(i - i_0)}{N}$ where i_0 is the index of the pixel located at the principal point in the image plane and FOV is the field of view of the pushbroom sensor ; the coefficients $r_{kl,j}$ are given by the 3-D rotation matrix for each scan line j :

$$
\begin{bmatrix} r_{11,j} & r_{12,j} & r_{13,j} \\ r_{21,j} & r_{22,j} & r_{23,j} \\ r_{31,j} & r_{32,j} & r_{33,j} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\omega_j & -\sin\omega_j \\ 0 & \sin\omega_j & \cos\omega_j \end{bmatrix} \cdot
$$
$$
\begin{bmatrix} \cos\phi_j & 0 & \sin\phi_j \\ 0 & 1 & 0 \\ -\sin\phi_j & 0 & \cos\phi_j \end{bmatrix} \cdot \begin{bmatrix} \cos\kappa_j & \sin\kappa_j & 0 \\ -\sin\kappa_j & \cos\kappa_j & 0 \\ 0 & 0 & 1 \end{bmatrix} , \tag{2.32}
$$

where ω_j, ϕ_j and κ_j are the instantaneous absolute values of respectively the roll, pitch and yaw angles of the aircraft which define the external orientation of the camera.

Summarizing, the georeferencing of every pixel in a raw (deformed) image is possible as soon as the whole deformation process $T_{[\mathbf{X},\mathbf{Y},\mathbf{Z},\omega,\phi,\kappa,\mathbf{z}]^T}$ is known. The ancillary data is assumed accurate, but the DEM (or equivalently $\mathbf{z}$) is assumed not available or inaccurate on the area where two or more flight tracks are to be mosaicked. Note that the $\mathbf{z}$ data is in fact an interpolation of the georeferenced DEM with respect to the current deformation process.

Let us define T_1 and T_2 as the deformations undergone by the (unknown) target scene during two adjacent flight tracks. These deformations produce the roughly parallel raw images I_1 and I_2, respectively. Optimizing the MI criterion with respect to the DEM can be performed by using a variational approach, approximating elevation variations by filtered weighted gradients of the current image $I_c(\mathbf{z}_1) = \left[I_2 \circ T_2 \circ T_1^{-1} \right](\mathbf{z}_1)$. Hence the gradient flow writes, with the notations above :

$$
\begin{cases} (\mathbf{z}_1)_t = -\left[\dfrac{1}{\mu(\Omega)} \left[\psi * \dfrac{\partial L^{\mathbf{z}_1}}{\partial v} \right] (I_1, I_c(\mathbf{z}_1)) \, \nabla_i I_c(\mathbf{z}_1) \right] * \mathbf{g} \\ \mathbf{z}_1|_{t=0} = \mathbf{z}_{1_0} \end{cases} \tag{2.33}
$$

where (from [HER 02]) :

$$
\frac{\partial L^{\mathbf{z}_1}}{\partial v} = \frac{1}{f_{I_1, I_c(\mathbf{z}_1)}(u,v)} \frac{\partial f_{I_1, I_c(\mathbf{z}_1)}(u,v)}{\partial v} - \frac{1}{f_{I_c(\mathbf{z}_1)}(v)} \frac{\partial f_{I_c(\mathbf{z}_1)}(v)}{\partial v} \quad ,
$$

ψ is the kernel used to estimate the joint pdf $f_{I_1, I_c(\mathbf{z}_1)}$, and $\mathbf{g}$ is the impulse response of a 2-D low-pass filter used to regularize the solution. The initial $\mathbf{z}_{1_0}$ can be chosen as
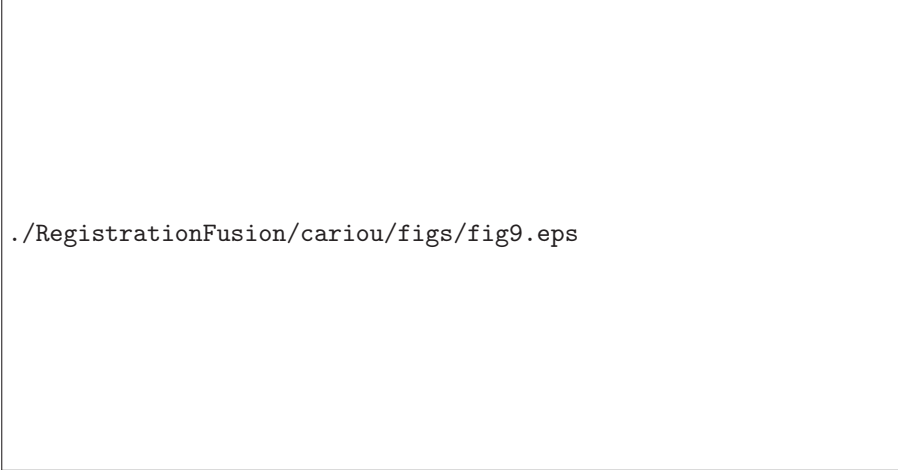
Figure 2.9. *Co-registration system and DEM estimation from two overlapping raw images acquired by a pushbroom imager, and associate ancillary measurements.*

either a constant or an interpolation of an approximate DEM when available. Note that the gradient used is relative to the coordinate index i, because the difference between I_1 and I_c which is induced by inaccurate elevation data is prominent in the across-track direction.

Fig. 2.9 depicts a system which was implemented for the fully automated co-registration of two adjacent flight tracks in [CAR 08b]. This system is divided into processing blocks. The `NAV` block computes ground map coordinates $(\mathbf{x}, \mathbf{y})$ from the ancillary and DEM data using Eqs. 2.30-2.32 for both images. The `GeoRef` block performs the georeferencing of raw images, i.e. drapes these into the chosen coordinate system. The `Interpol` block interpolates a georeferenced image with a given set of map coordinates and is in fact the reverse of the `GeoRef` function. Finally, the `Mutual Info` block computes the variations of $\mathbf{z}_1$ from Eqs. 2.33-2.34 which are then integrated and added to the initial DEM to provide the current estimate. Typically less than a hundred iterations of this system are sufficient to achieve convergence.

The mosaicking of overlapping remote sensing images is performed by selecting at each pixel location the radiance value(s) from one of the two registered images. This simple scheme avoids the mixing of data which can severely alter the expected mosaic. However, due to physical effects, the alignment of intensity values at registered pixel sites in a given spectral band is highly unlikely, but generally follows a linear function with a high correlation coefficient. A procedure then computes the mean slope

between individual pixel radiances belonging to adjacent images, and this slope is applied to normalize intensity values of one image with respect to the other. Fig. 2.10 shows a detail of a resulting mosaic of six parallel flight tracks acquired in July 2007 over a coastal zone located in Lanros, Brittany, France, with the airborne hyperspectral AISA Eagle sensor in multispectral mode (48 tuned bands). These images are part of a dataset acquired for coastal zone mapping on behalf of the CEVA (Centre d'Etude et de Valorisation des Algues, Pleubian, Brittany, France). The scan line is approximately 1000 pixels wide, and the nominal ground pixel size is 1 meter $\times$ 1 meter. The true elevation data lies the range $[0-30]$ meters. The DEM data was iteratively constructed by adding contributions of DEM corrections from track to track : the DEM estimate of one track was used as the initial condition for the registration of one other adjacent track, and the system is initialized using a zero DEM. The final planimetric RMS error computed from 100 control points on overlapping images is approximately 3 meters. The altimetric RMS error between the DEM estimate and the reference DEM provided by the Shuttle Radar Topographic Mission (SRTM, `http ://srtm.usgs.gov`) on a coarser grid (90 meters) is below 10 meters.

2.6. New trends in image registration

Historically, the research in image registration has faced a real and deep theoretical break at the end of the 1990's with the development of approaches allowing multimodality, multitemporality, and numerous geometric conditions and constraints. As previously mentioned, new advances in complex image registration problems were mostly driven by growing needs in medical imaging. This experience has been transferred with success and very rapidly to other application fields, the most important of which is probably remote sensing.

Recent research and future challenges in image registration include several aspects. From a general point of view, the only promising way to even more accurately register images is definitely to model the image acquisition process as close as possible to reality. Though complex and computationally demanding, such modeling effort is mandatory due to the increasing performance of affordable image acquisition systems (like built-in GPS/INS digital cameras) and the huge amount of data they can produce. This effort is quite well understood particularly by the remote sensing community which now can try to solve registration problems that could not be solved years ago without any human intervention. One such problem is the accurate estimation of the digital surface model (DSM) from high resolution stereo image pairs.

Another technical aspect relies in the possibility to mix up feature-based and the best of pixel-based approaches into a single statistical framework. After the works of Butz *et al.* [BUT 01] who have applied the MI criterion to spatial features like edges, several authors have proposed to extend the MI criterion to multidimensional feature vectors [HOL 04] [GAN 05]. However, the notion of MI for more than two random

./RegistrationFusion/cariou/figs/mosaic_originale.eps

./RegistrationFusion/cariou/figs/mosaic_finale.eps

Figure 2.10. *Automated mosaicking of six adjacent flight tracks. Top : mosaic without DEM estimation ; bottom : mosaic with DEM estimation.*

variables is not unique [JAK 04] and, depending on the definition chosen, cannot be used easily to derive a similarity criterion in a high dimensional feature space.

Finally, probably the most exciting research direction concerns the development of registration approaches which can mix up not only image data, but also several types of data, including qualitative variables. In the field of remote sensing, with the growing production of georeferenced data for widespread applications (transportation, administration, business, geostatistics, leisure), the registration and fusion of several data types in geographic information systems is now crucial [Le 06]. For instance, recent developments consider the joint registration of remote sensing images, technical drawings, and maps [FAN 08]. Our belief is that information theoretic tools still have an important role to play in the future for multivariate, multitemporal and multimodal image registration and fusion.

[ADA 01] ADAM S., ROUSSEAU F., OGIER J. M., CARIOU C., MULLOT R., LABICHE J., GARDES J., « A multi-scale and multi-orientation recognition technique applied to document interpretation application to French telephone network maps », *International Conference on Acoustics, Speech and Signal Processing*, p. 1509-1512, 2001.

[ARY 07] ARYA K., GUPTA P., KALRA P., MITRA P., « Image registration using robust M-estimators », *Pattern Recognition Letters*, vol. 28, n°15, p. 1957-1968, November 2007.

[BAJ 89] BAJCSY R., KOVAČIČ S., « Multiresolution elastic matching », *Computer Vision, Graphics, and Image Processing*, vol. 46, p. 1-21, 1989.

[BRA 65] BRACEWELL R., *The Fourier Transform and Its Applications*, McGraw-Hill, New York, 1965.

[BRO 92] BROWN L. G., « A survey of image registration techniques », *ACM Computing Surveys*, vol. 24, n°4, p. 325-376, 1992.

[BUT 01] BUTZ T., THIRAN J., « Affine Registration with Feature Space Mutual Information », NIESSEN W., VIERGEVER M., Eds., *MICCAI 2001*, vol. 2208 de *Lecture Notes in Computer Science*, Springer, p. 549-556, 2001.

[CAC 02] CACHIER P., Recalage non rigide d'images médicales volumiques - contribution aux approches iconiques et géométriques, Thèse de Doctorat, Ecole Centrale des Arts et Manufactures, January 2002.

[CAR 05] CARIOU C., CHEHDI K., « Recalage semi-rigide d'images de scanner aéroporté », *Actes du Colloque GRETSI 2005*, Louvain-la-Neuve, Belgique, p. 1121-1124, 2005.

[CAR 06] CARIOU C., CHEHDI K., « Semi-rigid registration of remote sensing airborne scanner images », *Proc. IEEE Int. Conference on Acoustics, Speech and Signal Processing, ICASSP 2006*, Toulouse, France, p. 825-828, May 2006.

[CAR 08a] CARIOU C., CHEHDI K., « Automatic georeferencing of airborne pushbroom scanner images with missing ancillary data using mutual information », *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, n°5, p. 1290-1300, May 2008.

[CAR 08b] CARIOU C., CHEHDI K., « Fully automated mosaicking of pushbroom aerial imagery », *Proc. IEEE Int. Conference on Acoustics, Speech and Signal Processing, ICASSP*

2008, Las Vegas, NV, USA, March 31 2008-April 4 2008, p. 1105-1108, April 2008.

[CHA 07] CHATELAIN F., TOURNERET J.-Y., INGLADA J., FERRARI A., « Bivariate gamma distributions for image registration and change detection », *IEEE Transactions on Image Processing*, vol. 16, n°7, p. 1796-1806, July 2007.

[CHE 03] CHEN H.-M., VARSHNEY P., ARORA M., « Performance of mutual information similarity measure for registration of multitemporal remote sensing images », *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, n°11, p. 2445-2454, November 2003.

[CHR 99] CHRISTENSEN G. E., « Consistent linear-elastic transformations for image matching », *Information Processing in Medical Imaging, LNCS*, vol. 1613, p. 224-237, 1999.

[COL 03] COLE-RHODES A., JOHNSON K., LE MOIGNE J., ZAVORIN I., ROBINSON D., MILANFAR P., « Multiresolution Registration of Remote Sensing Imagery by Optimization of Mutual Information Using a Stochastic Gradient », *IEEE Transactions on Image Processing*, vol. 12, n°12, p. 1495-1511, December 2003.

[COV 91] COVER T., THOMAS J., *Elements of information theory*, John Wiley & Sons, New York, NY, 1991.

[De 07] DE FALCO I., DELLA CIOPPA A., MAISTO D., SCAFURI U., TARANTINO E., « *Satellite image registration by distributed differential evolution* », vol. 4448, p. 251-260, Springer, Berlin-Heidelberg, 2007.

[de 08] DE BERG M., CHEONG O., VAN KREVELD M., OVERMARS M., *Computational Geometry : Algorithms and Applications*, Springer-Verlag, 2008.

[du 05] DU BOIS D'AISCHE A., DE CRAENE M., GEETS X., GREGOIRE V., MACQ B., WARFIELD S., « Efficient multi-modal dense field non-rigid registration : alignment of histological and section images », *Medical Image Analysis*, vol. 9, p. 538-546, 2005.

[FAN 05] FAN X., RHODY H., SABER E., « Automatic registration of multisensor airborne imagery », *Proc. 34th Applied Imagery and Pattern Recognition Workshop*, Washington, DC, USA, 2005.

[FAN 08] FAN X., RHODY H., SABER E., « A novel feature enhanced MMI based registration algorithm for automated maps and images », *Proc. IGARSS 2008*, Boston, MA, USA, 2008.

[FOR 02] FOROOSH H., ZERUBIA J., BERTHOD M., « Extension of phase correlation to subpixel registration », *IEEE Transactions on Image Processing*, vol. 11, p. 188-200, 2002.

[GAN 05] GAN R., CHUNG A. C., « Multi-dimensional mutual information based robust image registration using maximum distance-gradient-magnitude », CHRISTENSEN G., SONKA M., Eds., *IPMI 2005*, vol. 3565 de *Lecture Notes in Computer Science*, Springer, p. 210-221, 2005.

[GHO 94] GHORBEL F., « A complete invariant description for gray level images by the harmonic analysis approach », *Pattern Recognition Letters*, vol. 15, p. 1043-1051, 1994.

[GOS 05] GOSHTASBY A. A., *2-D and 3-D Image Registration for Medical, Remote Sensing, and Industrial Applications*, Wiley Publishers, Hoboken, New-Jersey, 2005.

[GUO 06] GUO Y., LU C., « Multi-modality image registration using mutual information based on gradient vector flow », *Proc. 18th Int. Conference on Pattern Recognition, ICPR*

2006, Hong-Kong, p. 697-700, 2006.

[HAA 00] HAALA N., FRITSCH D., STALLMANN D., CRAMER M., « On the performance of digital airborne pushbroom cameras for photogrammetric data processing - a case study », *International Archives of Photogrammetry and Remote Sensing*, vol. 33, part B4/1, Amsterdam, p. 320-331, 2000.

[HAJ 01] HAJNAL J., HILL D., (EDITORS) D. H., *Medical Image Registration*, CRC Press, 2001.

[HER 02] HERMOSILLO G., CHEFD'HOTEL C., FAUGERAS O., « Variational methods for multimodal image matching », *International Journal of Computer Vision*, vol. 50, n°3, p. 329-343, 2002.

[HOL 04] HOLDEN M., GRIFFIN L., SAEED N., HILL D., « Multichannel Mutual Information Using Scale Space », BARILLOT C., HAYNOR D., HELLIER P., Eds., *MICCAI 2004*, vol. 3216 de *Lecture Notes in Computer Science*, Springer, p. 797-804, 2004.

[HUB 81] HUBER P., *Robust Statistics*, Wiley, New York, 1981.

[ING 04] INGLADA J., GIROS A., « On the possibility of automatic multisensor image registration », *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, n°10, p. 2104-2120, October 2004.

[ING 07] INGLADA J., MERCIER G., « A new statistical similarity measure for change detection in multitemporal SAR images and its extension to multiscale change analysis », *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, n°5, p. 1432-1445, May 2007.

[JAK 04] JAKULIN A., BRATKO I., « Quantifying and visualizing attribute interactions : An approach based on entropy », *http ://arxiv.org/abs/cs.AI/0308002 v3*, vol. 308002, 2004.

[KER 07] KERN J., PATTICHIS M., « Robust multispectral image registration using mutual-information models », *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, n°5, p. 1494-1505, May 2007.

[LAN 02] LANDGREBE D., « Hyperspectral image data analysis », *IEEE Signal Processing Magazine*, vol. 19, n°1, p. 17-28, 2002.

[Le 06] LE MOIGNE J., COLE-RHODES A., EASTMAN R., JAIN P., JOSHUA A., MEMARSADEGHI N., MOUNT D., NETANYAHU N., MORISETTE J., UKO-OZORO E., « Image registration and fusion studies for the integration of multiple remote sensing data », *Proc. IEEE Int. Conference on Acoustics, Speech and Signal Processing, ICASSP 2006*, Toulouse, France, p. 1189-1192, May 2006.

[LEH 99] LEHMANN T., GÖNNER C., SPITZER K., « Survey : interpolation methods in medical image processing », *IEEE Transactions on Medical Imaging*, vol. 18, n°11, p. 1049-1075, November 1999.

[LEP 07] LEPRINCE S., BARBOT S., AYOUB F., AVOUAC J. P., « Automatic and precise orthorectification, coregistration, and subpixel correlation of satellite images, application to ground deformation measurements », *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, n°6, p. 1529-1558, June 2007.

[MAE 97] MAES F., COLLIGNON A., VANDERMEULEN D., MARCHAL G., SUETENS P., « Multimodality image registration by maximization of mutual information », *IEEE Transactions on Medical Imaging*, vol. 16, n°2, p. 187-198, 1997.

[MAI 98] MAINTZ J., VIERGEVER M., « A survey of medical image registration », *Medical Image Analysis*, vol. 2, n°1, p. 1-36, 1998.

[MAL 06] MALANDAIN G., Les mesures de similarité pour le recalage des images médicales, Habilitation à diriger des recherches, Université Nice Sophia-Antipolis, March 2006.

[MCG 98] MCGWIRE K., « Mosaicking airborne scanner data with the multiquadric rectification technique », *Photogrammetric Engineering and Remote Sensing*, vol. 64, n°6, p. 601-606, June 1998.

[MIL 93] MILLER M., CHRISTENSEN G., AMIT Y., GRENANDER U., « Mathematical textbook of deformable neuroanatomies », *Proceedings of the National Academy of Sciences*, vol. 90, n°24, p. 11944-11948, December 1993.

[MIL 01] MILLER M. I., YOUNES L., « Group actions, homeomorphisms, and matching : a general framework », *International Journal of Computer Vision*, vol. 41, p. 61-84, 2001.

[NAG 86] NAGEL H., ENKELMANN W., « An investigation of smoothness constraint for the estimation of displacement vector fields from images sequences », *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, p. 565-593, 1986.

[PAL 03] PALUBINSKAS G., MUELLER R., REINARTZ P., « Mosaicking of optical remote sensing images », *Proc. IEEE Int. Geoscience and Remote Sensing Symposium, IGARSS 2003*, Toulouse, France, p. 3955-3957, 2003.

[POW 87] POWELL M., « *Radial basis functions for multivariate interpolation : A review* », p. 143-167, J.C. Mason and M.G. Cox, 1987.

[RED 96] REDDY B., CHATTERJI B., « An FFT-based technique for translation, rotation and scale-invariant image registration. », *IEEE Transactions on Image Processing*, vol. 5, p. 1266-1271, 1996.

[RIC 02] RICHTER R., SCHLAEPFER D., « Geo-atmospheric Processing of Airborne Imaging Spectrometry Data Part 1 : Part 2 : Atmospheric/Topographic Correction. », *International Journal of Remote Sensing*, vol. 23, n°13, p. 2631-2649, 2002.

[ROB 04] ROBINSON D., MILANFAR P., « Fundamental performance limits in image registration », *IEEE Transactions on Image Processing*, vol. 13, n°9, p. 1185-1199, September 2004.

[ROC 98] ROCHE A., MALANDAIN G., AYACHE N., PENNEC X., Multimodal image registration by maximization of the correlation ratio, Rapport n°RR-3378, INRIA, August 1998.

[SCH 02] SCHLAEPFER D., RICHTER R., « Geo-atmospheric Processing of Airborne Imaging Spectrometry Data Part 1 : Parametric Orthorectification. », *International Journal of Remote Sensing*, vol. 23, n°13, p. 2609-2630, 2002.

[SKE 07] SKERL D., B. LIKAR J. F., PERNUŠ F., « Comparative evaluation of similarity measures for the rigid registration of multi-modal head images », *Physics in Medicine and Biology*, vol. 52, p. 5587-5601, 2007.

[STU 99] STUDHOLME C., HAWKES D., HILL D., « An overlap invariant entropy measure of 3D medical image alignment », *Pattern Recognition*, vol. 32, n°1, p. 71-86, 1999.

[THE 98] THEVENAZ P., RUTTIMAN U., UNSER M., « A pyramid approach to sub-pixel registration based on intensity », *IEEE Transactions on Image Processing*, vol. 7, n°1, p. 27-41, 1998.

[THI 98] THIRION J. P., « Image matching as a diffusion process : an analogy with Maxwell's demon », *Medical Image Analysis*, vol. 2, n°3, p. 243-260, 1998.

[TIK 77] TIKHONOV A., ARSENIN V., *Solutions of Ill-posed Problems*, Winston and Sons, Washington, D.C., 1977.

[TUO 04] TUOMINEN S., PEKKARINEN A., « Local radiometric correction of digital aerial photographs for multi source forest inventory », *Remote Sensing of Environment*, vol. 89, n°1, p. 72-82, January 2004.

[VIO 95] VIOLA P., Alignment by maximization of mutual information, PhD Thesis, Massachusetts Institute of Technology, 1995.

[VIO 97] VIOLA P., WELLS W., « Alignment by maximization of mutual information », *International Journal of Computer Vision*, vol. 24, p. 137-154, 1997.

[WEH 98] WEHN H., GOLDSTEIN N., BURKE I., BULT T., « Hyperspectral mosaicking, co-registration and change detection », *Proc. IEEE Aerospace Conference*, vol. 2, Snowmass, CO, USA, p. 13-20, 1998.

[WOL 74] WOLF P. R., *Elements of Photogrammetry (2nd ed.)*, McGraw-Hill, New-York, 1974.

[ZAV 05] ZAVORIN I., LE MOIGNE J., « Use of multiresolution wavelet feature pyramids for automatic registration of multisensor imagery », *IEEE Transactions on Image Processing*, vol. 14, n°6, p. 770-782, June 2005.

[ZEB 98] ZEBKER H., LU Y., « Phase unwrapping algorithms for radar interferometry : residue-cut, least-squares, and synthesis algorithms », *Journal of the Optical Society of America A*, vol. 15, p. 586-598, 1998.

[ZIT 03] ZITOVÁ B., FLUSSER J., « Image registration : a survey », *Image and Vision Computing*, vol. 21, n°11, p. 977-1000, October 2003.

Chapitre 3

Fusion of SAR and optical observation

3.1. Introduction and context

There are nowadays many remote sensing sensors : optical (by this we essentially mean the panchromatic image), multi or hyper-spectral sensors, SAR (Synthetic Aperture Radar), LIDAR,... They have all their own specificities and are adapted to different applications (land-use, urban planning, ground movement monitoring, Digital Elevation Model computation, etc.). But why using jointly SAR and optical sensors ? There are two main reasons : first, they hopefully provide complementary information ; secondly, SAR data only may be available in some crisis situations, but previously acquired optical data may help their interpretation.

The first point needs clarification. For human interpreters, optical images are usually really easier to interpret (see figures 3.1 and 3.2). Nevertheless, SAR data bring lots of information which are not available in optical data. For instance, the localization of urban areas is more easily seen on the SAR image (first row of figure 3.1). Beyond that, further information can be extracted if different combinations of polarization are used [CLO 97]. SAR is highly sensitive to geometrical configurations and can highlight objects appearing with a low contrast on the optical data, like flooded areas [CAL 96] or man-made objects in urban areas. Besides, polarimetric data have a high capability to discriminate phenological stages of plants like rice [ASC 96]. However, the speckle phenomenon strongly affects such signals, leading to imprecise object borders, which calls for a combination with optical data. The characteristics of optical and SAR data will be detailed and compared in the following section.

––––––––––––––

Figure 3.1. *Coarse resolution. Example of optical (SPOT, on the left) and SAR
(ERS-1, on the right) images of the city of Aix-en-Provence (France).
Resolution is approximately 10 meters for both sensors.*
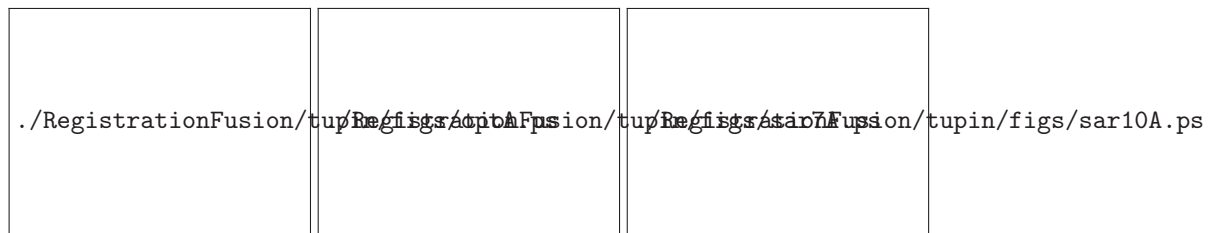
Figure 3.2. *Very high resolution (VHR). Example of optical (IGN, on the left)
and SAR (RAMSES, S-band in the middle and X-band on the right) images of a
building. Resolution is below one-meter. The speckle noise present on the SAR
images strongly affects the pixel radiometries, and the geometrical distorsions
lead to a difficult interpretation of the building.*

The second point is related to the all weather - all time capacity of SAR sensors.
Whereas many problems can be more easily solved with optical data, the availability
of such images is not guaranteed. Indeed, they can be strongly affected by atmosphe-
ric conditions and in many rainy or humid areas, useful optical images are not always
available due to the cloud cover. However, in emergency situations like natural disas-
ters (earth-quake, tsunami,...), the fast access to the data is a crucial point [WAN 05].
In such cases, additional information from optical data can drastically advance SAR
data processing, even if it is acquired at different dates and with different resolutions.
Indeed, object boundaries and area delimitations are usually stable in the landscape
and can be introduced in the SAR processing.

Nevertheless, optical and SAR fusion is not an easy task. The first step of such fusion is registration. Due to the very different appearence of the objects in both sensors, dedicated methods have been developed. This problem is studied in subsection 3.3. In the following section (3.4), some recent methods for joint classification of optical and SAR data are presented. In section 3.5, we deals with the introduction of optical information in the SAR processing. It is not "fusion" in the classical sense of the word, since both data are not considered at the same level. The application is 3D reconstruction in urban areas with high resolution data. Two different approaches based on a Markovian framework are described.

3.2. Comparison of optical and SAR sensors

SAR and optical sensors differ by essentially three points :

– both sensors are sensitive to very different features ; SAR backscattering strongly depends on the roughness of the object compared to the wavelength, the electromagnetic properties, the humidity,..., whereas the optical signal is influenced by the reflectance properties ;

– the "noise" is very different (additive for optical images and multiplicative for SAR ones) leading to different models for the radiometric distributions ;

– the geometrical distorsions caused by the acquisition systems are different, and the distance sampling of SAR sensors is very particular.

These points are enlarged when dealing with high resolution (HR) images (figure 3.2).

3.2.1. *Statistics*

Most of the optical images have some residual noise which can be well modeled by an additive white gaussian noise of zero mean. It is not at all the case for SAR signal. The interferences of the different waves reflected inside the resolution cell lead to the so-called "speckle" phenomenon strongly disturbing the SAR signal. It can be modeled as a multiplicative noise [GOO 76] following a Gamma distribution for intensity images and a Nakagami one for amplitude data. The Nakagami distribution has the following form (figure 3.3) :

$$p_A(u|L,\mu) = \frac{2}{\mu} \frac{\sqrt{L}}{\Gamma(L)} \left(\frac{\sqrt{L}u}{\mu} \right)^{2L-1} \left(e^{-\frac{\sqrt{L}u}{\mu}} \right)^2, u \geq 0 \qquad (3.1)$$

with $\mu = \sqrt{R}$ where R is proportionnal to the backscattering coefficient of the imaged pixel, and L is the number of looks (number of averaged samples to reduce the speckle effect). In case of textured areas like urban or vegetated ones, Fisher distributions are very appropriate models [TIS 04]. The shapes of such distributions with 3 parameters are illustrated in figure 3.3.

./RegistrationFusion/tupin/figs/loiNakai_a.eps ./RegistrationFusion/tupin/figs/loiFisher_a.eps

Figure 3.3. *Distribution of radiometric amplitudes in SAR images. On the left, the Nakagami distribution and on the right the Fisher distribution. Both of them have "heavy tails" [TIS 04].*

3.2.2. *Geometrical distorsions*

Due to the principle of the SAR acquisition system, the scene is distance sampled, which means that two points at the same distance from the sensors will be imaged in the same pixel. Besides, the higher an object is, the sooner it will appear in the image (see figure 3.2 and 3.4). The distance sampling leads to two effects. The first

./RegistrationFusion/tupin/figs/ov2.eps

Figure 3.4. *Geometrical distorsions due to distance sampling. The overlay part corresponds to mixed signals from ground, roof and facade of the building, whereas in the shadow area, no information is available.*

one is the lay-overs. It corresponds to areas where different signals are mixed since they are located at the same distance. The second one is the shadow areas, where no information is available due to the presence of obstacle on the electromagnetic wave path.

3.3. SAR and optical data registration

The first step before fusion is usually the registration step allowing to obtain the data in the same reference space. Two main situations can be distinguished : in the first one, the sensor parameters are well known and the projection equations can be used ; in the second one, they are not available and polynomial deformations are usually computed.

3.3.1. *Knowledge of the sensor parameters*

In this section we recall the geometrical equations of image formation for SAR and optical sensors. Let us also mention that the new products delivered by space agencies are more and more geo-coded, allowing a direct correspondance between the data but strongly relying on the used Digital Terrain Model and influenced by the interpolation function.

Transformation functions are used to project points from optical to SAR data and vice versa. They are based on the computation of the 3D coordinates of the point and on the knowledge of the sensor acquisition system parameters.

The principle of the SAR system is based on the emission of electromagnetic waves which are then backscattered by the elements lying on the ground. For a given time t of acquisition, the imaged points lie in the intersection of a sphere of range $R = ct$ and a cone related to the depointing of the antenna. More precisely, let us denote by S the sensor position, by $\vec{V}$ the speed of the sensor, and by θ_D the Doppler angle which is related to the Doppler frequency f_D and the speed by $\cos(\theta) = \frac{\lambda f_D}{2|\vec{V}|}$, the SAR equations are then given by [**?**] :

$$
\begin{aligned}
SM^2 &= R^2 \\
R\sin(\theta_D)V &= \vec{SM}.\vec{V}
\end{aligned}
$$

Knowing the line i and column j of a pixel and making a height hypothesis h, the 3D coordinates of the corresponding point M are recovered using the previous equations. R is given by the column number j, the resolution step δR, and the Nadir range R_o, by $R = j \times \delta R + R_o$. Thus the 3D point M is the intersection of a sphere with radius R, the Doppler cone of angle θ_D and a plane with altitude h. The coordinates are given as solutions of a system with 3 equations and 2 unknowns (since the height must be given).

Inversely, knowing the 3D point M allows to recover the (i, j) pixel image coordinates, by computing the sensor position for the corresponding Doppler angle (which

provides the line number) and then deducing the sensor - point distance, which permits to define the column number, since $j = \frac{R - R_o}{\delta R}$.

The geometrical model for optical image acquisition (in case of a pine-hole camera) is completely different and is based on the optical center. Each point of the image is obtained by the intersection of the image plan and the line joining the 3D point M and the optical center C. The collinear equations between the image coordinates (x_m, y_m) and the 3D point M (X_M, Y_M, Z_M) are given by :

$$
\begin{aligned}
x_m &= \frac{a_{11}X_M + a_{12}Y_M + a_{13}Z_M + a_{14}}{a_{31}X_M + a_{32}Y_M + a_{33}Z_M + a_{34}} \\
y_m &= \frac{a_{21}X_M + a_{22}Y_M + a_{23}Z_M + a_{24}}{a_{31}X_M + a_{32}Y_M + a_{33}Z_M + a_{34}}
\end{aligned}
$$

where the a_{ij} coefficients include internal and external parameters. Once again, a height hypothesis is necessary to obtain M from an image point (x_m, y_m). Figure 3.5 illustrates the two different acquisition systems.

./RegistrationFusion/tupin/figs/radar-optique.eps

Figure 3.5. *Illustration of the two different acquisition systems. A point of the SAR image is projected in the optical image for different heights.*

3.3.2. *Automatic registration*

The previous equations can only be used with a good knowledge of the sensor parameters. Many works have been dedicated to automatic registering of SAR and optical data with polynomial approaches [DAR 00] [MOI 03]. Most of them proceed in two steps : first some similarity measure between the two sensors is defined to

obtain a set of matching points ; then some optimization algorithm is used to compute the best parameters of the transformation.

The definition of similarity measures is not an easy task since as we have seen in section 3.2 the appearance of objects is very different in the two sensors. In the context of similarity measure based methods, two main approaches have been developed :

– feature-based approaches which rely on the extraction of edges or lines in both sensors [DAR 00][ING 01][LEH 08] ;

– signal-based approaches which rely on the computation of a radiometric similarity measure on local windows.

Concerning the feature-based approSaches, one of the main problem is that the features are not always similar in both sensors. For instance for VHR images, the corner between the wall and the ground of a building usually appear as a very bright line in the SAR data (see for instance figure 3.2). On the other hand, it corresponds to an edge in the optical image.

Concerning the radiometric similarity measures, different studies have been dedicated to the problem. In [ING 04] [SHA 07], some of them are analyzed and compared. One of the best criterion is the mutual entropy between the two signals.

3.4. Fusion of SAR and optical data for classification

Since the beginning of SAR imagery, there have been works dedicated to the problem of fusion with other sensors [AL. 96]. Some of them are dedicated to the extraction of some specific objects (like oil tanks [WAN 04] or buildings [TUP 03]), but the role of the SAR sensor is essentially to define region of interest or initialize the object search. In the classification framework, many different approaches have been investigated [CHE 08]. Different kinds of data can be used with SAR sensors : multitemporal series, polarimetric data, multi-frequencies data, interferometric (phase and coherence) images, depending on the application framework.

One family of methods is given by Maximum Likelihood based aproaches (and extensions) where the signals from the different sensors are concatenated in one vector. In this case, the main difficulty relies on establishing a good model for the multisource data distribution. In [LOM 03] a multivariate lognormal distribution seems to be an appropriate candidate, but multivariate gaussian distributions have also been used by other authors. More sophisticated methods introducing contextual knowledge inside Markovian framework have been developed [SOL 96]. Other works are based on the evidential theory of Dempster and Shafer to consider union of classes and represent both imprecision and uncertainty [HEG 02a] [BEN 02]. This is specially useful when

taking into account the "cloud" class in the optical images [HEG 02b]. Unsupervised approaches based on Isodata classification have also been proposed ([HIL 05] for agricultural types classification with polarimetric multi-band SAR).

Another family is given by neural networks which have been widely used for remote sensing applications [SER 95]. The 2007 data fusion contest on urban mapping using coarse SAR and optical data has been won using such a method with pre- and post-processing steps [PAC 08]. SVM approaches are also widely used for fusion [CAM 08] at the pixel level.

Instead of working at the pixel level, other methods have been developed to combine the sensor information at the decision level. The idea is to use an ensemble of classifiers and then merge them to improve the classification performances. Examples of such approaches can be found in [BRI 02] [WAS 08a] [WAS 08b].

It is not really easy to draw general conclusions concerning the performances of such methods, since the used data are usually different and also is the applicative framework. In the following section we will focus on 3D reconstruction using SAR and optical data.

3.5. Joint use of SAR interferometry and optical data for 3D reconstruction

SAR and optical data can be jointly exploited to derive 3D information. Indeed, using associated points and geometrical equations, it is possible to recover the point elevation (in [TOU 00] with manual interaction and using satellite images, in [TUP 04] with VHR images). In this part, we are interested in a different subject, i.e dealing with 3D SAR information (like interferometric or radargrammetric data) and an optical image of the same area. We have proposed a methodology based on a Markovian framework to merge information from both sensors. In such a situation, the optical data mainly provides the shapes of the building footprints whereas the SAR images provide their elevation. Let us note that the sensor parameters are supposed to be well known, and the optical data is acquired with an almost vertical viewing direction.

3.5.1. *Methodology*

The main idea of the proposed approach is to combine an over-segmentation of the optical image with 3D SAR features. Then the height of each region is computed using the SAR information and contextual knowledge expressed in a Markovian framework.

The first step is the extraction of 3D SAR information. It can be provided either by interferometric phases of points, or, as in this example, by matching of points in two SAR images (stereo-vision principle called radargrammetry). In [TUP 05], a feature

based approach is proposed. First, point-like and linear features are extracted in the two SAR images and matched. An associated height h_t is computed for each primitive t having a good matching score, defining a set S^{SAR}.

Starting from a set of regions computed on the optical data and denoted by S, a graph is defined. Each region corresponds to a node of the graph and the relationship between two regions is given by their adjacency, defining a set E of edges. The graph G is then $G = (S, E)$. For each region $s \in S$, $\mathcal{R}_s^{opt}$ is the corresponding part of the optical image. To each region s is associated a set of SAR primitives P_s such that their projection (or the projection of the middle point for segments) on the optical image belongs to $\mathcal{R}_s^{opt} : P_s = \{t \in S^{SAR}/I^{opt}(t, h_t) \in \mathcal{R}_s^{opt}\}$ with $I^{opt}(t, h_t)$ the image of the SAR primitive t projected in the optical image using the height information h_t. For segment projection, the two end-points are projected and then linked, which is not perfectly exact but is a good approximation.

One of our main assumption is that in urban areas the height surface is composed of planar patches. Because of the lack of information in our radargrammetric context, a model of flat patches (instead of planar or quadratic surfaces [MAI 92]) has been used. However, in the case of interferometric applications for instance, more complicated models could be easily introduced in the proposed framework. The problem of height reconstruction is modeled as the recovery of an height field H defined on the graph G, given a realization y of the random observation field $Y = (Y_s)_{s \in S}$. The observation y_s is given by the set of heights of $P_s : y_s = \{h_t, t \in P_s\}$. To clearly distinguish between the height field and the observation, we denote by $y_s(t)$ the height associated to $t \in P_s$, and therefore $y_s = \{y_s(t), t \in P_s\}$. To introduce contextual knowledge, H is supposed to be a Markov random field for the neighborhood defined by region adjacency. Although Markov random fields in image processing are mostly used on the pixel graph [GEM 84], they have also proved to be powerful models for feature based graph (region adjacency graph [MOD 92], characteristic point graph [REL 00], segment graph [TUP 98]). The searched realization $\hat{h}$ of H is defined to maximize the posterior probability $P(H|Y)$. Using the Bayes rule : $P(H|Y) = \frac{P(Y|H)P(H)}{P(Y)}$. If some independence assumptions are made (the observation for a region only depends on the true height of this region), the probability $P(Y|H)$ becomes :

$$P(Y|H) = \Pi_s P(Y_s|H_s) = \exp(-\sum_s -\log(P(Y_s|H_s))) = \exp(-\mathcal{U}(y|h))$$

As far as the prior $P(H)$ is concerned, we propose to use a Markovian model. Indeed, a local knowledge around a region is usually sufficient to predict its height. Therefore, H is supposed to be a Markov random field for the neighborhood defined by the adjacency relationship. This means that $P(H)$ is a Gibbs distribution and is written :

$$P(H) \propto \exp(-\mathcal{U}(h) = \exp(-\sum_{c \in C} V_c(h_s, s \in c))$$

with C the set of cliques of the graph. Using both results for $P(Y|H)$ and $P(H)$, the posterior field is also Markovian [GEM 84]. $\hat{h}$ minimizes an energy $\mathcal{U}(h, y) = \mathcal{U}(y|h) + \mathcal{U}(h)$ composed of 2 terms : a likelihood term $\mathcal{U}(y|h)$ and a prior term of regularization $\mathcal{U}(h)$.

Since the $(\mathcal{R}_s^{opt})_{s \in S}$ form a partition of the optical image, each SAR primitive belongs to a unique optical region. But many primitives can belong to the same region, and possibly with different heights. Due to the independence assumption of the regions, the likelihood term is written $\mathcal{U}(y|h) = \sum_s U_s(y_s, h_s)$. Another assumption is made about the independence of the SAR primitives conditionally to the region height h_s, which implies : $U_s(y_s, h_s) = \sum_{t \in P_s} u_s(y_s(t), h_s)$. Without real knowledge about the distribution of the SAR height primitive conditionally to h_s, a Gaussian distribution could be used, which leads to a quadratic energy. To take into account possible outliers in the height hypotheses, a truncated quadratic expression is chosen :

$$U_s(y_s, h_s) = \sum_{t \in P_s} \min[(h_s - y_s(t))^2, c]$$

This energy is zero if no SAR primitive belongs to the optical region.

The searched solution is constituted of objects (buildings) on a rather smooth ground. Besides, inside a building, the different parts should have a rather similar height. This knowledge is introduced in the definition of the clique potential of the graph. Only order 2 cliques are considered (the other clique potentials are set to zero). Two constraints are introduced in the potential definition. The first one is that the height field is naturally discontinuous. Although the height is regular inside a building or part of it, there are strong discontinuities between buildings and ground. Due to the height discontinuities, an implicit edge process is introduced. Different functions preserving discontinuities could have been used but once again a truncated quadratic function has been used.

The second constraint is related to the radiometry of the optical image. We would like to take into account the fact that a contrasted edge between two regions often implies a height discontinuity. Therefore, a weighting coefficient γ_{st} is associated to the graph edges st. This coefficient tends to 0 when the interaction between the two adjacent regions should be suppressed and 1 otherwise. The following prior energy is eventually used : $\mathcal{U}(h) = \beta \sum_{(s,t)} \gamma_{st} \min[(h_s - h_t)^2, k]$.

This energy favours configurations where adjacent regions have close heights, except if γ_{st} is small (edge presence between the 2 regions). If the the two heights are different, the penalty is limited to k, thus preserving the discontinuities naturally present in the image. The global energy is optimized using an Iterated Conditionnal Mode algorithm (ICM) [BES 86] with an initialization done by minimizing the likelihood term for each region.

The figure 3.7 shows some results obtained using the proposed methodology with the data of figure 3.6.

./RegistrationFusion/tupin/figs/bout_opt.ps ./RegistrationFusion/tupin/figs/bout_sar.ps

Figure 3.6. *Original optical image ©IGN and original SAR image ©DGA*

./RegistrationFusion/tupin/figs/resu.ps ./RegistrationFusion/tupin/figs/resu_sup.ps

Figure 3.7. *Perspective views of the result (radargrammetric framework); on the right with superimposition of the optical image.*

3.5.2. *Extension to the pixel-level*

In some recent work [DEN 08a], we have investigated a different approach working on the pixel level and more adapted to the interferometric case. In it the height

field is defined on the pixel graph and the regularization term is based on the minimization of the Total Variation. The idea is to introduce the discontinuities which are present on the optical image to weight the regularization potential. This way, the shapes of the objects on the optical image are introduced. Besides a new fast approximate optimization algorithm [DEN 08b] is used.

3.6. Conclusion

In spite of the improvement of sensor resolution, fusion of SAR and optical data remains a difficult problem. The interest on the subject has further increased with the launch of a new generation of sensors (TerraSAR-X, CosmoSkyMed, Pleiades). Although low level tools can help the interpretation process, to take the best of both sensors, high level methods have to be developed working at the object level, specially in urban areas. Indeed, the interactions of the scattering mechanisms and the geometrical distorsions require a full understanding of the local structures and approaches based on hypothesis testing and fed by SAR signal simulation tools could provide interesting answers.

[AL. 96] ET AL. R. B., « Complementary use of ERS-SAR and optical data for land cover mapping in Johor, Malaysia », *Second ERS Application Workshop, London, UK*, p. 31-35, 1996.

[ASC 96] ASCHBACHER J., PONGSRIHADULCHAI A., KARNCHANASUTHAM S., RODPROM C., PAUDYAL D., TOAN T. L., « ERS SAR data for rice crop mapping and monitoring », *Second ERS Application Workshop, London, UK*, p. 21-24, 1996.

[BEN 02] BENDJEBBOUR A., DELIGNON Y., FOUQUE L., SAMSON V., PIECZYNSKI W., « Multisensor Image segmentation Using dempster-Shafer Fusion in Markov Fields Context », *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, n°10, p. 2291-2299, 2002.

[BES 86] BESAG J., « On the statistical analysis of dirty pictures », *J. R. Statist. Soc. B*, vol. 48, n°3, p. 259-302, 1986.

[BRI 02] BRIEM G., BENEDIKTSSON J., SVEINSSON J., « Multiple classifiers applied to multisource remote sensing data », *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, n°10, p. 2291-2299, 2002.

[CAL 96] CALABRESI G., « The use of ERS data for flood monitoring : an overall assessment », *Second ERS Application Workshop, London, UK*, p. 237-241, 1996.

[CAM 08] CAMPS-VALLS G., GOMEZ-CHOVA L., MUNOZ-MARI J., ROJO-ALVAREZ J., MARTINEZ-RAMON M., SERPICO M., ROLI F., « Kernel-Based Framework for Multitemporal and Multisource Remote Sensing Data Classification and Change Detection », *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, n°6, p. 1822 - 1835, 2008.

[CHE 08] CHEN C., HO P., « Statistical Pattern Recognition in Remote Sensing », *Pattern Recognition*, vol. 41, n°9, p. 2731-2741, 2008.

[CLO 97] CLOUDE S. R., POTTIER E., « An Entropy Based Classification Scheme for Land Applications of polarimetric SAR », *IEEE Transactions on Geoscience and Remote Sensing*, vol. 35, n°1, p. 68-78, 1997.

[DAR 00] DARE P., DOWMAN I., « Automatic registration of SAR and SPOT imagery based on multiple feature extraction and matching », *IGARSS'00*, p. 24-28, 2000.

[DEN 08a] DENIS L., TUPIN F., DARBON J., SIGELLE M., « A regularization approach for InSAR and optical data fusion », *IGARSS'08*, Boston, USA, jul 2008.

[DEN 08b] DENIS L., TUPIN F., DARBON J., SIGELLE M., « SAR Image Regularization with Fast Approximate Discrete Minimization », *IEEE Transactions on Image Processing (submitted to)*, aug 2008.

[GEM 84] GEMAN S., GEMAN D., « Stochastic Relaxation, Gibbs Distribution, and the Bayesian Restoration of Images », *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-6, n°6, p. 721-741, novembre 1984.

[GOO 76] GOODMAN J., « Some fundamental properties of speckle », *Journal Optical Society of America*, vol. 66, n°11, p. 1145-1150, 1976.

[HEG 02a] HEGARAT-MASCLE S. L., BLOCH I., VIDAL-MADJAR D., « Application of Dempster-Shafer Evidence Theory to Unsupervised Classification in Multisource Remote Sensing », *IEEE Transactions on Geoscience and Remote Sensing*, vol. 35, n°4, p. 1018-1030, 2002.

[HEG 02b] HEGARAT-MASCLE S. L., BLOCH I., VIDAL-MADJAR D., « Introduction of neighborhood information in evidence theory and application to data fusion of radar and optical images with partial cloud cover », *Pattern Recognition*, vol. 40, n°10, p. 1811-1823, 2002.

[HIL 05] HILL M., TICEHURST C., LEE J.-S., GRUNES M., DONALD G., HENRY D., « Integration of optical and radar classifications for mapping pasture type in Western Australia », *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, p. 1665-1681, 2005.

[ING 01] INGLADA J., ADRAGNA F., « Automatic multi-sensor image registration by edge matching using genetic algorithms », *IGARSS'01*, p. 113-116, 2001.

[ING 04] INGLADA J., GIROS A., « On the possibility of automatic multisensor image registration », *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, n°10, p. 2104-2120, 2004.

[LEH 08] LEHUREAU G., TUPIN F., TISON C., OLLER G., PETIT D., « Registration of metric resolution SAR and Optical images in urban areas », *EUSAR 08*, jun 2008.

[LOM 03] LOMBARDO P., OLIVER C., PELLIZERI T., MELONI M., « A new Maximum-Likelihood Joint Segmentation Technique for multitemporal SAR and multiband optical images », *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, n°11, p. 2500-2518, 2003.

[MAI 92] MAITRE H., LUO W., « Using Models to Improve Stereo Reconstruction », *IEEE Transactions on Pattern Analysis and Machine Intelligence*, p. 269-277, 1992.

[MOD 92] MODESTINO J. W., ZHANG J., « A Markov Random Field Model-Based Approach to Image Interpretation », *IEEE Transactions on Pattern Analysis and Machine Intelligence*,

vol. 14, n°6, p. 606-615, 1992.

[MOI 03] MOIGNE J. L., MORISETTE J., COLE-RHODES A., NETANYAHU N., EASTMAN R., STONE H., « Earth Science Imagery Registration », *IGARSS'03*, p. 161-163, 2003.

[PAC 08] PACIFICI F., FRATE F. D., EMERY W., GAMBA P., CHANUSSOT J., « Urban Mapping Using Coarse SAR and Optical Data : Outcome of the 2007 GRSS Data Fusion Contest », *IEEE Geoscience and Remote Sensing Letters*, vol. 5, p. 331-335, 2008.

[REL 00] RELLIER G., DESCOMBES X., ZERUBIA J., « Deformation of a cartographic road network on a SPOT satellite image », *International Conference on Image Processing*, vol. 2, p. 736 - 739, 2000.

[SER 95] SERPICO S., ROLI F., « Classification of multisensor remote-sensing images by structured neural networks », *IEEE Transactions on Geoscience and Remote Sensing*, vol. 33, n°3, p. 562-578, 1995.

[SHA 07] SHABOU A., TUPIN F., CHAABANE F., « Similarity measures between SAR and optical images (submitted) », *IGARSS 2007*, 2007.

[SOL 96] SOLBERG A., TAXT T., JAIN A., « A Markov random field model for classification of multisource satellite imagery », *IEEE Transactions on Geoscience and Remote Sensing*, vol. 34, n°1, p. 100 - 113, 1996.

[TIS 04] TISON C., NICOLAS J., TUPIN F., MAÎTRE H., « A New Statistical Model of Urban Areas in High Resolution SAR Images for Markovian Segmentation », *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, n°10, p. 2046-2057, oct 2004.

[TOU 00] TOUTIN T., GRAY L., « State of the art of elevation extraction from satellite SAR data », *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 55, p. 13-33, 2000.

[TUP 98] TUPIN F., MAÎTRE H., MANGIN J.-F., NICOLAS J.-M., PECHERSKY E., « Detection of linear features in SAR images : application to road network extraction », *IEEE Transactions on Geoscience and Remote Sensing*, vol. 36, n°2, p. 434-453, mars 1998.

[TUP 03] TUPIN F., ROUX M., « Detection of building outlines based on the fusion of SAR and optical features », *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 58, n°1-2, p. 71-82, jun 2003.

[TUP 04] TUPIN F., ROUX M., « 3D information extraction by structural matching of SAR and optical features », *ISPRS'2004, Istanbul, Turquey*, 2004.

[TUP 05] TUPIN F., ROUX M., « Markov Random Field on Region Adjacency Graphs for the fusion of SAR and optical data in radargrammetric applications », *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, n°8, p. 1920-1928, aug 2005.

[WAN 04] WANG Y., TANG M., TAN T., TAI X., « Detection of circular oil tanks based on the fusion of SAR and optical images », *Third International Conference on Image and Graphics*, 2004.

[WAN 05] WANG X., WANG G., GUAN Y., CHEN Q., GAO L., « Small Satellite Constellation for Disaster Monitoring in China », *IGARSS'05*, 2005.

[WAS 08a] WASKE B., BENEDIKTSSON J., « Fusion of Support Vector Machines for Classification of Multisensor Data », *IEEE Transactions on Geoscience and Remote Sensing*,

vol. 45, n°12, p. 3858 - 3866, 2008.

[WAS 08b]   WASKE B., DER LINDEN S. V., « Classifying Multilevel Imagery From SAR and Optical Sensors by Decision Fusion », *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, n°5, p. 1457 - 1466, 2008.

Chapitre 4

Fusion of Satellite Images at Different Resolutions

4.1. Introduction

The advent of modern satellites has resulted in creation of passive sensors capable of capturing better spectral and spatial resolution than before. However, the images obtained, have either good spectral or spatial resolution. Spectral resolution, refers to the range of wavelengths captured by a sensor. A high spectral resolution indicates that the sensors can capture small intervals of the total wavelength range, thus dividing the total captured wavelength into smaller intervals. These images are referred to as Multispectral images (MS) or Hyperspectral images (HS) depending upon the spectral range of the sensor and the number of images provided by the sensor. Generally, the MS images divide the spectral range into four bands which are generally referred to as the Red, Green, Blue and Near-InfraRed (NIR) bands, whereas the HS images divide the spectral range into hundreds of bands, each capturing a very small spectral range. High spatial resolution indicates that the sensor provides sharp and detailed image, usually called a Panchromatic (Pan) image. The Pan image has high spatial resolution but no spectral diversity as compared to the MS image. However, it should be indicated here that high and low resolution are relative terms, i.e. generally the MS images have a higher spatial resolution than the HS images while a Pan image has a higher spatial resolution than the MS images. If the same object is seen in the two types of images, the Pan image with its higher spatial resolution presents the visual details of objects more accurately as compared to the MS images. Also, the human eye can see finer details in high spatial resolution. Hence Pan images are useful in applications

―――――――――

Chapitre rédigé par M. KHAN.

requiring human interpretation or supervised classification. Other important applications of remotely sensed images are unsupervised classification, object detection and object tracking. A high spectral resolution better serves applications not requiring human interpretation. Since, the satellites provide either one of the two, i.e. high spectral or spatial resolution. An interesting problem in remote sensing image processing revolves around developing images with both high spectral and spatial resolution and finding the quality of the developed image. The process of obtaining images with both high spectral and spatial resolution is either called "Fusion" or "Pansharpening".

The problem of Pansharpening has been studied for approximately three decades. However, the improving resolution of satellite sensors meant that with change of resolution the developed methods had to be changed. Methods which worked well at 30 m resolution may not necessarily work at 1m resolution. The literature on the subject is both rich and diverse. Hence, presenting a complete summary of the pansharpening methods is not easy. Also, due to different nature of the fusion algorithms, presenting a classification of the existing pansharpening methods, which is generally accepted by all the researchers is difficult. In this chapter different pansharpening algorithms are discussed based upon two criteria, i.e. (i)Detail extraction method used and (ii) Detail injection model used.

Another non-trivial problem associated with pansharpening is the quality assessment of fused images. The goal of pansharpening algorithms is to obtain a MS image which has high spatial quality, at the same scale as the Pan image, while retaining its spectral characteristics. In other words, it is now accepted in the pansharpening community that the goal of pansharpening is to obtain MS images which would have been captured if the sensor could have captured MS image at the same spatial resolution as the Pan image. Hence, for quality assessment we would like to compare our pansharpened image with the high resolution MS image. The problem is that if we had a high resolution MS image then there would be no need to construct one using pansharpening. Hence, the solution to assess the quality of fusion methods is to work at reduced resolution. However, recently quality assessment methods have been proposed which do not require a high resolution MS image as reference. Please note, that the term high resolution refers to high spatial resolution and it is assumed that the MS image is comprised of four bands namely Red, Green, Blue and NIR.

For presenting each aspect of pansharpening separately this chapter is divided into three parts. The first part presents the frequently used pansharpening model. The second part presents a classification of existing pansharpening methods, while the third part presents the different quality assessment methods.

4.2. Fusion

Pansharpening or Fusion as introduced earlier concerns improving the spatial resolution of the MS images using a high spatial resolution Pan image. The process

of fusion was defined by Lucien Wald in [WAL 02] as :" Data Fusion is a formal framework in which are expressed the means and tools for the alliance of data originating from different sources". The definition proposed by Wald covers a broad range of fusion methods including pansharpening. A more restricted and more oriented definition of pansharpening was presented by Pohl in [POH 98]" Image fusion is the combination of two or more different images to form a new image by using a certain algorithm". Hence, we can define pansharpening as " The process of extraction of spatial details from a high spatial resolution image and addition of these details into a low resolution image, using a detail injection model, with the goal of producing an image which should have been captured by the sensor if it had the high spatial resolution". The proposed definition incorporates the fact that the goal of the whole process is not to produce a MS image which is visually pleasant to the human eye but to create an image which would have been captured by the same sensor having better spatial characteristics. The fusion process is presented in Fig. 4.1.

Pansharpening can be simply defined as a process consisting of three main steps :

1) Co-registration of MS images to the Pan image

2) Extraction of spatial details from the Pan image

3) Addition of spatial details into the MS images

Before going into the details of different fusion methods it should be indicated that the pansharpening process being discussed in this chapter concerns only MS and Pan images from satellites. It is assumed that the images are already co-registered and obtained simultaneously (no temporal delay). Hence, the fusion process is reduced to only steps 2 and 3. These steps form the heart and soul of the pansharpening process. Based on the above mentioned steps we discuss a data fusion model. Most of the existing fusion methods make use of this pansharpening model to define and perform pansharpening.

4.2.1. *Pansharpening Model*

The pansharpening model proposes that the fused MS images can be obtained by adding the details, extracted from the high resolution Pan image, into upscaled MS images using a weight product. This is represented mathematically as :

$$\hat{M} = \tilde{M} + \alpha * \ddot{P} \tag{4.1}$$

For future reference $\hat{M}$ represents the pansharpened MS image, $\tilde{M}$ represents the upscaled MS image, $\ddot{P}$ represents the spatial details extracted from the Pan image and α controls the amount of details that are added to the upscaled MS image. Now the question arises why not add all the details extracted from the Pan image into the MS image ? The reason for not adding all the Pan details, with weight $\alpha = 1$, is that the

./RegistrationFusion/khan/figs/Pansharpening.eps

Figure 4.1. *Pansharpening.*

relationship between the details of the Pan image and the details of the MS images is non-linear. This occurs because different sensors are used for capturing the MS and Pan images and hence they have different optical responses. The satellite manufacturers do not provide the exact frequency response of the sensors but only the gain at Nyquist cutoff frequency. Using this value, an approximate modulation transfer function curve of the sensor response can be created. These responses are shown in Fig. 4.2 for Quickbird and simulated Pléiades satellites sensors. The frequency response for the Pléiades sensor is represented by the SMF (Starck and Murtagh Filter) frequency response. The Quickbird MS frequency responses are represented by Blue, Green, Red & Pink curves. Although not given in Fig. 4.2, the frequency response of the Pan sensor results in a Pan image which is more blurred as compared to a MS image at a given resolution. However, the Pan images provided by the satellite operators are post-processed, inverse filtered, before they are provided to the end-user. Hence, the Pan image provided is sharper than the image captured by the satellite. The resultant frequency response function for an inverse filtered panchromatic image can be represented by an approximately ideal low pass filter, which is presented as the black curve in Fig. 4.2.

The process of upscaling the MS images is also important. Comparing between zero order interpolation or pixel duplication and bicubic interpolation, it has been observed that zero order interpolation forms un-natural sharp edges in the upscaled

./RegistrationFusion/khan/figs/MTFfigure.eps

Figure 4.2. *Comparison of Frequency Response of Filters.*

low resolution MS image. Hence, it is better to use bicubic interpolation instead of zero order interpolation [AIA 02].

4.3. Fusion of Panchromatic and Multispectral Images

Most of the pansharpening methods make use of the pansharpening model discussed in Sect. 4.2.1. This model is prefered because it has been proven time and again that fusion methods using this model provide sufficiently accurate results. A comparison of fusion methods using different fusion models is not available on the same data. However, the fact that all of the methods presented in the recently conducted GRS-S 2006 IEEE Data Fusion Contest [ALP 07] used the presented pansharpening model, suggests that the pansharpening model is sufficient if not better than other complex fusion model while being computationally less expensive. Although the model used is similar, the literature on how to use this model is very rich. This diversity of different fusion methods makes it necessary to classify them. This classification helps in better understanding of the fusion process. A formal fusion method classification, which is agreed upon by the community, does not exist. The diversity makes it difficult to come up with a unanimously accepted classification. However the following classification of methods is generally accepted among the community.

1) Component Substitution (CS) based methods

2) Multiresolution Analysis (MRA) based methods

3) Hybrid methods

Making use of the above presented classification and the pansharpening model, different pansharpening methods are presented and discussed below.

4.3.1. *Component Substitution Methods*

All fusion methods which do not make use of a filtering process to extract the high frequency details from the Pan image fall in the category of component substitution methods. The principle idea is to add the details of the Pan image into the MS images making use of some transformation. Intensity-Hue-Saturation (IHS) based fusion methods, Brovey transform based fusion, PCA based fusion, all fall in the category of CS or Component Substitution Methods. We shall discuss each of these methods in more detail below.

4.3.1.1. *Brovey Transform*

The Brovey Transform is a method based on the chromaticity transform [GIL 87]. The method uses intensity modulation ratio to modulate the spatial details of the Pan image into the MS image [LIU 00]. Mathematically the fusion process can be defined as :

$$\hat{M} = \tilde{M} * \frac{P}{\tilde{I}} \tag{4.2}$$

Where, the upscaled low resolution Intensity image, $\tilde{I}$, is represented as :

$$\tilde{I} = \frac{\tilde{B} + \tilde{G} + \tilde{R}}{3} \tag{4.3}$$

The method in [GIL 87] proposes only how three band MS images can be converted into the intensity image. However, the method could be extended to fusion of satellite images which provide more than three MS bands. Since, the three band Brovey transform method is not efficient compared to the more recent fusion methods no literature is available on how the four band MS images can be converted into intensity component.

Rearranging Equ. 4.2 we get $\hat{M} = \tilde{M} + (\tilde{M} * \frac{P}{\tilde{I}} - \tilde{M})$

$$\hat{M} = \tilde{M} + \underbrace{\frac{\tilde{M}}{\tilde{I}}} (P - \tilde{I}) \tag{4.4}$$

Referring to the pansharpening model the underlined part of Equ. 4.4 represents α while the difference of high resolution Pan and the low resolution intensity image represents the Pan details.

The implementation of the method is simple :

1) Upscale the low spatial resolution MS images.

2) Generate the low resolution intensity image using the upscaled MS images using Equ. 4.3.

3) Using the modulation ratio given in Equ. 4.2 obtain each MS band separately.

Since no component is substituted during the pansharpening process this method is not a true CS method. However, the fact that the intensity image is generated from averaging the MS images and no filtering process is performed the brovey transform can be considered as a CS method.

4.3.1.2. *PCA Transform*

Principal Component Analysis is a mathematical projection of a given set of images into an orthogonal domain. The orthogonal components obtained by applying PCA are linearly uncorrelated to each other. Also, the total variance of the starting set of images is mapped onto the projected components, so that the first component corresponds to the largest amount of total variance. Further details on application of PCA to images can be obtained from [GON 92].

The implementation of the PCA method as given in [CHA 91] is as follows :

1) Upscale the low spatial resolution MS images.

2) Apply PCA to the upscaled MS images.

3) Perform histogram matching between the first principal component and the high resolution Pan image.

4) Replace the first Principal Component with the histogram matched high resolution Pan image.

5) Apply inverse PCA transform to obtain the high spatial resolution MS images.

The PCA based fusion process is graphically demonstrated in Fig. 4.3.

Proceeding mathematically the PCA transformation can be represented as :

$$\begin{bmatrix} C_1 \\ C_2 \\ ... \\ C_n \end{bmatrix} = \begin{bmatrix} a_{11} \ a_{12} \ ... \ a_{1n} \\ a_{21} \ a_{22} \ ... \ a_{2n} \\ ... \ ... \ ... \ ... \\ a_{n1} \ a_{n2} \ ... \ a_{nn} \end{bmatrix} \begin{bmatrix} \tilde{M}_1 \\ \tilde{M}_2 \\ ... \\ \tilde{M}_n \end{bmatrix} \tag{4.5}$$

Thus the PCA transformation process can be written as $C = A\,\tilde{M}$. Where *C* represents the principal components, *A* is the transformation matrix and $\tilde{M}$ is the matrix containing the upscaled low resolution MS images. For obtaining the high resolution MS images, replace the first principal component with the high resolution Pan image and perform inverse PCA transformation. The inverse transformation can be written

Figure 4.3. *Principal Component Analysis (PCA) based Fusion.*

as $\hat{M} = A^{-1}S'$. Where S' represents the matrix of principal components in which the first principal component has been replaced by the histogram matched Pan image. The transformation matrix A, composed of terms a_{ij}, consists of the eigenvectors of the covariance matrix of $\tilde{M}$ and its inverse is equal to taking its transpose. The inverse transformation matrix can be written as :

$$
\begin{bmatrix} \hat{M}_1 \\ \hat{M}_2 \\ ... \\ \hat{M}_n \end{bmatrix} = \begin{bmatrix} a_{11} \, a_{21} \, ... \, a_{n1} \\ a_{12} \, a_{22} \, ... \, a_{n2} \\ ... \, ... \, ... \, ... \\ a_{1n} \, a_{2n} \, ... \, a_{nn} \end{bmatrix} \begin{bmatrix} P \\ C_2 \\ ... \\ C_n \end{bmatrix} \tag{4.6}
$$

Where, $\hat{M}_1$ can be written as $\hat{M}_1 = a_{11}P + a_{21}C_2 + ... + a_{n1}C_n = \sum_{i=1}^{n} a_{i1}C_i$ and hence the general equation becomes :

$$
\hat{M}_j = \sum_{i=1}^{n} a_{ij}C_i \tag{4.7}
$$

It is worth noticing that $\hat{M}_1 = \underbrace{a_{11}P}_{} + a_{21}C_2 + ... + a_{n1}C_n$ helps us in explaining the fusion process. The high frequency part to be added comes from the underlined part of the equation where the complete Pan image is being added into the MS image

with the weight a_{11}. This is the reason why PCA tends to add spectral distortion in the final fused images. Only the weighted high frequency information should be added to the MS images but the PCA fusion process also adds the low frequency information of the Pan image into the fused MS image and hence the method produces sharp yet spectrally distorted fused images. It is generally accepted that before the high resolution Pan image replaces the first principal component it should be histogram matched with the first principal component. This results in an image which is generally less distorted spectrally.

4.3.1.3. *IHS Based Pansharpening*

IHS based pansharpening is among the most popular fusion methods because it is not computationally extensive. However, the IHS methods introduce spectral distortions in the fused images. Over the years the IHS based fusion methods have evolved but still the method suffers from spectral distortions. Starting with the basic IHS model

$$
\begin{bmatrix} I \\ H \\ S \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{-\sqrt{2}}{6} & \frac{-\sqrt{2}}{6} & \frac{2\sqrt{2}}{6} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{4.8}
$$

From the RGB images the Intensity, Hue and Saturation images can be obtained using Equ. 4.8. The inverse transformation, i.e. from IHS to RGB can be obtained using :

$$
\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1 & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 1 & \frac{-1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ 1 & \sqrt{2} & 0 \end{bmatrix} \begin{bmatrix} I \\ H \\ S \end{bmatrix} \tag{4.9}
$$

For preforming pansharpening, replace the R G B images with the upscaled MS images in Equ. 4.8. For obtaining the pansharpened images replace the Intensity image in Equ. 4.9 by the Pan image.

The implementation of the IHS based fusion method can be represented as :

1) Upscale the low spatial resolution MS images.

2) Perform RGB to IHS transformation.

3) Perform histogram matching between the high resolution Pan image and the upscaled low resolution intensity image.

4) Using the histogram matched Pan image perform the inverse PHS to RGB transformation.

The IHS fusion process can be demonstrated graphically as shown in Fig. 4.4.

Replacing the I, H, S and R, G, B images with their low resolution upscaled versions in Equ. 4.8 and Equ. 4.9, we can mathematically define the IHS fusion process

./RegistrationFusion/khan/figs/IHSFusion.eps
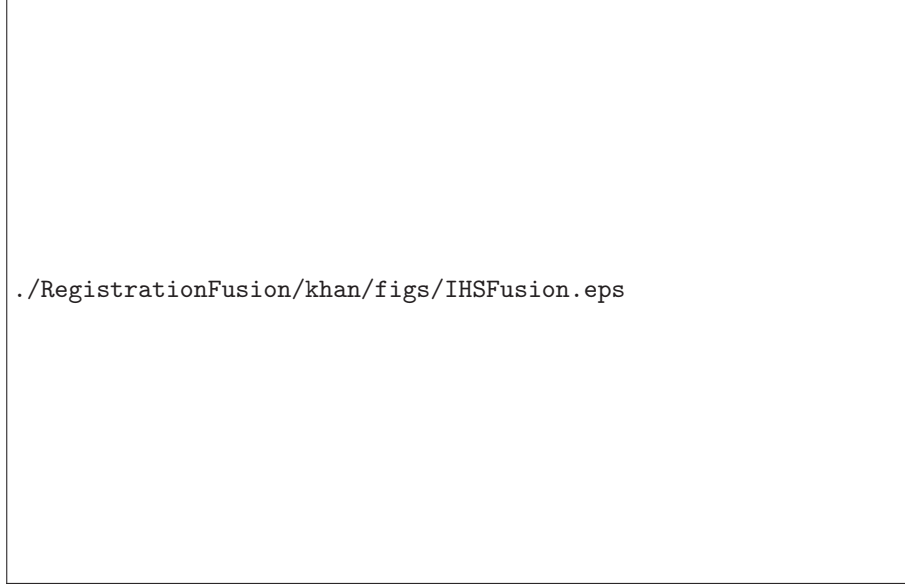
Figure 4.4. *Intensity-Hue-Saturation (IHS) based Fusion.*

without histogram matching as :

$\hat{R} = P - \frac{\tilde{H}}{\sqrt{2}} + \frac{\tilde{S}}{\sqrt{2}}$. Substituting value of Saturation and Hue from Equ. 4.8 we get

$\hat{R} = P - \frac{1}{\sqrt{2}}[\frac{\sqrt{2}}{6}(2\tilde{B} - \tilde{G} - \tilde{R})] + \frac{1}{\sqrt{2}}[\frac{1}{\sqrt{2}}(\tilde{R} - \tilde{G})]$. Solving and rearranging we get

$$\begin{bmatrix} \hat{R} \\ \hat{G} \\ \hat{B} \end{bmatrix} = \begin{bmatrix} \tilde{R} \\ \tilde{G} \\ \tilde{B} \end{bmatrix} + \begin{bmatrix} P - \frac{(\tilde{B}+\tilde{G}+\tilde{R})}{3} \\ P - \frac{(\tilde{B}+\tilde{G}+\tilde{R})}{3} \\ P - \frac{(\tilde{B}+\tilde{G}+\tilde{R})}{3} \end{bmatrix} \tag{4.10}$$

Comparing with the pansharpening model we conclude that $\alpha = 1$ and the panchromatic details are obtained as the difference between the high resolution Pan image and the intensity image formed by $(\tilde{B} + \tilde{G} + \tilde{R})/3$.

It was observed that the low resolution Pan image was introducing spectral distortion in the fused image. This occurred because of the difference in the mean and variance between the upscaled low resolution intensity image and the high resolution Pan images. Performing histogram matching the mean and variance of the high resolution Pan image were made similar to those of the low resolution intensity image. This significantly reduced the spectral distortion appearing in the fused images.

Since the original IHS method exists for only the three RGB bands there was a need to extend the method to incorporate more MS bands, as provided by the modern

satellites Quickbird and IKONOS. The solution was provided by Tu et al. in [TU 04]. For four band MS image, the fused image is obtained by :

$$\begin{bmatrix} \hat{R} \\ \hat{G} \\ \hat{B} \\ \hat{N} \end{bmatrix} = \begin{bmatrix} \tilde{R} \\ \tilde{G} \\ \tilde{B} \\ \tilde{N} \end{bmatrix} + \begin{bmatrix} P - \frac{(\tilde{R}+a*\tilde{G}+b*\tilde{B}+\tilde{N})}{3} \\ P - \frac{(\tilde{R}+a*\tilde{G}+b*\tilde{B}+\tilde{N})}{3} \\ P - \frac{(\tilde{R}+a*\tilde{G}+b*\tilde{B}+\tilde{N})}{3} \\ P - \frac{(\tilde{R}+a*\tilde{G}+b*\tilde{B}+\tilde{N})}{3} \end{bmatrix} \qquad (4.11)$$

Comparing with the pansharpening model we conclude that $\alpha = 1$ and the panchromatic details are obtained as the difference between the high resolution Pan image and the intensity image formed by $(\tilde{R} + a * \tilde{G} + b * \tilde{B} + \tilde{N})/3$. It should be noted that the authors presented their work for IKONOS sensor for which the Pan image partially covers the spectral response of the Green and Blue bands while totally covering the Red and NIR bands spectral responses. Hence, for other sensors this formula might have to be changed depending upon the spectral response of the Pan and MS sensors.

4.3.1.4. *Gram Schmidt Based Pansharpening with weighted Pan (GSWP)*

The Gram-Schmidt based fusion method can be categorized in both component substitution methods and in filter based hybrid methods. This is because two variations of the method exist, depending upon whether the low resolution Pan image is formed as a weighted sum of the low resolution MS images or by using a low pass filtered Pan image. Here only a brief discussion on the parameters obtained by Gram Schmidt (GS) fusion method is presented, with respect to the pansharpening model. For a detailed discussion on the GS orthogonalization process please refer to [LLO 97].

Assuming that the low resolution Pan image is obtained by averaging of the upscaled low resolution MS image and not by a filtering process, i.e :
$\tilde{P} = \frac{\tilde{B}+\tilde{G}+\tilde{R}+\tilde{N}}{4}$. A starting matrix can be considered as the vector representation of the low resolution Pan and the low resolution MS images. Let this matrix be a vector V. After the GS transformation the matrix obtained is G. Now matrix G is related to matrix V as : $G = VH$, where H is a square matrix equal to sum of number of Pan and MS images. The forward transformation returns us G containing the GS components. Replacing the first GS component with the high resolution Pan image and performing inverse GS transformation, $\hat{M} = G'H^{-1}$, would result in the desired high resolution MS images. Mathematically the inverse GS procedure can be defined as :

$$\begin{bmatrix} \hat{H} \\ \hat{B} \\ \hat{G} \\ \hat{R} \\ \hat{N} \end{bmatrix} = \begin{bmatrix} 1 & 0\ 0\ 0\ 0 \\ g_{bp} & 1\ 0\ 0\ 0 \\ g_{gp} & 0\ 1\ 0\ 0 \\ g_{rp} & 0\ 0\ 1\ 0 \\ g_{np} & 0\ 0\ 0\ 1 \end{bmatrix} \begin{bmatrix} P \\ GS_1 \\ GS_2 \\ GS_3 \\ GS_4 \end{bmatrix} \qquad (4.12)$$

Where $g_{bp} = \frac{\sigma(\tilde{B},\tilde{P})}{\sigma^2(\tilde{P})}$. We can generalize the terms $g_{bp}, g_{gp}, g_{rp}, g_{np}$ by g_{mp} representing the coefficients for each MS band with the Pan image. As defined in [AIA 07], the GS fusion process can be presented as :

$$\hat{M} = \tilde{M} + \frac{\sigma(\tilde{M},\tilde{P})}{\sigma^2(\tilde{P})} * (P - \tilde{P}) \tag{4.13}$$

Hence $\alpha = g_{mp}$ for each MS band and the low resolution Pan is a simple average of the upscaled low resolution MS images. $\sigma(\tilde{M},\tilde{P})$ represents covariance between the upscaled low resolution MS band and the upscaled low resolution Pan image and $\sigma^2(\tilde{P})$ represents the variance of the upscaled low resolution Pan image.

4.3.1.5. *Genetic Algorithm Based Pansharpening*

The use of genetic algorithms for pansharpening was proposed by Garzelli et al. in [GAR 06] & [GAR 07]. For clarity, the genetic algorithm does not perform the fusion process however, it is used to optimize the pansharpening algorithm output. The algorithm proposed in [GAR 07] comprises of two distinct parts, i.e. 1) Generation of a low resolution Pan image by using the coefficients determined by the genetic algorithm multiplied with their respective low resolution MS bands. This is similar to the procedure of calculating regression coefficients to generate a low resolution MS image, but in the proposed method a genetic algorithm is used to calculate the coefficients instead of a regression algorithm. 2) To determine the weight coefficients for the details, that are to be added in each MS band, the genetic algorithm is used to optimize the Q4 quality index.

The mathematical representation of the process as defined in [GAR 06] is :

$$\hat{M} = \tilde{M} + \delta(P - \sum_{k=1}^{N} \gamma_k \cdot \tilde{M}(k)) \tag{4.14}$$

Comparing with the pansharpening model $\alpha = \delta$, there are four values of δ for details that are to be added to each MS band separately and four values of γ, one each for the contribution of each MS band in forming the low resolution Pan image. Although not declared in the Data Fusion Contest [ALP 07], but looking at the results of the Data Fusion Contest it is clear that the results obtained by this method were a close third to the AWLP and CBD methods. However, the disadvantage of this method lies in the computation cost. A genetic algorithm is used to calculate the 2N coefficients required by the fusion process. For each computation the Q4 index is calculated and the process is repeated till a certain optimization of the Q4 index is reached. The method losses the computation speed that is generally associated with the CS methods. However, since no filtering process is involved the process can be classified as a component substitution method.

4.3.2. *Filter (Multiresolution) Based Methods*

The second type of fusion methods can be differentiated from the first class based upon the manner in which the details are extracted from the high resolution Pan image. Unlike the first class of methods, in these methods the Pan image is a low pass filtered image and not a weighted combination of the low resolution MS images. Also, no domain transformation is performed on the low resolution data. The methods in this class also employ a multiresolution filtering process. Thus degrading the images further than the low resolution to calculate the necessary detail injection model parameters. Finally, the estimated parameters are used to perform fusion at the desired resolution.

4.3.2.1. *SFIM*

Similar to the Brovey Transform, Smoothing Filter based Intensity Modulation (SFIM) [LIU 00] is also a method based on the use of intensity modulation ratio to modulate the spatial details of the Pan image into the MS image. Mathematically the fusion process can be defined as :

$$\hat{M} = \tilde{M} * \frac{P}{\tilde{I}} \tag{4.15}$$

Where, the Intensity image is obtained by low pass filtering of the high resolution Pan image. In [LIU 00], Liu proposes the use of the following filtering kernel for each high resolution Pan pixel :

$$\frac{1}{9} \begin{bmatrix} 1\,1\,1 \\ 1\,1\,1 \\ 1\,1\,1 \end{bmatrix} \tag{4.16}$$

It should be noted that Liu proposed to change the filtering kernel size with respect to the ratio between the low resolution and high resolution images. For example, it is proposed to use the 3x3 kernel, given above, for TM and SPOT fusion where the pixel size of the sensors is 10m and 30 m respectively. For images with a ratio of 4 the kernel size would change to 4x4.

Rewriting the equations for SFIM procedure it can be seen that the mathematical representation for SFIM and Brovey fusion processes is exactly the same. However, the difference lies in the construction of the low resolution Pan image. The process with which the low resolution Pan image is obtained directly effects the details that are extracted from the high resolution Pan image. Thus it becomes important in the context of classification of fusion methods. Hence, SFIM pansharpening process can also be defined as :

$$\hat{M} = \tilde{M} + \frac{\tilde{M}}{\tilde{I}} * (P - \tilde{I}) \tag{4.17}$$

which is similar to Equ. 4.4. However, the difference lies in the manner by which the low resolution intensity image $\tilde{I}$ is obtained.

4.3.2.2. *Wavelet Based Fusion*

In 1999, Nunez et al. [N´ 99] proposed a fusion method based on multi-resolution analysis. The authors used wavelets for decomposition of both Pan and MS images into high and low frequency components. Since, the fusion is concerned with the addition of high frequency information from the Pan image into the MS image, they replaced the high frequency component of the decomposed MS images with that of the Pan image. The high resolution MS image was obtained by an inverse wavelet transformation. There are three main variations of the wavelet based fusion proposed by Nunez et al. i) Substitution Wavelet Fusion ii)Additive Wavelet Fusion (AWRGB) iii)Additive Wavelet Fusion (AWL).

Substitution Wavelet Fusion
The fusion process is as follows :

1) Upscale the low resolution MS images.

2) Construct a low resolution Pan image, $Intensity = \frac{\tilde{B}+\tilde{G}+\tilde{R}}{3}$.

3) Perform histogram matching between the intensity image and the high resolution Pan image.

4) Decompose the high resolution Pan image into high and low frequency components, using wavelet decomposition.

5) Decompose the low resolution MS images into high and low frequency components, using wavelet decomposition.

6) Replace the high frequency component of the MS images with the high frequency component of the Pan image.

7) Perform inverse wavelet transformation to obtain the high resolution MS images.

Mathematically the wavelet transformation can be represented as follows :

$$M = \dot{M} + \sum_{k=1}^{N} \ddot{M}_k \tag{4.18}$$

$$P = \dot{P} + \sum_{k=1}^{N} \ddot{P}_k \tag{4.19}$$

Finally the fused MS image can be obtained as follows :

$$\hat{M} = \dot{M} + \sum_{k=1}^{N} \ddot{P}_k \tag{4.20}$$

Comparing with the pansharpening model it is evident that $\alpha = 1$ and $\ddot{P}$ are obtained directly from the filtered high resolution panchromatic image and k represents the different levels of the hierarchy at which the high frequency information is extracted using the wavelet transform. Hence the $Pan_{details}$ would depend upon the type of wavelet filter used. The problem with the substitution method is that the details of the MS image are discarded in the pansharpening process.

Additive Wavelet Fusion RGB

This fusion process is called AWRGB in literature and is as follows :

1) Upscale the low resolution MS images.

2) Construct a low resolution Pan image by $Intensity = \frac{\tilde{B}+\tilde{G}+\tilde{R}}{3}$.

3) Perform histogram matching between the intensity image and the high resolution Pan image.

4) Decompose the high resolution Pan image into high and low frequency components, using wavelet decomposition.

5) Add the high frequency component of the Pan image to the details of each MS image.

6) Perform inverse wavelet transformation to obtain the high resolution MS images.

Mathematically the wavelet transformation can be represented as follows :

$$M = \dot{M} + \sum_{k=1}^{N} \ddot{M}_k, P = \dot{P} + \sum_{k=1}^{N} \ddot{P}_k \qquad (4.21)$$

Finally the fused MS image can be obtained as follows :

$$\hat{M} = \dot{M} + \sum_{k=1}^{N} \ddot{P}_k + \ddot{M}_k \qquad (4.22)$$

Comparing with the pansharpening model it is evident that again $\alpha = 1$ and $\ddot{P}$ & $\ddot{M}$ are obtained directly from the filtered high resolution panchromatic & low resolution MS images. The difference from the substitution method is that the high frequency details of the Pan image are added to the MS image directly and not only to the low frequency component of the MS image. The AWRGB based fusion process is shown in Fig. 4.5. The dotted line indicates that the MS image does not exist at the highest resolution and hence we do not have the MS details at this resolution. However, the low resolution information can be generated by an upscaling process.

Additive Wavelet Fusion L

This fusion process is called AWL in literature and is as follows :

1) Upscale the low resolution MS images.

2) Transform the RGB images to LHS images.

3) Perform histogram matching between the intensity image L and the high resolution Pan image

4) Decompose the high resolution Pan image into high and low frequency components, using wavelet decomposition

5) Add the high frequency component of the Pan image to intensity image.

6) Perform LHS to RGB transformation to obtain the high resolution MS images.

The high resolution intensity image can be mathematically defined as :

$$L = \dot{L} + \sum_{k=1}^{N} \ddot{P}_k \qquad (4.23)$$

./RegistrationFusion/khan/figs/AWLPFusion.eps

Figure 4.5. *Additive Wavelet RGB (AWRGB) Fusion.*

Comparing with the pansharpening model it is evident that again $\alpha = 1$ and $\ddot{P}$ are obtained directly from the filtered high resolution panchromatic image. However,

these details are added into the intensity image which provides the details of the fused MS image on inverse transformation. The advantage of the LHS transformation is that the details of Pan image are added into an equivalent intensity image instead of the MS images separately. In [N´ 99] the authors have tested all of the three above proposed variations and found AWL as the best.

Other variations of the wavelet transform fusion have also been proposed. In [KHA 08c] the authors have used the CDF9/7 filter, the same filter used in JPEG2000 compression, for fusion. In [CHI 03] the authors have used the redundant complex wavelet transform to perform fusion. Another variation of wavelet type transformation is the Curvelet transform [GAR 05].

4.3.2.3. *WISPER*

In [OTA 05] the authors present a generalization to the AWL fusion method. The AWL method, when proposed in 1999, could only be used to fuse MS images with only three bands. However, modern satellites provide four MS images. Hence the AWL algorithm was generalized to incorporate all the MS bands and the new version of the fusion method is called AWLP. The AWLP method was an improvement over the AWL method and was rated as the joint winner alongside Context Based Decision (CBD) fusion method of the GRS-S Data Fusion Contest held in 2006 [ALP 07]. Using the AWLP method fusion can be accomplished as follows :

1) Upscale the low resolution MS images.

2) Decompose the high resolution Pan image into high and low frequency components, using wavelet decomposition.

3) Obtain high resolution MS image using $\hat{M} = \frac{\tilde{M}}{\sum_{k=1}^{N} \tilde{M}_k} \ddot{P}$.

Also, the authors proposed to use the spectral responses of sensors and the physical properties of the object during the pansharpening process. The spectral response of the sensor determines the probability that a photon of a given frequency is detected by the sensor. Also, there are different types of spectral responses, i.e. one for the Pan sensor and the other for each of the MS bands. From the spectral response of these sensors and given the number of photons detected by the high resolution Pan sensor the number of photons that the low resolution MS sensor would detect can be calculated. This can be defined mathematically as :

$$n_{MS_k} = \frac{\alpha_{Pan} \, P(p_{MS_k} | p_{PanMS})}{P(p_{PanMS} | p_{MS_k})} \ n_{Pan} \qquad (4.24)$$

Where n_{MS_k} is the number of photons detected by the MS band, α_{Pan} represents the ratio of number of photons detected by all MS bands that are also present in the residual plane of the Pan image. $\alpha_{Pan} = \frac{\sum_{k=1}^{N} n_{MS_k}}{n_{Pan_r}}$, n_{Pan} is the number of photons

detected by the Pan image, p_{PanMS} represents the probability that a photon was detected by both Pan and MS band and p_{MS_i} represents the probability that a photon was detected by the MS image. For more details on how to obtain 4.24, please refer to [OTA 05].

Returning to physical properties of the object, from the physical emission of the observed object the spectral signature factor of the low resolution MS band can be represented as :

$$s_k = \frac{\rho_{PanMS_k}}{\overline{\rho_{MS}}} \quad (4.25)$$

Where ρ_{PanMS_k} represents the photons below both MS and Pan spectral functions while $\overline{\rho_{MS}}$ is the mean for all of the MS images.

The final fused image can be obtained by using :

$$MS_{Fused_k} = MS_{Upscaled_k} + \sum_{r=1}^{n} s_k \underbrace{\frac{\alpha_{Pan}\ P(p_{MS_k}|p_{PanMS})}{P(p_{PanMS}|p_{MS_k})} \frac{(2-\beta_k)}{2}\ \ddot{Pan}_r}$$
$$(4.26)$$

Where, β_k represents the fraction of area of a MS bands spectral response function shared with its adjacent MS bands spectral response function. Comparing with the simple data fusion model the underlined part of the equation represents α, while the Pan details, $\ddot{Pan}_r$, are obtained from the wavelet filter.

4.3.2.4. *ARSIS*

Improved Spatial Resolution by Injection of Structures (ARSIS) proposed by Wald et al. [RAN 99] presents a general framework for multiresolution fusion. The ARSIS concept defines the frame work for extraction of high frequency information from the Pan image and also proposes different models which can be used for injection of high frequency details into the MS images. ARSIS in itself may not be a fusion method but a generalized description of all the methods using MRA. Hence, the MRA methods can be explained in the ARSIS framework. However, since there are a couple of methods which use ARSIS framework as their starting point and propose detail injection citing the ARSIS concept, this chapter treats ARSIS as a fusion procedure.

It is interesting to note that the ARSIS concept was presented in 1999, the same year that Nunez et al. proposed the AWRGB and AWL MRA based pansharpening methods. In the ARSIS concept the Pan details are extracted using the wavelet filter. However, the ARSIS concept does not define the number of decomposition levels necessary to obtain a good fused image. The number of wavelet decomposition levels necessary to get good fusion quality were explicitly determined by Pradhan et al. in [PRA 06]. It was noted that the best results were obtained when the number of decomposition levels were equal to or greater than the ratio of scales between the Pan and

MS images. Returning back to ARSIS concept, the main interest of the model appears in the presentation of different detail injection models. The interest of the injection model lies in the fact that the detail injection model is calculated at a degraded scale and it is assumed that this model would hold at the higher resolution. The two different scale models are referred to as the Inter Modality Model(IMM) and High Resolution Inter Modality Model (HRIMM) in the ARSIS concept.

Based upon the detail injection models the different variations of the ARSIS concept are as follows :

ARSIS-M1 [RAN 99]
All the high frequency details extracted from the high resolution Pan image are added to each upscaled MS band. This model satisfies the simple wavelet based fusion process, equivalent to a fusion process without a detail injection model. Mathematically :

$$\hat{M} = \tilde{M} + \ddot{P} \tag{4.27}$$

ARSIS-M2 [RAN 99]
The model can be defined Mathematically as :

$$\hat{M} = \tilde{M} + \underbrace{a\ddot{P} + (\overline{\ddot{M}} - a\overline{\ddot{P}})} \tag{4.28}$$

Where $a = \sqrt{\frac{\sigma^2_{\ddot{M}}}{\sigma^2_{\ddot{P}}}}$. $\ddot{M}$ represents details of the low resolution MS image and $\ddot{P}$ represents details of the low resolution Pan image. $\sigma^2_{\ddot{M}}$ represents the variance of the details of the low resolution MS image and $\overline{\ddot{M}}$ represents the mean of the details of the low resolution MS image. The underlined part of the equation represents the details that are to be added to the low frequency component of the MS image during inverse wavelet transformation.

ARSIS-M3
The model can be defined Mathematically as :

$$\hat{M} = \tilde{M} + \underbrace{a\ddot{P} + (\overline{\ddot{M}} - a\overline{\ddot{P}})} \tag{4.29}$$

Where $a = \frac{\sigma^2_{\ddot{M}} - \sigma^2_{\ddot{P}} + \sqrt{(\sigma^2_{\ddot{M}} - \sigma^2_{\ddot{P}})^2 + 4*\sigma(\ddot{M},\ddot{P})}}{2*\sigma(\ddot{M},\ddot{P})}$ and $\sigma(\ddot{M}, \ddot{P})$ represents covariance & $\sigma^2_{\ddot{M}}$ represents variance.

ARSIS-RWM [RAN 03]
Based upon two thresholds the injection model can be divided into three parts. If the

correlation between the MS and Pan image is less than the lower threshold, θ_1, then the fusion process can be defined Mathematically as :

$$\hat{M} = \tilde{M} + a_1 \ddot{P} + (\overline{\ddot{\tilde{M}}} - a_1 \overline{\ddot{\tilde{P}}}) \tag{4.30}$$

Where $a_1 = (sign)\frac{\sigma_{\ddot{\tilde{M}}}}{\sigma_{\ddot{\tilde{P}}}}$ represents α in the pansharpening model. If the correlation between the MS and Pan image is greater than the upper threshold, θ_2, then the fusion process can be defined Mathematically as :

$$\hat{M} = \tilde{M} + a_2 \ddot{P} + (\overline{\ddot{\tilde{M}}} - a_2 \overline{\ddot{\tilde{P}}}) \tag{4.31}$$

Where $a_2 = \frac{\sigma_{\ddot{\tilde{M}}}^2 - \sigma_{\ddot{\tilde{P}}}^2 + \sqrt{(\sigma_{\ddot{\tilde{M}}}^2 - \sigma_{\ddot{\tilde{P}}}^2)^2 + 4*\sigma(\ddot{\tilde{M}},\ddot{\tilde{P}})}}{2*\sigma(\ddot{\tilde{M}},\ddot{\tilde{P}})}$ represents α in the pansharpening model. If the correlation between the MS and Pan image is between the upper and lower threshold then the fusion process can be defined Mathematically as :

$$\hat{M} = \tilde{M} + a \ddot{P} + (\overline{\ddot{\tilde{M}}} - a \overline{\ddot{\tilde{P}}}) \tag{4.32}$$

Where $a = a_1 + (a_2 - a_1)(\frac{|\rho|-\theta_2}{\theta_1-\theta_2})$ and ρ represents the correlation, while θ_1 and $\theta2$ are the thresholds.

4.3.2.5. *CBD*

In [AIA 02], Aiazzi et al. proposed a filter based multiresolution fusion process called Context Driven Fusion. The authors propose the use of Laplacian pyramid for the extraction of high frequency component from the Pan image and its addition to the low resolution MS images using a context driven model. The Fusion process can be described as follows :

1) Upscale low resolution MS images.

2) Downscale high resolution Pan image.

3) Upscale the downscaled low resolution Pan images.

4) Calculate the local correlation, $LCC = \frac{\sigma(\tilde{M},\tilde{P})}{\sigma_{\tilde{M}}\sigma_{\tilde{P}}}$ between the Upscaled low resolution MS images and the Upscaled low resolution Pan image.

5) If $LCC > \theta$, θ being the threshold, then the local gain $LG = \frac{\sigma_{\tilde{M}}}{\sigma_{\tilde{P}}}$. IF $LCC < \theta$, then $LG = 0$.

6) The fused MS image can be obtained by adding the high frequency details obtained from the Pan image, multiplied by LG, to the upscaled low resolution MS image.

Hence, the fused image can be represented mathematically as :

$$\hat{M} = \tilde{M} + \frac{\sigma_{\tilde{M}}}{\sigma_{\tilde{P}}} * \ddot{P} \tag{4.33}$$

Where $\ddot{P}$ depend upon the Laplacian pyramid used to extract the high frequency information. Also, with correspondence to the pansharpening model $LG = \alpha$ if the upscaled low resolution MS image is highly correlated to the upscaled low resolution Pan image. Otherwise $\alpha = 0$ and no high frequency details are added to the upscaled low resolution MS image.

In the recently held Data Fusion Contest [ALP 07], both CBD and AWLP methods were declared as the joint winners. This emphasizes the fact that model based methods, in collaboration with MRA based detail extraction are the most performing. Although, the fused images produced by these methods are not sharp as compared to the CS methods however the fact that they do not add spectral distortions makes them more interesting methods rather than the CS methods.

4.3.3. *Hybrid Methods*

4.3.3.1. *PCA/IHS and Wavelet Merger*

In [GON 04] Audicana et al. propose the use of a component substitution method merged with a filter based multiresolution approach. The authors propose to merge PCA or IHS with a MRA based fusion process to obtain the fused images. The fused images obtained by merging the two methods are quantitatively better than the fused images obtained by PCA or IHS based methods. The fusion process can be described as :

1) Upscale the MS images.

2) Either perform RGB to IHS transformation (3 MS bands) or apply PCA transformation to the MS bands.

3) Perform histogram matching between the I intensity and the high resolution Pan image or between the first principal component C_1 and the high resolution Pan image.

4) Apply wavelet decomposition to the I or C_1 and to the high resolution Pan image.

5) Add the details of the high resolution Pan image to the details of the I or C_1 image.

6) Apply inverse wavelet transform to obtain high resolution I or C_1 components.

7) Apply IHS to RGB conversion or inverse PCA transform to obtain the high resolution MS images.

The hybrid fusion process is shown in Fig. 4.6.

From Equ. 4.10 it is known that $\hat{R} = \tilde{R} + P - \left[\frac{\tilde{R}+\tilde{G}+\tilde{B}}{3}\right]$ where $\left[\frac{\tilde{R}+\tilde{G}+\tilde{B}}{3}\right] = \tilde{I}$. For the wavelet based decomposition, the low resolution intensity Image $\tilde{I}$ and the

Figure 4.6. *PCA/IHS and Wavelet merger based fusion.*

high resolution Pan image P can be represented as $\tilde{I} = \tilde{\tilde{I}} + \sum_{k=1}^{N} \tilde{\tilde{I}}_k, P = \dot{P} + \sum_{k=1}^{N} \ddot{P}_k$. Now adding the high frequency part of the Pan image into the low resolution intensity details and inverse filtering results in a high resolution Intensity image $\hat{I} = \tilde{\tilde{I}} + \sum_{k=1}^{N}(\tilde{\tilde{I}}_k + \ddot{P}_k)$. The high resolution MS image can also be represented as $\hat{R} = \hat{I} - \frac{\tilde{H}}{\sqrt{2}} - \frac{\tilde{S}}{\sqrt{2}}$. Substituting the values of $\tilde{H}$ and $\tilde{S}$ results in the final form :

$$\hat{R} = \tilde{R} + \hat{I} - \left[\frac{\tilde{R} + \tilde{G} + \tilde{B}}{3} \right] \tag{4.34}$$

Thus, the process has $\alpha = 1$ and the details are extracted from the high resolution Pan image via wavelet filters and added to the low resolution intensity image to form a high resolution intensity image. Finally, this high resolution intensity image is used to form the high resolution MS images using an inverse IHS process.

4.3.3.2. *Gram Schmidt Based Pansharpening with filtered Pan (GSFP)*

As discussed in the previous section the only difference between this method and $GSWP$ presented in 4.3.1.4, is that the Pan image is obtained by a filtering process rather than by averaging of the upscaled low resolution MS images. In [AIA 07], Aiazzi et al. show that using a filtered Pan image results in spectrally better fused

images than when the low resolution Pan image is generated by averaging the low resolution MS images. The GSFP method uses a transformation for the actual fusion process but it uses a filter based approach to extract the high frequency information from the Pan image. Thus, it can be concluded that it is a hybrid method.

Presented above was a brief introduction to the various types of fusion methods available along with a classification of these methods. In the framework of a pansharpening model different detail extraction methods along with their injection models were discussed. It is obvious that the visual fusion results would depend upon how much of the details are extracted from the Pan image and what percentage of these details is added to the MS images. If all the details are added then the resultant image will be visually sharp but may suffer from spectral distortion. To the contrary, if the details extracted are added using a model and only a specific percentage is added to the MS image the final image would be slightly blurred but there would be no spectral distortion. From this discussion it would seem that there is a trade-off between spectral and spatial quality. However, this is not true. If satellites capture a high resolution MS image, it has both high spectral and spatial quality. Hence, there is no trade-off between the two in the original image. The synthesized (fused) MS image may demonstrate this trade-off because of the inadequacy of either the filters used for extraction of details or upscaling of MS images or because of the poor model selection for the injection of details.

The question may arise what are the advantages of pansharpening ? The answer is, pansharpened images not only improve the visual interpretation of the MS images but also improve the performance of automated feature extraction tasks such as classification and segmentation. In [SIR 08] the authors have demonstrated how the use of pansharpening helps in a better mapping of snow. The authors used MODIS data for mapping snow in the region of Mount Cook in New Zealand. The authors had a reference high resolution MS image of the same region obtained from ASTER satellite. So a comparison of improvement of mapping accuracy could be obtained. The authors used (Receiver Operating Characteristics) ROC curves and Q index type measures to assess the improvement and found that ROC demonstrated an improvement of approximately 9% where as Q index demonstrated an improvement of approximately 4% in the mapping accuracy. This is also clear from the images presented in Fig. 4.7. In [COL 06] the authors have demonstrated the influence of pansharpening on the accuracy of classification.

4.4. Fusion Quality Assessment

The second part of the fusion process is to determine the quality of the fused images. In image processing quality of the processed image is usually determined by comparing it with a reference image. The problem with determining the quality of pansharpened images is that no reference image is available for making a comparison.

./RegistrationFusion/khan/figs/ieee_grsl_08_sirguey_snowcover.eps

Figure 4.7. *Visual performance in estimating the snow fraction (Mount Cook Area, South Island, New Zealand ; December 2002). (a) False color composite MODIS bands 1, 2, 3 (fused to 250 m). (b) Reference snow map extracted from ASTER at degraded (250 m). (c) MODIS-derived snow map without fusion at 500 m. (d) MODIS-derived snow map with fused bands at 250 m. ©IEEE 2008*

Hence, the quality of fusion is determined at reduced resolution and it is assumed that this quality would hold at the higher resolution. Determining the quality at reduced resolution refers to the fact that the reference images are reduced to a lower resolution and then fusion is performed. Now the produced fused images have the same resolution as the starting reference MS images. Hence, a comparison between the fused and reference MS images can be made. Before providing the methods for determining the quality of fused images it is important to present the three properties that should be satisfied by fused images of good quality. These properties were presented by Lucien Wald in [WAL 02].

1) Consistency Property : The fused image when degraded to the low resolution should be identical to the starting reference.

2) Synthesis Property : Each fused image should be similar to the high resolution MS image, which would have been captured if the sensor had the capability to capture

MS images at high resolution.

3) Synthesis Set Property : The set of fused images should be similar to the set of high resolution MS images, which would have been captured if the sensor had the capability to capture the set of MS images at high resolution.

The fusion quality assessment methods are generally divided into methods which require a reference MS image for determining the quality and those methods which do not require a reference MS image. The later methods have appeared just recently and are not yet mature. However, they provide the advantage of determining the quality of the fused image at high resolution rather than working at degraded scale. Both types of quality assessment methods are presented below.

4.4.1. *Methods Requiring a Reference Image*

The quality assessment methods which require a reference MS image fall in the category of conventional quality assessment methods. In the Data Fusion Community these methods are unanimously/undisputably accepted as the standard methods for determining the quality of fused images. However, as mentioned earlier, these methods require a reference image and hence require that the fusion quality is assessed by starting at a degraded resolution. Some examples are Spectral Angle Mapper (SAM), Global Relative Error in Synthesis (ERGAS) and Quaternion Index (Q4). These methods are presented in detail below.

4.4.1.1. *SAM*

SAM or Spectral Angle Mapper is used to determine if the fused image has spectral distortions. SAM calculates an angle between the spectral vector of the fused image and the reference MS image. Since, applications like classification require that the spectral information does not change after fusion, fused images having low SAM value tend to produce better classification results [THO 06]. The ideal value of SAM index is zero. Since this index is a distance measure, the larger the value of SAM the poorer is the spectral quality of the fused image with respect to the reference MS image. SAM can be calculated for each pixel using :

$$SAM = \arccos\left(\frac{\left\langle \hat{M}, M \right\rangle}{\left\|\hat{M}\right\|_2 \|M\|_2}\right) \tag{4.35}$$

Where, $\hat{M}$ represents the fused MS images and M represents the reference high resolution MS images.

4.4.1.2. *ERGAS*

ERGAS represents " Erreur Relative Globable Adimensionnelle de Synthèse" or " Global Relative Error in Synthesis". Presented in [WAL 02], ERGAS calculation is based upon RMSE "Root Mean Square Error". To determine the quality of fused images by using RMSE it was proposed that the RMSE value should be calculated for each pixel and each band separately :

$$RMSE(Band) = \frac{1}{NP} \sqrt{\sum^{Forallpixels} (M(Band) - \hat{M}(Band))^2} \qquad (4.36)$$

Where NP=total number of pixels in one MS band. Since RMSE is a number it is difficult to determine the quality of fused image without making a comparison with a reference. Also, RMSE is only sensitive to changes in radiance values. Hence, a modification to RMSE named RASE was proposed in [RAN 99]. The RASE "Relative Average Spectral Error" proposed can be calculated using :

$$RASE = \frac{100}{\mu(Band)} \sqrt{\frac{1}{N} \sum_{Band=1}^{N} RMSE(Band)^2} \qquad (4.37)$$

Where μ is the mean value of the reference MS image. Although, RASE satisfies the Synthesis Property proposed by Wald, however, problems occurred for images in which the difference in gray levels is quite large. Hence, a further modification was proposed. This modified quality assessment method is called ERGAS and it can be represented as :

$$ERGAS = 100 \frac{h}{l} \sqrt{\frac{1}{N} \sum_{Band=1}^{N} \frac{[RMSE(Band)]^2}{[\mu(Band)]^2}} \qquad (4.38)$$

Where the ratio between h and l represents the difference between the high and low resolution scales. It should also be noticed that both RASE and ERGAS are distortion indexes and hence the smaller the two values the better is the quality of the fused image. Hence, ERGAS=0 represents perfect fusion. The drawback of ERGAS is that it does not give an absolute index for fused images. This means that there is no absolute scale to determine the quality of fused images. Hence, it is similar to SAM i.e. the ideal value of the index is zero but the maximum value possible is not defined.

4.4.1.3. *Q4*

In 2004, Alparone et al. [ALP 04] proposed a global quality measurement index, Q4. Although the index is calculated on block basis but the final result is an average of all the blocks. Hence, the end result of the index is a global value. The Q4 index is now generally accepted to be one of the most reliable fusion quality assessment methods. The index is derived from the quality index Q proposed by Wang and Bovik in [WAN 02]. The Q index can be calculated using :

$$Q = 4 * \frac{\sigma(\hat{M}, M)}{\sigma_{\hat{M}}^2 + \sigma_M^2} * \frac{\overline{\hat{M}}\,\overline{M}}{\overline{\hat{M}}^2 + \overline{M}^2} \tag{4.39}$$

The modification proposed by Alparone et al. transforms the index as follows :

$$Q4 = 4 * \frac{\sigma(\hat{Z}, Z)}{\sigma_{\hat{Z}}\sigma_Z} * \frac{\sigma_{\hat{Z}}\sigma_Z}{\sigma_{\hat{Z}}^2 + \sigma_Z^2} * \frac{\overline{\hat{Z}}\,\overline{Z}}{\overline{\hat{Z}}^2 + \overline{Z}^2} \tag{4.40}$$

Where Z=ai+bj+ck+d, represents a quaternion and a=Blue, b=Green, c=Red and d=NIR MS bands. The first part of the index calculates the correlation between the fused MS image and the reference MS image. The second part of the index determines the change in contrast between the fused and reference MS images. Finally, the right most part of the index determines changes in the mean between the fused and reference MS images. Hence the index presented as Equ. 4.40 can detect changes in correlation, mean and contrast of the fused images with respect to the reference MS image. Another important aspect of Q4 is that it calculates these indexes between each of the MS bands, thus satisfying the third property, along with the second, proposed by Wald. The advantage of Q4 is that the resultant index is always within a fixed range of zero to one, with one being the ideal value (best result).

4.4.2. *Methods Not-Requiring a Reference Image*

Unlike methods which require a reference MS image methods not-requiring a reference image are non-conventional and are in the process of evolution. Since, most of these methods are relatively new they are not frequently used by the Data Fusion community. However, the advantage of these methods is that they actually determine the quality of the fused image at the desired high resolution, removing the inconvenience of working at reduced resolution.

4.4.2.1. *Zhous Protocol*

Proposed in 1998 [ZHO 98], the index is divided into two parts. One measures spectral distortion and the other spatial distortion. Spectral distortion is calculated as the sum of absolute difference between each pixel of the fused MS image and the upscaled low resolution MS image. This is represented mathematically as :

$$ZhouSpectralDistortion = \frac{1}{N} \sum_{Band=1}^{N} \left| \tilde{M}(Band) - \hat{M}(Band) \right| \qquad (4.41)$$

The ideal value of Zhous Spectral Distortion is zero. This means that the fused image is identical to the upscaled low resolution MS image. The fact that spectral distortion is only zero or minimal when the fused MS image is equal to the upscaled low resolution MS image, i.e. has no details, suggests that the index may be inconsistent. The spatial distortion is calculated between the details extracted from the fused MS image and the high resolution Pan image. The details are extracted using the Laplacian high pass filter given below. The spatial distortion is based upon the correlation, calculated between the details extracted from both the high resolution Pan and the fused MS images. A correlation value of one indicates ideal results and hence minimum spatial distortion.

$$\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix} \qquad (4.42)$$

The problem with calculating the spatial distortion using Zhous index is that the correlation between the details of MS and Pan images is ideally not one. This is because of the fact that the Pan and MS images are captured using different sensors, having different modulation transfer functions. Hence, the assumption that the correlation should be one is incorrect, as mentioned in [THO 07].

Although Zhou's protocol provides two separate indexes, one each for spectral and spatial distortion, the fact that the indexes are not on the same scale causes the problem of combining the two indexes together to obtain a single quality index. Also, the consistency or correctness of the index is not satisfactory.

4.4.2.2. *QNR*

Quality Without Reference, QNR, recently proposed by Alparone et al. [ALP 08] is a new method to calculate the quality of the fused images without requiring a reference high resolution MS image. The index like Zhou's quality index can provide

spectral and spatial distortion separately and can even provide a single combined quality index for both the distortions. The QNR index calculates the Q index, at both low (starting) and high (fused) resolution, between each of the MS bands among themselves i.e. red-green, red-blue, red-NIR, green-blue, green-NIR and blue-NIR (inter band calculations) and also between each MS band and Pan image (intra band calculations). The interband calculations at the two scales help determine if there is a difference between the spectral content among MS bands across scale, indicating spectral distortion. The intraband calculations at the two scales help determine the difference in the Q index between MS bands and Pan image across scale, determining if there is spatial distortion. Hence, the principle idea of QNR index revolves around the concept that the Q index calculated across scale should remain constant. The QNR index can be defined mathematically as :

$$QNR = (1 - D_\lambda)^\alpha (1 - D_s)^\beta \qquad (4.43)$$

Where D_λ represents spectral distortion and D_s represents spatial distortion. Further the spectral and spatial distortions can be represented mathematically as :

$$D_\lambda = \sqrt{\frac{1}{N(N-1)} \sum_{l=1}^{N} \sum_{r=1, r \neq l}^{N} \left| \underbrace{Q(\tilde{M}_l, \tilde{M}_r) - Q(\hat{M}_l, \hat{M}_r)} \right|} \qquad (4.44)$$

$$D_s = \sqrt{\frac{1}{N} \sum_{l=1}^{N} \left| \underbrace{Q(\tilde{M}_l, \dot{P}) - Q(\hat{M}_l, P)} \right|} \qquad (4.45)$$

The ideal values of both spectral D_λ and spatial D_s distortion is zero. Thus, the ideal value of the QNR index is 1. The terms α and β can be used to give more significance to either the spectral or spatial quality index. It should be noted that Equ. 4.44 and Equ. 4.45 represent distortions, however $(1 - D_\lambda)$ and $(1 - D_s)$ represent spectral and spatial quality.

QNR calculates the Q index at different block sizes however the final Q index is a global index as all of the block results are added before calculating the difference. A slight modification to the QNR index was proposed in [KHA 08a] to make the QNR index work on local blocks, where the size of the block used for QNR calculations was set equal to the size used for calculation of the Q index. In this way the QNR index can be calculated separately for each block, thus providing a local QNR index, and if required the final global QNR can be calculated by summing the local QNR indexes of each block rather than summing the absolute difference of Q index for each block.

4.4.2.3. *Modified Zhou Wald Protocol*

Proposed recently, the method proposes a hybrid quality assessment approach based upon Wald's spectral consistency property and inspired by Zhou's spatial quality. In [KHA 08b] Khan et al. have used the same principle of using two separate quality indexes. The novelty of the method exists in using approximate modulation transfer function (MTF) filters to assess the fusion quality. However, the proposed index does not provide a single comprehensive quality index, like QNR.

The principle of the proposed quality assessment index revolves around separation of high frequency information from the fused images and Pan image at both the low and high resolution. The low frequency information is used to assess the spectral quality of fused MS image. Using Wald's consistency protocol, the fused low pass filtered image is decimated to obtain a MS image at the same resolution as the starting reference. A simple Q4 index based quality assessment can be performed between this reduced, MTF filtered, image and the reference low resolution MS image. Hence, the MTF filters are used to obtain the low pass filtered image and their complements provide the details for assessing the spatial quality. To assess spectral quality making use of Wald's consistency property the following equation can be used.

$$D_\lambda = 1 - Q4(\hat{M} \downarrow, \dot{M}) \tag{4.46}$$

Where, $\hat{M}$ represents the fused image, $\hat{M} \downarrow$ represents the MTF filtered and decimated fused MS image and $\dot{M}$ represents the low resolution reference MS image.

The high frequency details extracted from the fused MS images are used for spatial quality assessment. A box type filter is used to extract the high frequency details from the high resolution Pan image. The MS image details are obtained by using their respective MTF filters. Next the same procedure is repeated on the low resolution reference MS images and the low resolution Pan image. The low resolution Pan image is obtained by filtering the high resolution Pan image by an approximately ideal low pass filter (since a Pan image at low resolution would have been inverse filtered after passing through its respective MTF filter). In this way the Q index is calculated between the details of MS images and the Pan image at both low resolution and high resolution. The absolute difference between the Q index obtained before and after fusion provides a spatial quality index. Hence unlike Zhou's protocol it is not assumed that the correlation coefficient between the details of the Pan and MS image is one. Instead, a relationship is calculated between the details of the MS and Pan images at low resolution and it is assumed that this relationship remains constant across scale since the MTF of the sensors is scale invariant.

$$D_s = \frac{1}{N} * \sum_{Band=1}^{N} \left| Q(\ddot{\hat{M}}(Band), \ddot{P}) - Q(\ddot{\dot{M}}(Band), \ddot{P}) \right| \tag{4.47}$$

4.4.3. *Methods Not to be used*

Fusion quality assessment is not a trivial task. Attention should be paid to the methods used for assessing fusion quality. Simple indexes like mean, variance, correlation, entropy etc. alone, do not serve the purpose of quality assessment. It has been shown in [THO 06] that an image and a noise sequence may have the same mean, variance and entropy. Hence, these indexes alone are non-sufficient to determine the quality of fused images. Also, it has been shown that two identical images having different mean could return a correlation coefficient of 1. Another important point to note is that RMSE type calculations should never be performed between the fused MS images and the Pan image to assess fusion quality. Some authors have used this type of index to determine the spatial quality of fused images. This index would lead to incorrect conclusions since in RMSE, difference is calculated at pixel level between the MS and Pan images. This difference should never be zero since the MS and Pan images are obtained using different sensors and have different spectral ranges and hence different photon count and radiance measurements. Hence this difference does not provide any meaningful information and would thus lead to inconclusive results, since the ideal RMSE vale of zero would be obtained if all MS bands are equal to the Pan image.

4.5. Conclusion

Pansharpening produces MS images at higher spatial quality while preventing spectral distortion. In the context of pansharpening this chapter presented a brief overview of different fusion indexes and the fusion quality assessment methods. Instead of testing different algorithms and presenting their results, it is better to present the results of the recently held IEEE GRS-S Data Fusion Contest to give an idea of the fusion results obtained by different pansharpening algorithms and their respective scores using different quality indexes. The visual results of both CS and MRA based fusion methods are presented in Fig. 4.8. The image selected is a portion of the Pléiades Toulouse image provided by CNES. A visual comparison of the fused images with the reference suggests that the image obtained by Context Based Decision Fusion is the best fused image. It also demonstrates the ability of pansharpened images to be similar to the image that a satellite would have captured if it had the desired higher resolution. Thus, demonstrating the usefulness of pansharpening. The quantitative comparison between the fusion methods is presented in Tab. 4.1. All four indexes suggest that CBD has the best quantitative results. Thus, it can be concluded that with reference to the current literature the fusion methods making using of Multiresolution based analysis and which use a detail injection model are considered to be the best fusion methods.

The future direction of work seems to be in the direction of improving CS methods by developing better detail injection models for them. Since, these methods are

Index	AWLP	GLP-CBD	GIHS-GA	GIHS-TP
SAM	4.56	3.67	4.08	5.30
ERGAS	3.12	2.78	3.63	5.34
Q4	0.96	0.96	0.93	0.84

Tableau 4.1. *Quantitative comparison of Pansharpening methods. ©IEEE 2007*

computationally non-intensive and result in sharper fused images a lot of work is being done to improve their spectral quality. Also, a trend toward further extension of context based fusion is being sought. The work entails using different fusion models for different objects in an image thus providing sharper results for objects in urban scene while slightly less sharper objects in images containing country side. Another aspect of fusion, not considered in this chapter, are the result of different fusion algorithms for different sensors and different types of scenes. This is an interesting topic and demands a complete chapter on its own, however, it should be made clear here that each fusion algorithm performs differently for different sensor as the signal to noise ratio for each sensor changes. Hence, an area of interest regarding pansharpening would be a detail extraction and injection model which would be sensor adaptive.

[AIA 02] AIAZZI B., ALPARONE L., BARONTI S., GARZELLI A., « Context-Driven Fusion of High Spatial and Spectral Resolution Images Based on Oversampled Multiresolution Analysis », *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, n°10, p. 2300–2312, 2002.

[AIA 07] AIAZZI B., BARONTI S., SELVA M., « Improving Component Substitution Pansharpening Through Multivariate Regression of MS+Pan Data », *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, n°10, p. 3230-3239, October 2007.

[ALP 04] ALPARONE L., BARONTI S., GARZELLI A., NENCINI F., « A global quality measurement of pan-sharpened multispectral imagery », *IEEE Geoscience and Remote Sensing Letters*, vol. 1, n°4, p. 313–317, 2004.

[ALP 07] ALPARONE L., WALD L., CHANUSSOT J., THOMAS C., GAMBA P., BRUCE L. M., « Comparison of Pansharpening Algorithms : Outcome of the 2006 GRS-S Data Fusion Contest », *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, p. 3012-3021, 2007.

[ALP 08] ALPARONE L., AIAZZI B., BARONTI S., GARZELLI A., NENCINI F., SELVA M., « Multispectral and Panchromatic Data Fusion Assessment Without Reference », *Photogrammetric Engineering and Remote Sensing*, vol. 74, p. 193-200, 2008.

[CHA 91] CHAVEZ P. S., STUART J., SIDES C., ANDERSON J. A., « Comparison of three different methods to merge multiresolution and multispectral data : Landsat TM and SPOT panchromatic », *Photogrammetric Engineering and Remote Sensing*, vol. 57, p. 259–303, 1991.

[CHI 03] Chibani Y., « Multisource Image Fusion using Redundant Wavelet Decomposition », IEEE Geoscience and Remote Sensing Symposium, 2003.

[CHO 06] Choi M., « A new intensity-hue-saturation fusion approach to image fusion with a tradeoff parameter », *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, n°6, p. 1672-1682, June 2006.

[COL 06] Colditz R., Wehrmann T., M. Bachmann K. S., Schmidt M., Strunz G., Dech S., « Influence of Image Fusion Approaches on Classification Accuracy », *International Journal of Remote Sensing*, vol. 27, n°15, p. 3311 - 3335, August 2006.

[GAR 05] Garzelli A., Nencini F., Alparone L., Baronti S., « Multiresolution Fusion of Multispectral and Panchromatic Images through the Curvelet Transform », IEEE Geoscience and Remote Sensing Symposium, 2005.

[GAR 06] Garzelli A., Nencini F., « PAN-sharpening of very high resolution multispectral images using genetic algorithm », *International Journal of Remote Sensing*, vol. 27, p. 3273-3292, 2006.

[GAR 07] Garzelli A. ., Nencini F., « A Genetic Approach to Pan-sharpening of Multispectral Images », Bochenek Z., Ed., *New Developments and Challenges in Remote Sensing*, Proceedings of the 26th Symposium of The European Association of Remote Sensing Laboratories, p. 91 - 100, May 2007.

[GIL 87] Gillespie A. R., Kahle A. B., Walker R. E., « Color Enhancement of Highly Correlated Images-II. Channel Ratio and "Chromaticity" Transform Techniques », *Remote Sensing of Environment*, vol. 22, n°3, p. 343-365, 1987.

[GON 92] Gonzalez R. C., Woods R. E., *Digital Image Processing*, Addison Wessley Publishing Company, 1992.

[GON 04] González-Audícana M., Saleta J. L., Catalán R. G., García R., « Fusion of Multispectral and Panchromatic Images Using Iproved IHS and PCA Mergers Based on Wavelet Decomposition », *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, p. 1291-1299, 2004.

[KHA 08a] Khan M. M., Alparone L., Chanussot J., « Pansharpening based on QNR Optimization », *Proceedings of IEEE Geoscience and Remote Sensing Symposium*, IEEE Seoscience and Remote Sensing Symposium, July 2008.

[KHA 08b] Khan M. M., Alparone L., Chanussot J., « Pansharpening Quality Assessment using Modulation Transfer Function Filters », *Proceedings of IEEE Geoscience and Remote Sensing Symposium*, IEEE Geoscience and Remote Sensing Symposium, July 2008.

[KHA 08c] Khan M. M., Chanussot J., Condat L., Montanvert A., « Indusion : Fusion of Multispectral and Panchromatic Images using Induction Scaling Technique », *IEEE Geoscience and Remote Sensing Letters*, vol. 5, n°1, p. 98-102, 2008.

[LIU 00] Liu J. G., « Smoothing Filter Based Intensity Modulation : A spectral preserve image fusion technique for improving spatial details », *International Journal of Remote Sensing*, vol. 21, n°18, p. 3461–3472, 2000.

[LLO 97] Lloyd N. T., Bau D., *Numerical Linear Algebra*, SIAM : Society for Industiral and Applied Mathematics, 1997.

[N´ 99] NÚNEZ J., OTAZU X., FORS O., PRADES A., PALÀ V., ARBIOL R., « Multiresolution-based image fusion with additive wavelet decomposition », *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, n°3, p. 1204–1211, 1999.

[OTA 05] OTAZU X., GONZÁLEZ-AUDÍCANA M., FORS O., NÚNEZ J., « Induction of Sensor Spectral Response Into Image Fusion Methods. Application to Wavelet-Based Methods », *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, p. 2376-2385, 2005.

[POH 98] POHL C., GENDEREN J. L. V., « Multisensor image fusion in remote sensing : concepts, methods and applications », *International Journal of Remote Sensing*, vol. 19, n°5, p. 823-854, 1998.

[PRA 06] PRADHAN P. S., KING R. L., YOUNAN N. H., HOLCOMB D. W., « Estimation of the Number of Decomposition Levels for a Wavelet-Based Multiresolution Multisensor Image Fusion », *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, n°12, p. 3674-3686, December 2006.

[RAN 99] RANCHIN T., WALD L., « Fusion of high spatial and spectral resolution images : The ARSIS concept and its implementation », *Photogrammetric Engineering and Remote Sensing*, vol. 66, n°1, p. 49–61, 1999.

[RAN 03] RANCHIN T., AIAZZI B., ALPARONE L., BARONTI S., WALD L., « Image fusion - the ARSIS concept and some successful implmentation schemes », *ISPRS Journal of Photogrammtry and Remote Sensing*, vol. 58, p. 4-18, 2003.

[SIR 08] SIRGUEY P., MATHIEU R., ARNAUD Y., KHAN M. M., CHANUSSOT J., « Improving MODIS Spatial Resolution for Snow Mapping Using Wavelet Fusion and ARSIS Concept », *IEEE Geoscience and Remote Sensing Letters*, vol. 5, p. 78-82, 2008.

[THO 06] THOMAS C., Fusion d'images de résolutions spatiales différentes, PhD thesis, École des Mines de Paris, Sophia Antipolis, 2006.

[THO 07] THOMAS C., WALD L., « Comparing Distances for Quality Assessment of Fused Images », *New Developments and Challenges in Remote Sensing*, Proceedings of the 26th Symposium of the European Association of Remote Sensing Laboratories, p. 101-111, 2007.

[TU 04] TU T. M., HUANG P. S., HUNG C. L., CHANG C. P., « A fast intensity-hue-saturation fusion technique with spectral adjustment for IKONOS imagery », *IEEE Geoscience and Remote Sensing Letters*, vol. 1, n°4, p. 309–312, 2004.

[TU 07] TU T. M., CHENG W. C., CHANG C. P., HUANG P. S., CHANG J. C., « Best Tradeoff for High-Resolution Image Fusion to Preserve Spatial Details and Minimize Color Distortion », *IEEE Geoscience and Remote Sensing Letters*, vol. 4, n°2, p. 302-306, 2007.

[WAL 02] WALD L., *Data Fusion : Definitions and Architectures — Fusion of images of different spatial resolutions*, ENSMP, 2002.

[WAN 02] WANG Z., BOVIK A. C., « A Universal Image Quality index », *IEEE Signal Processing Letters*, vol. 9, n°3, p. 81-84, March 2002.

[ZHO 98] ZHOU J., CIVCO D. L., SILANDER J. A., « A Wavelet Transform Method to Merge Landsat TM and SPOT Panchromatic Data », *International Journal of Remote Sensing*, vol. 19, n°4, p. 743-757, 1998.

```
./RegistrationFusion/khan/figs/DataContestPaper.eps
```

Figure 4.8. *The results of fusion are presented using different fusion algorithms. The Panchromatic and Original MS images are presented for reference. The fused images obtained by Additive Wavelet Luminance Pansharpening Method (AWLP), Context Based Decision (GLP-CBD) Pansharpening Method, Genetic Algorithm Based Pansharpening (GIHS-GA) and Intensity-Hue-Saturation Fusion using Choi's method (GIHS-TP) are presented. ©IEEE 2007*

Change detection

Chapitre 5

Change detection in remote sensing observation

The aim of this chapter is to draw an overview of change detection techniques in remote sensing earth observation. In remote sensing domain, a change may be considered to as an alteration of the surface components. But temporal analysis of remote sensing image is facing several difficulties, among them the large amount of data to be processed and also the very few number of temporal observations. Despite the lack of temporal model, the natural evolution of landscape and the evolution induced by the sensors, many valuable techniques exist that perform change detection from two or more images acquired from the same or from different sensors.

This overview has been drawn from the CACHANT initiative supported by the GdR ISIS in 2006 and the presentation that have been given during the two meetings[1].

5.1. Abrupt Change Detection due to Major Disasters

Remote sensing imagery is a precious tool for rapid mapping applications. In this context, one of the main uses of remote sensing is the detection of changes occurring after a natural or anthropic disaster. Since they are abrupt and seldom predictable, these events cannot be well temporally sampled – in the Shannon sense – by the polar orbit satellites which provide the medium, high and very high resolution imagery

1. CACHANT : *Cartographie Automatique des ChAngements Naturels par Télédetection* http ://gdr-isis.org/cachant.

needed for an accurate analysis of the land cover. Therefore, rapid mapping is often produced by detecting the changes between an acquisition after the event and available archive data.

5.1.1. *Constraints*

This change detection procedure is made difficult due to the time constraints imposed by the emergency context. During a natural or industrial hazard, it is usually difficult to provide hazard assessment using satellite imagery in a relatively short delay. Operational problems can affect the task of hazard assessment, like the absence of a dedicated satellite constellation for this purpose as well as the delay between the satellite programming and the actual image acquisition. Indeed, the first available acquisition after the event has to be used whatever its modality which is more likely to be a radar image due to weather and daylight constraints.

The kind of changes produced by the event of interest are often difficult to model. The same kind of event – *e.g.* a flood – can have different signatures depending on where it happens – *e.g.* high density built-up areas, agricultural areas... – and on the characteristics of the sensor (spatial resolution, frequency, polarimetry, angle of sight and ascending/descending orbits...). Also, the changes of interest are all mixed up with *normal changes*, which can be the majority if the time gap between the two acquisitions is too long. A very difficult problem arises from these issues : detecting abrupt unmodeled transitions in a temporal series with only two dates[2].

From this position of the problem, one can make the straightforward deduction that pixel-wise comparison between the two images will not be robust enough. The first problem that arises is the automatic fine registration before performing the change detection itself. This problem is even more difficult when images are acquired with different sensors. By different sensors, it is understood sensors that produce images with different radiometric properties, *i.e.* sensors which measure different physical magnitudes : optical sensors operating in different spectral bands, radar and optical sensors... For this kind of image pairs, the classical approach of fine correlation [KUG 75, DVO 83] cannot always be used to provide the required accuracy. Indeed, the correlation coefficient can only measure similarities up to an affine transformation.

There are two main questions :

1) What does the similarity mean between, *e.g.* a radar and an optical image ?

---

2. In the case where a sequence of several images is to be processed, the approaches presented in [BRU 04, LOM 02] may be applied.

2) What does fine registration mean in the case where the geometric distortions are so big and the source of information can be located in different places (*e.g.* the same kind of edge can be produced by the edge of the roof of a building in an optical image and by the wall-ground bounce in a radar image) ?

An answer may be drawn by saying that the images of the same object obtained by different sensors are two different representations of the same reality. For the same spatial location, we have two different measures. Both items of information come from the same source, and thus, they have a lot of common information. This relationship may not be perfect, but it can be evaluated in a relative way : different geometrical distortions are compared, and the one leading to the strongest link between the two measures is kept. In [ING 04], the mutual information is used to perform such a multi-sensor-based image registration. It is worth noting that the *similarity measures* that are used for image registration purposes are not too far from the *similarity measures* used for change detection, mostly for Synthetic Aperture Radar (SAR) images of medium resolution.

Figures 5.1 and 5.2 show a typical examples of images to be processed from a real case [ING 03b], for which the International Charter "Space and Major Disaster" has been engaged[3]. Due to the large amount of data to be processed and the delay demanded to deliver the impacted zone characterization, automatic or unsupervized approaches are required. Unfortunately, due to the lack of *a priori* knowledge in the kind of change and the high variability in the sensors, an expert has to directed the processing chain by using supervized or semi-supervized techniques.

A pair of Radarsat images, acquired before and after the eruption of the Nyira-gongo volcano (D.R. of Congo) which occurred in January 2002, were used. Fig. 5.1 shows the two images to be compared and a change map produced using ground mea-sure. The images have a ground resolution of 10 m and cover an area of $4km \times 8km$. The images were ortho-rectified by IGN-F, the French National Geographic Institute, to a UTM35S projection, which was the same as the one used for the reference map. No filtering or calibration was applied to the data. The 16 to 8 bit conversion was performed using a 3σ thresholding followed by a linear intensity rescaling. It is worth noting that the image resampling applied in the ortho-registration step modifies the local statistics of the image. Indeed, the image resampling implies local image inter-polation which is based on approximation. A bicubic interpolation was used in this case. This type of filter has a smoothing effect which depends on the local shift [**?**]. Because of these radiometric artifacts introduced during the geometric pre-processing, the theoretical models for SAR statistics may not hold locally. The area at the bottom right hand corner of the ground truth mask corresponds to an area where a severe mis-registration exists due to the lack of a proper digital terrain model. Finally, one has

----

3. http ://www.disasterscharter.org/

(a) Before image (Radarsat F5)   (b) After image (Radarsat F2)   (c) Mask

Figure 5.1. *Data and ground truth for the Nyiragongo volcanic eruption of January 2002.*

to take into account the fact that the ground truth is not perfect and that all results presented in this section should be analyzed rather in a relative manner – one detector with respect to another – rather than in an absolute one – absolute value of detection probabilities.

5.1.2. *Usual strategies for measuring changes*

The scientific community provides a wide variety of methods for change detection using remote sensing imagery. From the study of the different algorithms it was noted that most of these methods are structured in two major steps :

– **Feature extraction** Feature extraction consists in the processing of the available imagery in order to extract useful information and to emphasize the input required by the decision step. These features can be intrinsic features extracted from the images or change indicators like the difference or ratio between the input images.

– **Decision function** The decision function consists of the operations applied on the extracted features in order to produce the decision (*i.e.* change *vs.* no change). Finally based on this decision, the output change detection map is obtained.

(a) Before SPOT image   (b) After SPOT image

Figure 5.2. *SPOT data used to detect impacted zone after the food in Gloucester, UK, in 2000.*

This is the most generic scheme for change detection, but not all methods follow it [Dre 93]. For instance methods can directly use the available imagery in the decision function without the need for feature extraction.

In the literature, change detection problems are generally addressed for specific types of sensors (*i.e.* passive or active). The remaining of this section follows this classification.

5.1.2.1. *Measuring changes in-between optical images*

The most commonly used input for the decision function in change detection methods using passive sensors is the difference image. The difference image is interesting since the noise level in this type of imagery is additive and hence taking the difference between image at time t_1 (denoted to as I_1) and the image at time t_2 (I_2) gives a valid change measure. This kind of *difference* measure may be applied directly to the pixel value of the images [COP 04, LU 04] or on a transformed representation (*e.g.* the texture). In fact, textural measures may be calculated separately on each image in order to evaluate their difference to measure a change between the two acquisitions [LI 01].

Other derivations from the difference image can be also used, for example methods based on the Change Vector Analysis (CVA) as in [YOO 03, AVR 02] as well as methods using different versions of the difference image as in [BEA 04] where the authors use low pass versions of the difference image or the difference between local

means in order to reduce noise. In [BOV 07], an extent of the CVA technique is presented which takes in consideration not only the magnitude of the change vector but also its phase when the change vector is represented in the polar domain.

The Multivariate Alteration Detection (MAD) is proposed in [NIE 98] as a multivariate change detection technique (which produces a change measure using multiband images at the input). It may be used also when the number of bands in the two input images are not equal or even when the bands proceed from different sensors. The MAD method is based on a linear transformation of bands of I_1 and I_2 (namely $\boldsymbol{a}I_1 - \boldsymbol{b}I_2$) such as the variance is minimized (given the variances of $\boldsymbol{a}I_1$ and $\boldsymbol{b}I_2$). MAD is insensitive to differences in sensor gain settings, or to linear radiometric and atmospheric correction schemes. MAD, followed by Maximum Autocorrelation Factor (MAF) transformation [NIE 98], gave significantly better results than PCA of the *difference* measure in the detection of coherent patterns of spatial change in urban zones in Australia. Moreover, the MAD transformation provides a way to combine different data types that may be useful in historical change detection studies. [ZHA 04] *et* al. propose to improve the estimation of data means and covariance matrix by using two methods : (i) the exclusion of outliers from the computation and (ii) the robust estimation of parameters not to be affected by the outliers. Promising results have been found, however, not quantitative validation information has been given.

Also of great interest, PCA-based techniques have been reported in many applications. Considering now the simple idea of detecting changes between two bands of I_1 and I_2 using the PCA transform. The basic hypothesis here is that change pixels between the two dates are fewer or much fewer than pixels that represent no change. Therefore, the scatter plot between pixels of the same spectral bands but from two different dates, is similar in shape to a 2D correlated distribution. Then, points along the first Principal Component (PC1) may be considered to as pixels that have not been changed between the two dates, because they are highly correlated. On the other hand, points that have large values of PC2 may be considered to as change pixels. Normally, one may use as many spectral bands as necessary instead of only one to improve the separation of change and no-change areas [FUN 87, BYR 98, FUN 90].

5.1.2.2. *Measuring changes in-between radar images*

The main difficulty in the multitemporal analysis of SAR images is the presence of speckle noise. When moving away from interferometric configurations, the speckle is different from one image to the other and it can induce a high number of false alarms in the change detection procedure. Because of the multiplicative nature of speckle, the classical approach in SAR remote sensing involves using the ratio of the local means in the neighborhood of each pair of co-located pixels [RIG 93]. The Mean Ratio Detector, MRD, is usually implemented as the following normalized quantity :

$$r_{MRD} = 1 - \min\left\{\frac{\mu_X}{\mu_Y}, \frac{\mu_Y}{\mu_X}\right\}, \tag{5.1}$$

where μ_X and μ_Y stand for the local mean values of the images before and after the event of interest. The logarithm of eq. (5.1) may also be used. Nevertheless, this operation does not modify the performance of the detector in terms of ROC[4] even if the contrast of the image of change indicator is modified. However, the logarithm is used since it modifies the initial pdf of the image of change indicator and then facilitates the development of Bayesian thresholding approaches [BAZ 05].

This detector is robust to speckle noise but limited to the comparison of first order statistics. The classical model for SAR intensity introduced by Ulaby *et* al. [ULA 86] assumes that the texture is a zero-mean multiplicative contribution. Therefore, changes taking place at the texture level which preserve the mean value will not be detected by the mean ratio detector. This remark invites a more accurate analysis of the local statistics of the images to be compared. Bujor *et* al. [BUJ 04] did an interesting work by analyzing the interest of higher order statistics for change detection in SAR images. They concluded that the ratio of means was useful for step changes and that the second and third order log-cumulants were useful for progressive changes appearing in consecutive images in multi-temporal series. Since higher order statistics seem to be helpful, one may want to compare the local probability density functions (pdfs) of the neighborhood of the homologous pixels of the pair of images used for the change detection.

Then a measure derived from Information Theory called the Kullback-Leibler divergence [KUL 51] may be used for change detection purpose [ING 07a]. Let P_X and P_Y be two probability laws of the random variables X and Y. The Kullback-Leibler divergence from Y to X, in the case where these two laws have the densities f_X and f_Y, is given by :

$$K(Y|X) = \int \log \frac{f_X(x)}{f_Y(x)} \, f_X(x) \, dx. \tag{5.2}$$

The measure $\log \frac{f_X(x)}{f_Y(x)}$ can be thought of as the information on x for discrimination between hypothesis $\mathcal{H}_X$ and $\mathcal{H}_Y$ if hypothesis $\mathcal{H}_X$ is associated with pdf $f_X(x)$, and $\mathcal{H}_Y$ with $f_Y(x)$. Therefore, the Kullback-Leibler divergence $K(Y|X)$ can be understood as the mean information for discriminating $\mathcal{H}_X$ and $\mathcal{H}_Y$ per observation. This divergence appears to be an appropriate tool when it is considered that changes on the ground induce different shapes of the local pdf. This measure is not symmetric as it stands ; *i.e.* $K(Y|X) \neq K(X|Y)$, but a symmetric version may be defined by :

$$D(X,Y) = D(Y,X) = K(Y|X) + K(X|Y), \tag{5.3}$$

called the Kullback-Leibler *distance* (KLD).

In the absence of texture, it is known that the radar intensity follows a Gamma distribution. In the presence of texture, the local statistics can deviate from the Gamma

----

4. ROC : Receiver Operating Characteristics

distribution. For instance, if the texture is modeled by a Gamma distribution with a shape parameter, the resulting intensity distribution follows a K-law [WAR 81]. More generally, it is now accepted that the statistics of SAR images can be well modeled by the family of probability distributions known as the Pearson system [JOH 94, DEL 97]. It is made up of eight types of distributions among which the Gaussian and the Gamma distributions may be found. The Pearson-based Kullback-Leibler Detector (PKLD) was originally introduced in [ING 03a]. It does not have a unique analytic expression, since 8 different types of distribution may be hold. Therefore, 64 different possibilities for the couples of pdf exist. Once the couple of pdfs is identified, the detection can be performed by numerical integration.

Instead of considering a parameterization of a given density, or set of densities, it may be of interest to *describe* the shape of the distribution. Such a description is based on quantitative terms that may approximate the pdf itself. The cumulants themselves do not provide such a pdf estimation directly but are necessary to *describe* its shape : for example, third order (κ_3) is linked to the symmetry (*i.e.* skewness), while the fourth (κ_4) is linked to the flatness (*i.e.* kurtosis). The density is then estimated through a series expansion. In fact, the cumulant generating function is used for such an estimation.

Let us assume that the density to be approximated is not *too far* from a Gaussian pdf (denoted as $\mathcal{G}_X$ to underline the fact that it has the same mean and variance as X), that is, with a shape similar to the Gaussian distribution. The density may be expressed by a formal Taylor-like series :

$$f_X(x) = \mathcal{G}_X(x) + c_1 \frac{d\mathcal{G}_X(x)}{dx} + c_2 \frac{d^2\mathcal{G}_X(x)}{dx^2} + \cdots = \sum_{r=0}^{\infty} c_r H_r(x) \mathcal{G}_X(x),$$

with $H_r(x)$ known as the Chebyshev-Hermite polynomial of order r [STU 91]. When choosing a Gaussian law so that its first and second cumulants agree with those of X, the number of terms of the series expansion is greatly reduced. This is the so-called Edgeworth series expansion. Its expression, when truncated to order 6, is the following :

$$f_X(x) = \Big(1 + \frac{\kappa_{X';3}}{6} H_3(x) + \frac{\kappa_{X';4}}{24} H_4(x) + \frac{\kappa_{X';5}}{120} H_5(x)$$
$$+ \frac{\kappa_{X';6} + 10\kappa_{X';3}^2}{720} H_6(x) \Big) \mathcal{G}_X(x). \tag{5.4}$$

It can be thought of as a model of the form $X = X_{\mathcal{G}} + X'$ where $X_{\mathcal{G}}$ is a random variable with Gaussian density with same mean and variance as X, and X', a standardized version of X [MCC 87] with :

$$X' = (X - \kappa_{X;1}) \kappa_{X;2}^{-1/2}.$$

The Edgeworth series expansion of the two pdfs f_X and f_Y may be introduced into the Kullback-Leibler divergence (eq. (5.2)). It yields an approximation of the Kullback-Leibler divergence by Edgeworth series, truncated at a given order. In [LIN 99], such an approximation has been truncated to order 4 by using the equality $\frac{f_X}{f_Y} = \frac{f_X}{\mathcal{G}_X} \frac{\mathcal{G}_X}{\mathcal{G}_Y} \frac{\mathcal{G}_Y}{f_Y}$, where $\mathcal{G}_X$ (resp. $\mathcal{G}_Y$) is a Gaussian density of same mean and variance as f_X (resp. f_Y). Then, the cumulant-based Kullback-Leibler detector (CKLD) between two observations X and Y is obtain [ING 07a] :

$$r_{CKLD} = KL_{\text{Edgeworth}}(X, Y) + KL_{\text{Edgeworth}}(Y, X). \tag{5.5}$$

The reader should note that, like for the Pearson-based detector, despite the apparent complexity of the formulas, only moments up to order 4 have to be computed.

Figure 5.3 shows the change detections, performed by those algorithms, on the Goma case (from Fig. 5.1). The comparisons between the result coming from the

(a) Intensity ratio   (b) Pearson Kulback-Leibler   (c) Cumulant-based Kullback-Leibler
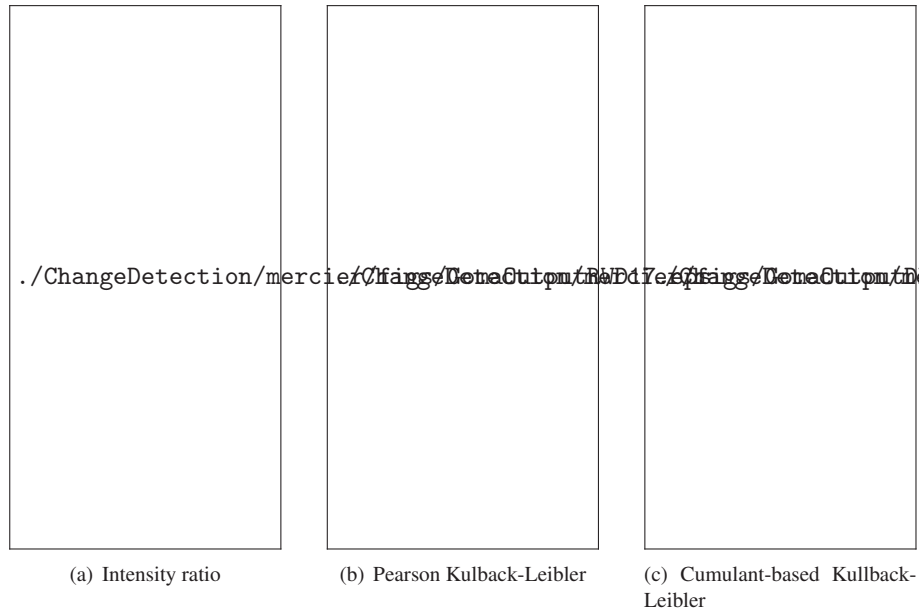
Figure 5.3. *Change detection : comparison between the different change indicators using the same window size (35×35 pixels).*

MRD and the CKLD are shown in fig. 5.3. Fig. 5.4 gives the ROC plots using the

ground truth of fig. 5.1(c)[5]. It shows that the use of KL approximation by the Edgeworth series outperforms any other methods such as model-based (Gaussian-based or Pearson-based) KL distance, or a miss-detection behavior of this detector can be observed because it uses the mean pixel values only. It is interesting to underline the fact that the ratio criterion is not always worse than pdf-based criteria. In fact, a density model has to fit the data in order to yield pertinent results.

./ChangeDetection/mercier/figs/rocw.eps

Figure 5.4. *ROC plots for the different detectors. The cumulant-based Kullback-Leibler detector (CKLD) outperforms all other detectors. The Pearson-based detector (PKLD) gives results identical to the classical mean ratio (MRD). The Gaussian-based detector shows the worst behavior.*

5.1.2.3. *Taking the decision to yield a change map*

After extracting an adequate change measure, the key of a change detection algorithm lies in the decision function. This is the function that "decides" whether a given unit (*i.e.* pixel, region, cluster, object, . . .) belongs to the change or no-change class.

The most intuitive approach to take a decision is to threshold the feature images, as performed in [LI 01, ITA 05]. This blind decision, *i.e.* decision taken pixel by pixel by comparing the feature with the threshold, often suffers from mis- or over-detection. One way to deal with this problem is to take into account the neighborhood of each

---

5. In order to the ROC plot, the evolution of the detection probability, P_{det} as a function of the false alarm probability, P_{fa}, has be evaluated with a set of constant thresholds applied to the whole detection image.

pixel for decision making about that pixel. A common approach is to use Markov Random Fields (MRF) model, this type of analysis is proposed in [LIU 05, BRU 00, BRU 02, WIE 97, BRU 99, KAS 02, Ami 04]. The MRF information is then integrated in a larger statistical framework, and the decision is taken in favor of the class with the highest probability. The difference between these methods lie within the statistical model (the pdf of the classes, the Bayesian criterion such as MAP or MPM...) and the estimation of the model parameters to make decision unsupervised. A slight modification of this reasoning is presented in [HWA 04] where the authors propose the pdf estimation of the noise in multi-spectral imagery and use this estimation in a model aimed for identifying the behavior of unchanged pixels. For maximum of genericity, no *a priori* hypothesis should be taken into account. The most flexible algorithms then to be used are the learning based algorithms. These algorithms, which are supervized, are made of two phases : learning and decision making. For this category of methods, the user defines a set of areas that the algorithm uses to learn and then generalize on new unknown data. The two most commonly used algorithms are Neural Networks (NN) [XUE 04, D'A 04, VEL 02] and more recently Support Vector Machines (SVM) [GUO 06, ZHI 06, LI 04, MER 06].

Blind decision, *i.e.* decision taken pixel by pixel by comparing the feature with the threshold, often suffers from mis- or over-detection. One way to deal with this problem is to take into account the neighborhood of each pixel for decision making about that pixel. A common approach is to consider that the change map follows a Markov model, which has to be recovered from the noisy observation given by the feature image. This type of analysis is considered for example in [LIU 05, BRU 00, BRU 02, WIE 97, BRU 99, KAS 02, Ami 04, SER 06]. The difference between these methods lie within the kind of feature image used, the estimation of the model parameters to make decision unsupervised and the statistical model (the shape of class-conditional densities, the Bayesian criterion such as MAP or MPM...). A slight modification of this reasoning is presented in [HWA 04] where the authors propose the pdf estimation of the noise in multi-spectral imagery and use this estimation in a model aimed for identifying the behavior of unchanged pixels.

Most of the existing methods for automatic change detection employ crisp models and consequently neglect the fuzzy aspect of the scene behavior (change in illumination, sensor degradation, natural evolution of the scene). To cope with such situations, fuzzy set theory and classical change detection techniques have been combined to perform change detection [DEE 98, NEM 04], e.g. in spectral-spatial features maps [HIL 00], as well as in the post-classification comparison context [DEE 02]. An imprecision measure, characteristic of fuzzy approaches, can also be added to the uncertainty measure of statistical models, resulting in original models that are able to take into account both the uncertainty and imprecision of changes. This introduction has been achieved in different statistical contexts : blind and contextual [CAI 97], Hidden Markov Random Field (HMRF) [CAI 93, SAL 97, RUA 02] and in Hidden Markov Chain (HMC) [CAR 06, SAL 06]. In the latter case, the 2D image is first

converted to a 1D sequence by using the Hilbert-Peano scan [SKA 92]. The way used to introduce the fuzzy measure in these models was to consider that the Markovian process does not take its values in a discrete set {0=no change, 1=change} anymore, but in the continuous interval $[0, 1]$. Parameter estimation is made unsupervised by original variants of iterative methods such as Expectation-Maximisation (EM), Stochastic EM (SEM) or Iterated Conditional Estimation (ICE).

The fuzzy HMC model proposed in [CAR 06] is now illustrated with the two SAR images in fig. 5.5. The data are two geo-referenced and co-registered 3-look ERS-2 PRI images with pixel sizes of $12.5m$ in both azimuth and range directions. They have been acquired in September and October 2000, around the town of Gloucester, England (SPOT images of the same scene are shown in fig.5.2). The scene corresponds to agricultural regions before and during a flood. Flooded areas, which appear dark due to the specular reflectivity of water, are visible along the river in the middle upper part of fig. 5.5(b). In addition to the classical SAR images difficulties, several water surfaces of the scene were exposed to the wind, whereas others were sheltered from it. Frictional stress at the interface between the water and the wind causes the water to move in the direction of the wind, which produces changes of different signs in the log-ratio image, see fig. 5.6(a). Indeed, one can distinguish the main areas corresponding to the flood, as well as some unexpected areas where strong changes seem to have occurred.

(a) September 9, 2000

(b) October 21, 2000
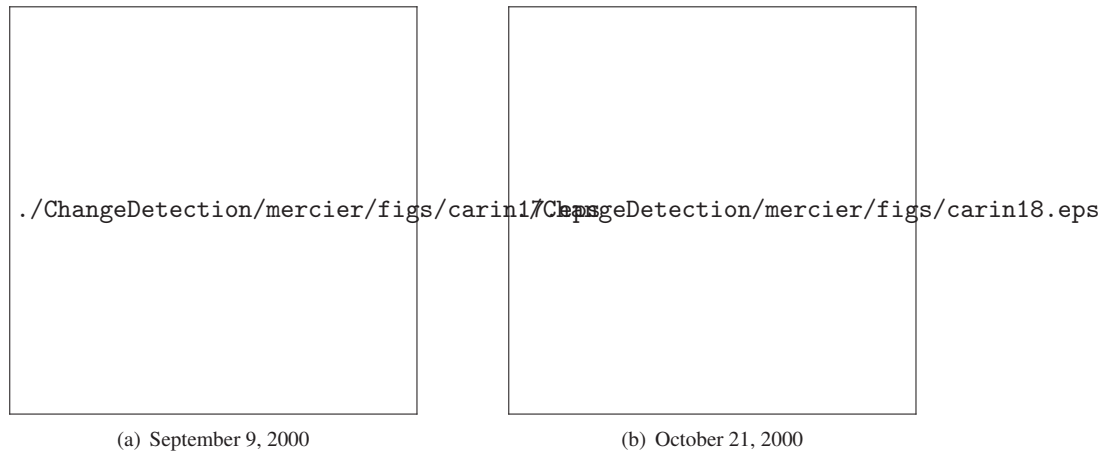
Figure 5.5. *Two co-registered ERS-2 PRI images of agricultural regions near Gloucester, England (512×512), © ESA, Eurimage distribution.*

To highlight the interest of this approach, fig. 5.6(b) presents the result obtained with the fuzzy HMC model under Gaussian class-conditional densities assumption. The segmentation result produced by the model is not a binary map –"change" (white)

(a)   (b)
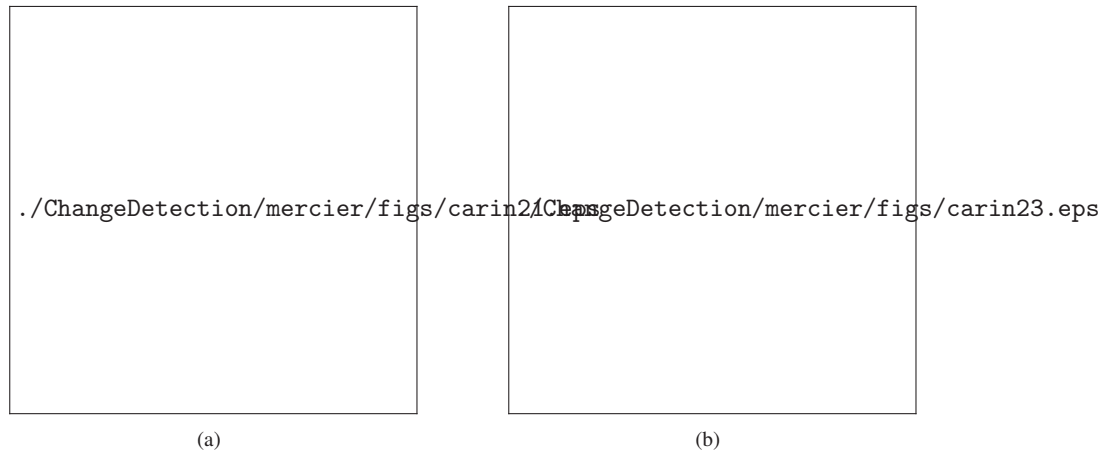
Figure 5.6. *Left : Log-ratio feature image. Right : Change detection map.*

vs "no change" (black)–, but a fuzzy map with grey levels. Pixels intensities are proportional to the attributed fuzzy membership degrees. Bright intensities thus represent high proportion of "no change", and dark ones represent high proportion of "change". The fuzzy HMC model is thus visually interesting on its own, since it allows an expert to interpret the change detection map with its proper knowledge of the scene.

The fuzzy HMC defuzzification result in fig. 5.7(b) has been obtained by performing a threshold at $\varepsilon = 0.5$ on the fuzzy classification result. The corresponding HMC result is presented in fig. 5.7(a). One can easily distinguish in fig. 5.7(b) the main areas corresponding to the flood, whereas the HMC model produces a lot of false alarms.

5.1.2.4. *GIS updating from high resolution images*

Demand for up-to-date geographic data is increasing due to the fast growth of Geographic Information Systems (GIS). With the improvement of spatial resolution, satellite images are becoming a reliable source of information to gather knowledge and to track changes on the ground for map updating. The problem of GIS revision can be seen as to detect changes between heterogeneous datasets, *i.e.* the vectorial database made of objects (polygonal shapes, *e.g.* building, road, lake, fields...) and an image (matrix of pixels). Vector data provides a classified abstract and simplified representation of the landscape, while imagery is an unclassified, continuous and "exhaustive" but resolution-dependent representation of the scene [ALB 08, ARM 02, WAL 04]. Map revision operations for existing topographic databases fall into two parts :

– Change detection, encompassing feature extraction and classification ;

– Change insertion into the GIS, integrating the old and new vectorial data.

(a) HMC
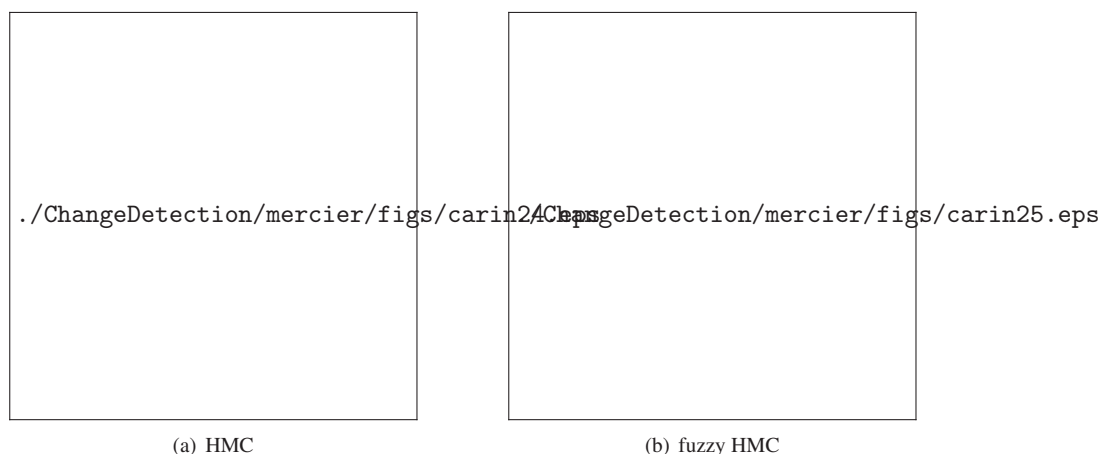
(b) fuzzy HMC

Figure 5.7. *Change detection maps obtained with HMC and fuzzy HMC models with Gaussian distributions.*

The particular case where the GIS must be set up *in extenso* is not considered here. Let t_1 be the date of the already existing GIS, denoted by V_1. Let t_2 be the date of the newly acquired image (*e.g.* radar or optical image, mono or multi-spectral data), denoted by I_2 ($t_1 < t_2$). Revising the GIS consists in including new information extracted from I_2 (object appearance/disappearance, change in object topology or thematic class) in V_1 to get V_2.

Sometimes additional information are also available from archives and can be used to improve the detection process. For example an image I_1, with the same or with a different modality than I_2, can also be included to update the GIS. When I_1 is known, one can think of two strategies to improve updating :

1) applying directly change detection techniques between I_1 and I_2. The so-obtained changes are used to update V_1.

2) learning precisely the rules (geometric, statistic, thematic...) which conduct to V_1, from I_1. These rules can then be applied on I_2 to get directly V_2 (no updating step is required, at least theoretically).

When I_1 is not known, these rules must be estimated from available data, *i.e.* V_1 and I_2 only. The initial GIS V_1 can be seen as a thematic classification of data collected at time t_1, in which each object is assigned to a class. Most of designed algorithms fall into the Post-Classification Comparison (PCC) methods for which change detection is examined over time between suites of independently characterized thematic categories (map-to-map comparison). In this way, it is possible to detect changes and to understand the kinds of transitions that have taken place. The way PCC is done is based on

spatial reasoning which described spatial relation [ALB 08], or by mean of fuzzy set theory [FOO 99]. But it is important to note that this method depends on the results of the classification of both images and data stored in GIS database. In particular, the final change detection map exhibits an accuracy close to the product of the accuracies yielded at the two times [COP 04]. This is due to the fact that PCC does not take into account the dependence existing between two images of the same area acquired at two different times, while Direct Multidata Classification (DMC) [SCH 03] (also called Joint Characterization in [DER 03]) is able to overcome this problem. In this technique, pixels are characterized by a vector obtained by stacking the feature vectors related to the images acquired at two times. Then, change detection is performed by considering each transition as a class and by training a classifier to recognize the transitions.

Before PCC be applied, I_2 must be segmented, and numerous existing algorithms are available. Segmentation algorithms can be applied on the image itself or on some valuable criteria such as, on optical images, Normalized Difference Vegetation Index (NDVI), Soil Brightness Index (SBI), Wetness Index (WeI) and or Water Index (WaI) to exhibit objects of interest in the scene. Segmentation methods are categorized into

– *supervised*, and the available GIS is used to defined learning areas for parameter estimation (*e.g.* maximum likelihood classification, Neural network [UTO 07, PAC 07], SVM [GUO 06]), or

– *unsupervised* and all needed parameters are learned automatically (*e.g.* Isodata, also known as k-means, or EM-based statistical classifier).

The object-based approaches are different from the pixel ones in the sense that classification is performed on groups of pixels that represent already existing objects in the GIS database [WAL 04]. [IM 05] *et al.* proposed a change detection model based on neighborhood correlation image analysis and decision tree classification. [FRA 05] *et al.* used coarse spatial resolution imagery while [DES 06] *et al.* used statistical object-based method for forest change detection. [BON 08] *et al.* developed a method accounting for temporal dependencies in time series with medium to coarse spatial resolution.

5.1.3. *New constraints due to high resolution images*

Many techniques have been proposed for change detection in medium resolution radar or multispectral data. Most of them produce reliable change detection maps by analyzing locally the two images I_1 and I_2. However, the availability of very high resolution data (formally under 1m, *e.g.* images acquired by Quikbird and Ikonos satellites) results in a new set of possible applications, which requires the generation of change detection maps characterized by both a high geometrical precision and the capability of properly modeling the complex objects/areas present in the scene. Nonetheless, classical change-detection techniques are ineffective on very high resolution

images, as they often assume spatial independence among pixels, which is not a reasonable approximation in high geometrical resolution data.

In fact, the increase in the resolution induces new constraints that are more complexe to mitigate :

1) Geometrical considerations. The increase in the resolution makes the evolution of the viewing angle very sensitive to the shape of objects, especially on urban area. Some characteristic points may appear differently or may be hidden from one image to the other. Then, image superposition is becoming difficult. Moreover, it is a challenging task without knowledge of Digital Elevation Model (DEM) or in presence of changes [CHE 07].

2) Radiometrical considerations. The increase in the resolution increases the impact of shadows in objects definition. On the other hand, observation may be affected by BRDF (Bidirectional Reflectance Distribution Function) consideration. Moreover, a class that is homogeneous in its definition (*e.g.* a field) may appear strongly heterogeneous in high resolution (*e.g.* grass or bare soil).

3) Dynamical consideration. The increase in the resolution better characterizes objects of meter-scale (cars, animals, trees. . .) that may move or appear differently from one date to an other.

For those reasons, it is necessary to develop advanced context-sensitive change-detection methods capable to properly exploit the specific properties of very high resolution images [DES 04, ING 07b].

5.2. Time Series Analysis in Remote Sensing Images

5.2.1. *Constraints*

In this context, a change may be considered to as an alteration in the surface components [MIL 88]. This definition may be extended to any type of changes, not only those related to vegetation cover. [SIN 89] defines the process of detecting a change as the process of identifying differences in the state of an object or phenomenon by observing it at different times. The rate of change can be either gradual, such as biomass accumulation or abrupt, as in cases of floods or fires. In a change process, if a type of cover is substituted by another one, then it concerns land-cover *conversion* process. An example is the conversion of forest areas to agricultural ones. A change may also occurs also without modifying the class of the vegetation entity. In this later case, it is related to land-cover *modification*. As an example, a forest may evolute from unmanaged to managed one.

A change may occur at a variety of temporal scales. From this fact, it is critical for [HOB 90] to take into account variability at one scale when trying to interpret

changes at another. As an example, examination of year-to-year variability must take seasonal changes into account. The same author cited the types of variability that have to be considered as :

1) Seasonal response, distinct seasonal activities of the vegetation covers.

2) Interannual variability, changes from one year to another within same seasons.

3) Directional vegetation change, a progressive or irreversible change in the vegetation cover which may be caused by :

    a) Intrinsic vegetation processes (*ie, succession*) ;

    b) Land-use and other human-induced changes ;

    c) Changes in global climatic patterns.

Also, a change may occur at a variety of spatial scales. Researchers studied changes at different spatial scales according to their research objectives. According to these different types of change and change scale, satellite images used to reveal these changes have to be selected with care. Temporal trajectory analysis is used when a seasonal study is to be conducted. Therefore, a time profile of remotely sensed features is constructed from daily images provided by sensors that have a high revisit frequency.

Change detection based on such profiles has been proved to be appropriate for studies at regional scales [LAM 97]. An advantage of this method of change detection is that, changes related to season-scale can be separated from other changes. However, satellites that provide data at high temporal frequency has moderate (*e.g.* $250m$) to low ground spatial resolution (greater than $1km$). At such spatial scales, some of changes are not detectable, which is the case with fragmented surfaces as those of the Brittany region in western France for instance.

Since end-members are considered to be representative to the spectral signature of known natural spectral signals (*e.g.* canopy, soil, shadow...), they have the advantage of providing physically based standardized measures of fractional abundance. Accordingly, temporal changes are reflected in the proportion of each end-member, particularly changes that modify the cover and do not replace it totally [COP 04].

Multi-temporal Spectral Mixture Analysis (MSMA) algorithms were implemented with good results by Adams *et* al. [ADA 95] to monitor the physical nature of land cover changes in the Brazilian Amazon with four time-sequential Landsat TM images. They used a four end-member model consisting of green vegetation, soil, shade and non-photosynthetic vegetation. Fractions of end-members were used to define seven categories, each of which consisted of one or more classes of ground cover. To detect changes from year to year, they measured the differences in the end-members fractions. Roberts *et* al. [ROB 98] introduced the Multiple End-member SMA (MESMA). By this method, the number of end-members and types are allowed to vary for each pixel in the image. They implemented an iterative process whereby a reference spectral library was manipulated in order to derive candidate reference end-members for each

pixel. Although this method was applied to a hyper-spectral data, which allows large number of end-members to be modeled across the scene, it may be applied to multi-spectral data where the number of end-members to be modeled are larger than the dimensionality of the data. Coca *et* al. [Cam 04] presented the Variable End-member SMA (VESMA) to map the fractional vegetation cover and to study the seasonal changes of the vegetation cover. Instead of modeling each pixel in the scene by a sub-set of end-members, as in [ROB 98], they classified the Landsat images and allowed each class to be modeled by a subset of the reference end-members. The major advantage of SMA and its variations (MSMA, MESMA, VESMA...), when compared with other change detection algorithms, lies in its capability to recover natural ecosystem change signals at much finer event scales, *e.g.* thinning in forest ecosystems.

The main difficulty remains in the definition of normal change without prior information. An other difficulty that appears in operational monitoring is the use of heterogeneous data. Three of them will be detailed in this chapter :

1) Time sampling. The general difficulty is related to non-uniform sampling for time series analysis. The problem focuses here on the similarity measure of two time series that are not sampled in the same way.

2) Data validation. Due to geometrical aspects or weather conditions, some samples may not be usable and need first to be recovered to complete the time series before analysis.

3) Data resolution. At present time, spaceborne observations cannot perform fine temporal sampling associated to high spatial resolution observations. Specific studies have to be developed to perform change detection through observations at several resolutions.

5.2.2. *Non-homogeneous time series*

5.2.2.1. *Similarity measure between time series*

The information in a time series may be used to detect not only changes from one season (or from one year to another), but also the drifts from the normal behavior in one season (or one year). This idea is used in [PLI 00] about the impact of the El Niño-Southern Oscillation (ESNO) on East African ecosystem. A temporal trajectory of Advanced Very High Resolution Radiometer (AVHRR) $8 \times 8\ km$ data has been used to cover the period from 1981 to 1994, and two biophysical indicators from this huge source of information has been extracted : NDVI and Surface brightness temperature. Afterward, the monthly anomalies have been calculated by using the Z-score $\frac{x_i - \mu}{\sigma}$, where x_i stands for the data value of a given month i, μ for the mean data value for that month across all years and σ for its the standard deviation. Thereafter, the time series of monthly anomalies is used to study the effects of the ENSO on the vegetation productivity and the surface temperature of the East Africa.

In [DEF 95] many other metrics may be found. They have been used in analyzing a time series AVHRR images. These metrics have been derived from the NDVI temporal profile, observations in red, infrared, and thermal bands. Authors have shown that the separability between classes is improved by using these metrics and recommended 6 robust metrics out of the 16 metrics investigated, *i.e.* mean NDVI, maximum NDVI, NDVI amplitude (the difference between the maximum and minimum NDVI), near-infrared reflectance at maximum NDVI, red reflectance at maximum NDVI and maximum land surface temperature.

An alternative strategy in change detection could be the definition of a similarity measure for "similar evolution" instead of a full characterization of the temporal dynamics of the surface components. This is specifically interesting when the temporal sampling is not similar from a time series to an other one to be compared. Through this point of view, orders in which events appear are more relevant than values of the events measure (*i.e.* spectral signature). Then the exact position and multiplicity of states along the time is ignored : only the order is considered relevant. This means that two sequences can be considered similar as soon as their various elements denote successive states that are similar enough, without taking into account the position of the states in the sequence. As an intuitive example, consider two sequences whose states are denoted A and B, with the first sequence being A - A - B and the second A - B - B. These two sequences can be considered very similar, and their common "behavior" can be summed up as "first A then B". If the data come from multi-date remote sensing images, such a notion of similarity captures the notion of ground evolution, regardless of the rhythm of evolution. Of course, capturing sequence similarity requires to be able to compare pixels from different (single-date) images, as well as being able to "realign" sequences during comparison.

Figure 5.8. *Similarity measure between two series of different size*
(from [KET 07]).

The simple example in Fig. 5.8 illustrates the previous definition [KET 07]. Note that the two sequences $\{s_1, \ldots, s_4\}$ and $\{t_1, \ldots, t_5\}$ of 2D points are of different lengths, and that straight lines connecting the points are here only to illustrate the order

in which the points appear, and are not considered part of the sequences. One possible coupling procedure can be written down as $\{(1,1),(2,1),(3,2),(3,3),(4,4),(4,5)\}$. There are several ways to interpret this coupling. The most intuitive would be to consider that the data went through three distinct "states", the first being drawn by data points s_1, s_2 and t_1, the second by s_3, t_2 and t_3, and the third by s_4, t_4 and t_5. Another interpretation would simply use the six couples as a characterization of the evolution of both sequences. In that case, a sequence may have a single observation to represent several successive "states" (*e.g.* t_1 is part of two successive couples). Finally, the overall similarity measure is based on the sum of each elementary distance between coupled "states". This point of view has been developed in [KET 07] for SPOT-based temporal-spectral clustering approach. It yielded valuable results with promising perspectives on progressive change detection from long time series and also on anomaly detection.

5.2.2.2. *Recovering missing data in a time series*

Images from passive sensors are sensitive to weather conditions. When present, coulds and their shadow blur the corresponding pixel and make it unsuable. When cloud coverage is rare, a rejection procedure of contaminated data may be used (it may concern the affected pixels, inclusive tiles or the complete image.) When the impact of cloud contamination is more important, some compositing techniques have been proposed for low-resolution data (such as AVHRR, MODIS or MERIS...) But it remains a severe problem for higher resolution data when cloud coverage makes a large amount of data unusable.

Therefore, many papers have been published to deal with the problem of clouds and cloud shadows detection, removal and reconstructions. High resolution images (Landsat/ETM+, SPOT/HRVIR, Aster...) have two main differences from low resolution images. Firstly, they hold more information in their spatial domain. Secondly, they are obtained at infrequent time intervals. Accordingly, a recovery process from a high resolution image contaminated by clouds or shadows depends on either the characteristics of the image itself [CHE 05, CHU 04] or on a limited number of images [MEL 06]. Whereas, low resolution data have the advantage of high temporal resolution. Accordingly, one can process many images that encompass a short period of time and select best observations from these images, considering that no changes on the landscape have taken place. Here, we focus on the low resolution data recovering when affected by weather conditions.

Modeling and recovering from a specific type of clouds, the cirrus clouds which normally exist in the upper troposphere and sometimes extend to the stratosphere, has been treated in [Bo- 02]. Authors have found that narrow channels near the center of the strong $1.38\mu m$ water vapor band are effective in detecting thin cirrus clouds. Therefore, they have developed a mathematical model to correct the attenuation caused by these clouds in the $1.38\mu m$ and visible bands. They have obtained good results with MODIS and AVIRIS data.

Most algorithms used in the literature to deal with different types of clouds are those based on using a time series of spatially low resolution data. There are mainly two types of such algorithms : interpolating and compositing ones.

An ecosystem-dependent interpolating algorithms is used in [MOO 05] to fill missing or seasonally snow-covered data in the official MOD43B3 albedo product. It is based on the computation of pixel-level and regional ecosystem-dependent phenological curves. Missing temporal data associated with a given pixel are reconstructed from the most appropriate curve among the available pixel-level and regional curves.

Compositing procedures are much wider used for recovering data from bad weather conditions. They do not require neither modeling different land covers nor *a priori* information on the phenological curves of different ecosystem. The most commonly used compositing method, applied on optical sensors, is the Maximum Value Composite (MVC) [HOL 86, LAS 06], which is mainly applied to AVHRR or SPOT VEGE-TATION. MVC intends to compose the NDVI by taking into consideration the overall data for a predefined time period. The goal is to keep the best observation considering that, for a given period, the maximum NDVI value is related to land surfaces. This method is applied to dense time series of images with often one image per day. As a matter of example, if we consider a 16-day compositing period, the algorithm assumes that no significant change on the land cover may occur during this short period. Hence, choosing the highest NDVI observation from this period, into the time series, will automatically eliminate pixels affected by clouds and other atmospheric phenomena. Even if it is considered as the reference method for image compositing, MVC has several drawbacks [CIH 94, EKL 95, CHE 03]. Some observations are often replaced by the NDVI maximum value, which is not necessarily the best NDVI value available. Moreover, MVC prefers forward scattering observations and increases NDVI values. As a result, the association of pixels acquired with different observation angles creates patchwork artifacts. MODIS vegetation index compositing method is an improvement of the MVC method that considers a Constrained View (CV) angle in addition to the original MVC. The CV-MVC method selects the highest vegetation index value with the lowest sensor viewing zenithal angle [XIA 03]. Although this method reduces effects of high sensor viewing angles, some artifacts remain in the time series [HUE 02]. Many other criteria have been introduced as an alternative to the MVC method, which may be more suitable for some specific problems, such as the Maximum Surface Temperature (MaxTs) for burned scar mapping. A discussion and assessment on different compositing algorithms may be found in [CHU 05]. The Best Index Slope Extraction (BISE) algorithm was proposed in [VIV 92] as an alternative to the MVC method for reducing noise in NDVI time series. It is based on the assumption that sudden falls of vegetation indices are not related to vegetation phenology and should depend on clouds or cloud shadows. Furthermore, a decrease in the NDVI values cannot be immediately followed by an increase of NDVI, because vegetation growth is relatively slow. This method is generally used to monitor inter-annual changes at global scale [WHI 97, REI 07].

Whatever the compositing method, artifacts still remain in the images. Therefore, it is generally necessary to denoise time series as a preprocessing step. The most commonly used techniques for removing local errors are the temporal averaging or median smoothing of time series [REE 94]. The goal is to derive annual phenology at a global or a continental scale [DEF 94, DEF 95, LOV 00]. This preprocessing step can also be applied by using the Fourier transform or the wavelet decomposition technique in order to extract phenological metrics [SAK 05, ROE 00, MOO 01, JAK 01]. These last two methods are used to extract a representative curve of vegetation phenology by vanishing changes of high frequency. Similarly, a *robust* estimator was used in [LEC 05] to recover a time series of NDVI, issued from MODIS images, to monitor land use and land cover changes in the Brittany region of France. Denoising methods based on the Fourier transform, the wavelet decomposition or any *robust* estimator cannot be used to distinguish artifacts from natural changes because they are based on a frequency threshold. Consequently, small variations are rejected even if changes are related to land surface. The use of a robust estimator to recover time series renders the same drawbacks, but it is possible to recover time series based on a contextual model [MEL 06].

Interpolation methods depend either on a well defined mathematical model or on a previously acquired phenological curves. On the other hand, compositing and denoising methods are generally applied to NDVI time series or biophysical indices to observe inter-annual changes in the vegetation growth. Although NDVI has proved to be robust enough for assessing the total amount of green vegetation and the phenology of vegetation (*e.g.* plant greenup and senescence), it has shown limitations in the evaluation of the vegetation cover fraction, since vegetation cover and NDVI are not linearly dependent [BAR 93, GIT 02].

Therefore, a data recovering technique has been selected for its capability to use complete spectral signature of the MODIS observations. This technique is required also to avoid the need of regularly sampled data over time. Then, reflectance denoising from clouds and associated shadows has been investigated, in this study, without NDVI composition or interpolation procedures.

A nonparametric regression procedure adapted from the Kohonen's Self Organizing Map (SOM [KOH 97]) is deeper detailed in this chapter. It uses the ability of the SOM to process incomplete data, data that has some missing values, which has been validated in [COT 05, FES 02]. The problem of processing a MODIS time series affected by atmospheric perturbations may be formulated as a problem of incomplete time series. The erroneous values, pixels contaminated by clouds or shadows, has to be detected first by means of an outlier detector. Detected erroneous values are marked as missing values and then are recovered by the SOM algorithm for missing values. A SOM may be trained on incomplete or complete data, when available. Once trained until convergence, the SOM can be used to estimate the missing values in the time series.

Observations (*i.e.* temporal profiles associated with each pixel) are supposed to be clustered into M classes of R^n (here, n is the number of dates in the time series). x denotes the temoral signature of a given pixel. When the input observation x is an incomplete vector, the set $\mathcal{M}_x$, a sub-set of $\{t_1, t_2, \ldots, t_n\}$, is introduced to define the indices of the missing values. The winning neuron $C_{m_x}(\ell)$ related to x is usually found by using

$$\|x - C_{m_x}\| = \min_{m \in \{1,\ldots,M\}} \|x - C_m\| \tag{5.6}$$

at iteration ℓ. When facing incomplete vector x, the distance $\|xC_m(\ell)\|^2$ is computed with the valid components of x only.

If incomplete observations are to be found in the training set, the update of the nodes (the best-matching one, m_x, and its neighbors in $N_{m_x}(\ell)$) affects the valid components only. By denoting $C_m(\ell) = (C_{m;1}, \ldots, C_{m;k}, \ldots, C_{m;n})^t$ the components of vector $C_m(\ell)$ and $x = (x_1, \ldots, x_n)^t$, the standard Kohonen's SOM training procedure becomes [ABD 08] :

$$C_{m;k}(\ell+1) = \begin{cases} C_{m;k}(\ell) + h_{m,m_x}(\ell)\Big[x_k - C_{m;k}(\ell)\Big] & \text{for } k \notin \mathcal{M}_x, \\ C_{m;k}(\ell) & \text{otherwise.} \end{cases} \tag{5.7}$$

This technique has been applied to a set of MODIS data dedicated to identifying bare soils in the Brittany region during the winter season. Unfortunately, most of observations are not usable due to the presence of clouds or shadows. In order to make such a monitoring available, it is necessary to process as most data as possible.

Ten images are available from November 25, 2002 to April 16, 2003. Almost all the images captured between these dates are contaminated by clouds, but these dates were selected indeed for their minimal amount of cloud coverage. Only bands 1 and 2 of the MODIS data have been processed due to their $250m$ spatial resolution. Therefore, the SOM algorithm for missing values will be applied twice : one process dedicated to the data set constituted by the 10 dates of red reflectances and another independent process applied to the data set constituted by the 10 dates of near-infrared reflectances. The method can be applied in the same way to any other reflectance channel. Fig. 5.9 shows the near-infrared bands at 4 dates during the year 2002 over the Brittany region and the reconstruted data which have been validated by COSTEL for land-use application [ABD 08].

The number of neurons in the SOM grid has to be fixed to come to a compromise between the processing time and the required quantization error (average distance between each vector in the test data to its Best Matching Unit (BMU) in the SOM grid.) One can begin by a reasonably small map, say 12×8, and increase the size of the map to reach a satisfactory quantization error. In our application, an average quantization error of 0.0327 (in the Mean Square Error sense) was obtained in a reasonable time

(a) Original MODIS at t_1

(b) Recovered data at t_1

(c) Original MODIS at t_3

(d) Recovered data at t_3

(e) Original MODIS at t_6

(f) Recovered data at t_6

(g) Original MODIS at t_8

(h) Recovered data at t_8

Figure 5.9. *Typical examples of MODIS data affected by clouds at near infrared band (250m of spatial resolution). The column on the left show some example the original data at time t_1, t_3, t_6 and t_8. The column on the left show the recovered images (by using 10 dates. All of them are contaminated but one).*

of 45 sec. This average quantization error was considered to be reasonably small with respect to the reflectance values in the data set and the dimension of the time series

(10 dates). It is also advisable to have a rectangular shape map if the data are correlated. Then, a rectangular shape of 50×20 of the SOM grid has been choosen for this application.

5.2.2.3. *Super-resolution*

In change detection application for vegetation monitoring or classification map update, a high frequency of data acquisition is often required, leading usually to the use of coarse spatial resolution time series. Moreover, if a high frequency of acquisition is mandatory for vegetation monitoring, spatial resolution is also necessary in order to locate objects of interest. As the remote sensing images dedicated to vegetation can not be acquired with both a high spatial resolution and a high temporal frequency, high temporal frequency is often chosen first. Then it is necessary to detect land-cover changes at subpixel scale, by comparing a coarse resolution (CR) time series to a former high resolution (HR) reference classification. Indeed, such a comparison allows to follow objects that are observable in the HR classification but impossible to distinguish at a CR. The classification is hence used as a mask enabling to study the coherence between the time series and the classification.

In [ROB 07], an *a contrario* detection model has been built that enables to combine linear mixture models. It allows to manage the change in resolution as well as the change vs. no change decision balance with a single probabilistic criterion. This criterion allows the detection of the most coherent solution according to a given classification (that is, a given high-resolution label map), controlling the expected number of false alarms. The complementary of this sub-domain is then considered as the set of pixels about to represent changes. The *a-contrario* detection [DES 07] enables to compute a level of significance without modeling changes nor quantifying the expected differences (noise, distortions, intrinsic variability...), that is with very few *a priori* information. It relies on the idea that a given structure is to be detected if its occurrence is a very unlikely event according to a naive random model on the data. In the context of change detection, this idea is particularly interesting as changes cannot be reasonably modeled through a list of all possible changes (crop rotation, forest cuts or fires can occur on the Earth surface with various size and radiometric intensities). Moreover, an *a priori* model of the no-change surface is also physically difficult to define as temporal evolution profiles of different land-covers vary from a year to another and from a geographical area to another. The aim of *a contrario* modeling is not to reasonably model the data but rather to define a noise model against which significant structures in the data will be detected. Most of the time, this naive model is defined as Gaussian random field of given variance. Performances resulting from a theoretical study of the model according to its main parameters are mentioned in [ROB 07]. Some results have been presented for an agricultural site of the Danubian plain (Rumania, ADAM database).

5.3. Temporal Analysis of SAR Interferometric data

The high potential of Synthetic Aperture Radar (SAR) images for change detection applications is related to the capability of acquiring images night and days, independently from the weather conditions. This allows to obtain images very fast after an event, for instance in the case of a disaster as illustrated in Fig. 5.1, and also to built time series to monitor progressive changes and temporal evolutions as described in the previous section. Moreover, SAR sensors provide complex images and both amplitude and phase carry information which may be useful for change analysis. In this section, we will fucus on the use of the phase information and more specifically on repeat pass SAR Interferometry (InSAR) which measures phase differences between two SAR images acquired at different time. We first recall some constraints to use interferometric data and introduce the two main products : the interferogram and the coherence images. The potential and the limitations of these images for change analysis is illustrated with an application : the monitoring of Alpine glaciers by SAR differential interferometry.

5.3.1. *Constraints*

SAR interferometry is based on the computation of the phase difference between two SAR images acquired either simultaneously with a spatial baseline between two antennas in a bistatic configuration (data provided up to now by airborne systems or by the SRTM mission and starting 2009 by the satellite TerraSAR-X tandem mission) or by repeated passes of the sensor. The spatial baseline creates a stereoscopic configuration which allows to measure the topography whereas the time intervale between the two acquisitions (also called temporal baseline) allows to observe surface change and target displacement between the two dates. Differential SAR interferometry (D-InSAR) consists in removing the geometrical phase component due to the topography (by using an external Digital Elevation Model or several interferograms) in order to measure the displacement field associated to abrupt or progressive changes such as earthquakes, volcano activity, urban subsidence, glacier motion...

The InSAR information is extracted from a pair of complex SAR images, u_1 and u_2, by a processing chain which includes the following "critical" steps :

– SAR synthesis where interferometric constraints are taken into account in order to reduce the noise (the same zero Doppler has to be used to observe targets from the same viewing angle and the non-overlapping parts of the spectra are removed),

– sub-pixel cross-registration of the two resulting Single Look Complex images in order to measure the phase difference from the same resolution cell,

– computation of the normalized complex correlation between these two images :

$$\rho e^{i\phi} = \frac{\langle u_1 u_2^* \rangle}{\sqrt{\langle |u_1|^2 \rangle \langle |u_2|^2 \rangle}} \tag{5.8}$$

by replacing the mathematical expectancy $\langle . \rangle$ by a spatial averaging under the assumption of local stationarity :

$$\hat{\rho}(m,n)e^{i\hat{\phi}(m,n)} = \frac{\left| \sum_{i=1}^{L} u_1(i)u_2^*(i) \right|}{\sqrt{\sum_{i=1}^{L} |u_1(i)|^2 \sum_{i=1}^{L} |u_2(i)|^2}}. \tag{5.9}$$

The magnitude of the correlation provides the so-called coherence image $\hat{\rho}(m,n)$ and the argument $\hat{\phi}(m,n)$ the interferogram.

– in the case of differential interferometry, the removal of the topographic fringes by subtracting the estimated path differences due to the target distance to the two sensor positions provided by the orbital data.

– phase filtering to reduce the noise, especially in low coherence areas.

– phase unwrapping (in the areas where the phase signal is reliable) to retrieve the 2π multiple to be added to the fringes to obtain a continuous measure.

The phase of the complex SAR image can be decomposed in two terms, one due to the propagation between the antenna and the "phase center" of the resolution cell, and one due to the backscattering process which depends on the surface characteristics (dielectric constants, surface roughness, wave penetration...) The main constraint of InSAR technique is that the "backscattering" term is the same in the two acquisitions and disappear when the phase difference is computed. This is necessary to access to the "propagation" term and retrieve the position or displacement information. If surface change occurs between the two acquisitions (vegetation natural evolution, deforestation, flood, etc.), the "backscattering" term is different between the two acquisitions and introduce noise in the interferogram. Small changes at the scale of the wavelength create a noise level which can be recovered by filtering techniques whereas strong changes results in a uniform phase distribution which totally hides the "propagation" term. The coherence level reveals this phase instability and can be considered as an indicator of changes in the backscattering surface as illustrated in section 5.3.2. The coherence image, usually after appropriate filtering, is also used as a confidence measure to reveal the areas where the phase information is reliable.

Even when the coherence level is sufficient to observe fringe patterns, the interferogram measures a propagation path delay variation which may suffer from different artefacts. The mains error sources are the correctness of the DEM (the sensitivity increases with the baseline), the lack of precision in the orbits and different propagation speeds due to variations of the atmospheric conditions within the scene between the two acquisitions. The artefacts known as "atmospheric perturbations" are actually the main limitations in the precision of D-InSAR measurements. Different approaches based on the use of interferogram time series have been introduced to overcome these difficulties. Some methods dedicated to the analysis of interferogram time series will be briefly introduced in subsection 5.3.3.

5.3.2. *Analysis of coherence information*

5.3.2.1. *Sources of decorrelation*

A major limitation on the applicability of interferometry is the preservation of coherence. An analysis of the sources of decorrelation in interferometric radar echoes may be found in [ZEB 92]. Zebker *et* al. have proposed to separate decorrelation due sensor geometry from that dependent on temporal changes. The observed coherence (denoted ρ_{total}) can be written as the product of three terms as :

$$\rho_{total} = \rho_{SNR} \cdot \rho_{temporal} \cdot \rho_{spatial} \qquad (5.10)$$

where ρ_{SNR} is the decorrelation term due to the signal to noise ratio. $\rho_{temporal}$ corresponds to decorrelation induced by changes in the target surface between the two SAR acquisitions. It includes decorrelation due to surface change (*e.g.* vegetation growth, ice freezing or melting) and motion (*e.g.* subsidence, glacier displacement, landslide). $\rho_{spatial}$ corresponds to decorrelation induced by the sensor geometry. Indeed, for an interferometric configuration, the two sensors are separated by a spatial baseline. This leads to observe the target with a slightly different incidence angle and causes a spatial baseline decorrelation.

5.3.2.2. *Multi-date coherence analysis*

The potential of SAR interferometry to monitor changes depends on the combination of these decorrelation sources in the available data sets. For glacier monitoring for instance, most of the results have been obtained by using 1-day interferograms. In [TRO 07], a series of tandem interferograms has been analyzed to investigate the complementarity of InSAR data with summer optical data and ground measurements. Fig. 5.10 illustrates 5 interferometric coherence computed from July 1995 to April 1996 over the Chamonix/Mont-Blanc test-site.

Among the five interferometric couples shown in Fig. 5.10, the highest coherence is observed in March 1996. Indeed, almost 45% of the Argentière glacier has a coherence greater than 0.5. This couple has the smallest perpendicular baseline ($B_\perp = 9m$). Moreover, the interferograms acquired in winter show a better coherence than those acquired in the other seasons. Compared to March, the 31 Dec.95/1 Jan.96 interferogram shows a loss of coherence. This loss is observed on glacier (2%) and non glacier (7%) areas. The large perpendicular baseline ($B_\perp = 208m$) causes a significant spatial decorrelation, especially in the valley forested areas. In warmer seasons (October and April), coherence is preserved only on the upper parts of the glaciers. Indeed, for low altitudes, the higher temperature leads to a relevant change of the glacier surface state and causes a loss of the interferometric coherence. For the same reason, during the hot season, coherence is not preserved on the whole surface of the nine observed glaciers.
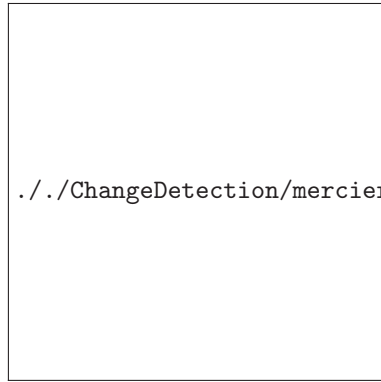
(a) Mask

(b) 9/10 Jul.95, $B_\perp = 52$m
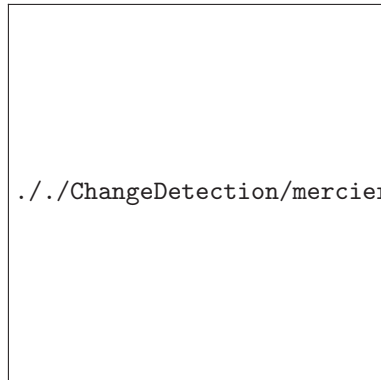
(c) 22/23 Oct.95, $B_\perp = -107$m
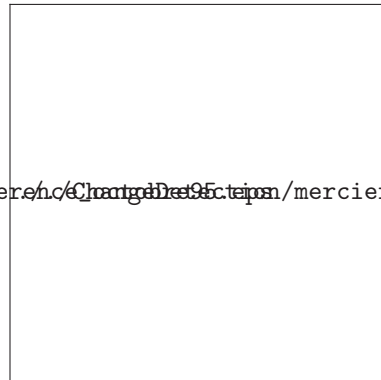
(d) 31 Dec.95/1 Jan.96, $B_\perp = 208$m
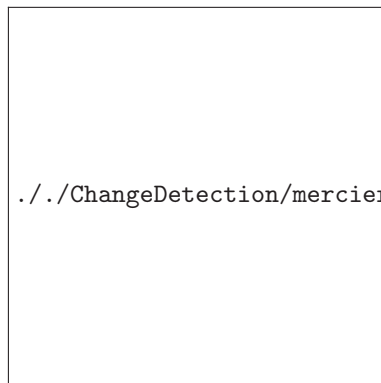
(e) 10/11 Mar.96, $B_\perp = 9$m

(f) 14/15 Apr.96, $B_\perp = 93$m

Figure 5.10. *ERS-1/2 tandem interferograms over Chamonix Mont-blanc area
(1024×1024 pixels). (a) glaciers mask ; (b)-(f) Coherence filtered by
amplitude driven adaptive neighborhood, from summer 1995 to spring 1996.*

./ChangeDetection/mercier/figs/coherence_altitude_argentiere_anglais.eps

(a)
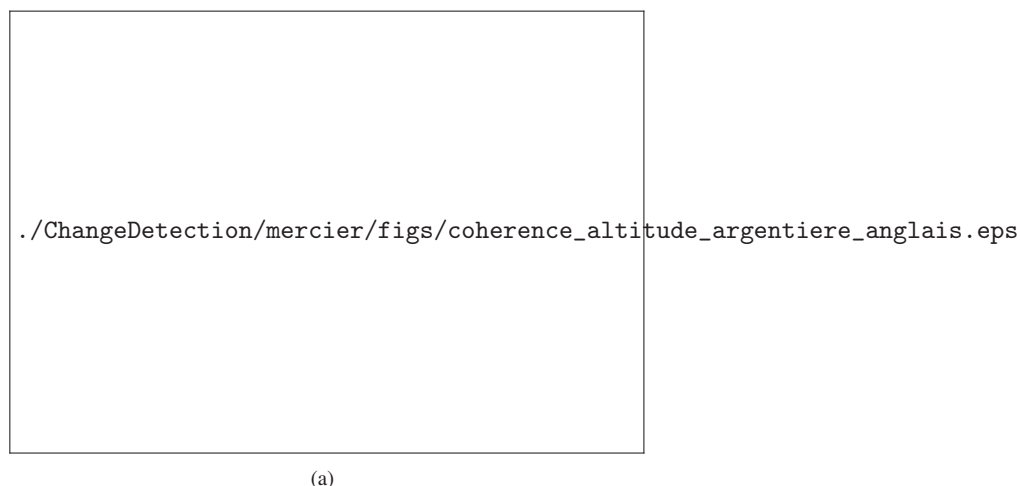
Figure 5.11. *Influence of altitude and slope on the interferometric coherence.*

In April 1996, coherence is preserved only on the upper part of the Argentière glacier. 35% of the pixels have a coherence greater than 0.5, whereas on the Mer de Glace/Leschaux glacier, only 1% of the pixels preserve the coherence. As the altitude is higher on the top of Argentière glacier than on the Mer de Glace/Leschaux glacier, coherence preservation may be dependent on glacier elevation. To investigate this relation, a Digital Terrain Model (DTM) has been used to build a mask of visibility. This mask determines which regions are not visible (foldover, shadow) on the ERS-1/2 images. Then, the transformation from ground geometry to the SAR slant range geometry is coded in a look up table which is used to convert the mask of visibility and resample the DTM in slant range. Only visible regions are taken into account in the plot of the coherence as a function of altitude. Fig. 5.11 shows the average coherence as a function of the altitude and slope on the Argentière glacier. A low coherence can be observed in December 1995 (green line) due to the large baseline of this interferometric pair. The significant change of the glacier surface state in summer does not allow a coherence preservation. Note that in October 1995, March and April 1996, coherence increases with altitude on the Argentière glacier. Nevertheless, for high altitude ($\geq$ 2900m), coherence falls down. This decrease is explained by the increasing of slope at those altitude.

5.3.3. *Analysis of phase information*

5.3.3.1. *Sources of fringe patterns*

The phase difference $\hat{\phi}(m, n)$ estimated by the computation of the interferogram according to equation (5.9) includes several contributions which can be decomposed

as :

$$\hat{\phi}(m,n) = [\phi_{orbit} + \phi_{topo} + \phi_{disp} + \phi_{atmo} + \phi_{noise}]_{\text{modulo } 2\pi} \qquad (5.11)$$

where :

$-\ \phi_{orbit}$ is the phase variation corresponding to the path difference induced by the baseline between the two passes. Normally, this term should be removed in the processing by using the satellite orbits (or the flight trajectories for airborne data). However, a very low frequency fringe pattern may subsist across the interferogram due to the imprecisions of the orbits. A posteriori corrections are often performed to remove the remaining orbital fringes by subtracting a first order phase model estimated from the areas where the phase signal should be constant.

$-\ \phi_{topo}$ is the phase variation due to the topography. If the elevation z is known, this term can be compensated according to the baseline component perpendicular to the line of sight $B_\perp$, the local incidence angle θ and the distance to the sensor R_1 :

$$\phi_{topo} = \frac{4\pi}{\lambda} \frac{B_\perp}{R_1 \sin\theta} \Delta z = \frac{2\pi}{e_a} z \qquad \text{with} \qquad e_a = \frac{\lambda}{2} \frac{R_1 \sin\theta}{B_\perp} \qquad (5.12)$$

where e_a denotes the height of ambiguity, which is the height difference corresponding to a 2π phase rotation. If the DEM used to remove the topographic phase contribution has error, every differential interferogram built with this DEM contains residual fringes at the same location and the number of fringes is inversely proportional to the height of ambiguity. With a single interferogram, it is difficult to discriminate this artefact from other phase signal whereas by using a series of interferograms, the linear relation with the baseline can be used to detect the DEM errors as illustrated in Figures 5.12–e,–f.

$-\ \phi_{disp}$ is the sought-after phase signal which measures the projection of the surface displacement in the direction of the SAR line of sight. The detection of this signal provides both a quantitative and qualitative information regarding the studied phenomena. Qualitatively, the detection of a displacement informs on the presence of the phenomena (urban areas affected by subsidence for instance) and potential sources of risk : building displacements, landslides... Quantitatively, the precision of the displacement measure can reach centimeters in X and C band and even millimeters by using interferogram time series. This allows to derive information regarding the amplitude of the deformation or to invert the geophysical model parameters of the phenomena which creates the measured surface displacements (slip distribution on a fault during an earthquake [PED 03], magma transport and storage [FRO 04], etc.),

$-\ \phi_{atmo}$ is the delay due to variations of the atmospheric conditions between the two acquisitions. The variations of the propagation in the ionosphere are usually considered as very low frequency artifacts and can be corrected by the same approach as orbital errors. The variations of the propagation in the troposphere are mostly related to the water vapor pressure and temperature (Smith and Weintraub formula) [TAR 96]

which can vary locally in the scene and change between the two acquisition according to the meteorological conditions. These artifacts create "undesirable changes" in the interferogram which have to be discriminated from the phase signal related to the studied deformation.

$- \phi_{noise}$ is the noise which affects the wrapped phase. Mis-registration and temporal decorrelation are among the main source of noise. The main consequence of this noise is the increased difficulty of the phase unwrapping step since the wrapped phase gradient is corrupted with 2π errors which can propagate or create a biased estimation of the unwrapped phase [TRO 98]. The initial spatial averaging performed in equation (5.9) or adaptive filtering techniques [VAS 06] can be used to strongly reduce this noise term.

5.3.3.2. *Different approaches to retrieve displacement information*

Different approaches have been proposed to discriminate the deformation phase signal from the atmospheric perturbations which strongly reduces the D-InSAR reliability in terms of "displacement detection" and D-InSAR precision in terms of "displacement measurement". They are based on the use of interferogram time series and the hypothesis that the displacement appears as a deterministic function of the temporal baseline whereas the atmospheric perturbations are considered as having a random behavior, spatially correlated but temporally decorrelated. The different approaches which can be used to try to retrieve the displacement information in interferogram time series depends on the target characteristics. Two kinds of targets are usually discriminated.

1) **Coherent targets** - Coherent scatterers are characterized by a stable response which may be detected and recognized at different dates. The term "Coherent" refers to the stability of the target response regarding the viewing angles : the "backscattering" part of the phase is the same when the target is seen from different positions along the synthetic antenna, and the coherent summation performed in the SAR synthesis results in a high amplitude which is used to detect such targets. They usually correspond to pixels where the signal is dominated by a single target with a stronger backscattering coefficient than surrounding targets. When coherent scatterers are also stable in time, they are called "Permanent" or "Persistent" scatterers (PS). Such targets are the best "candidates" to extract the displacement information from phase difference time series since the viewing angle stability makes them more robust to the baseline decorrelation and the temporal stability makes them more robust to the temporal decorrelation. The first method based on these Permanent Scatterers has been proposed by Ferretti et al. [FER 01]. It has been successfully applied to detect and measure ground deformations in urban areas where the buildings and man made features create numerous PS. Different versions of this approach have been proposed to improve the PS selection and separate the atmospheric artefact and the displacement phase signal by using linear or more complex models. A comparative analysis of their performances has been conducted by the European Space Agency [RAU 09].

2) **Non-coherent targets** - Distributed media have a response which results from the sum of the responses of the elementary targets over the surface (or volume) covered by the resolution cell. This case corresponds to the conventional fully developed speckle model [MAI 08] and makes the phase of the backscattered signal very sensitive to the viewing angle and temporal evolution of the elementary targets. When the speckle is temporally correlated, the coherence is high and the "propagation part" of the phase difference can be estimated. The temporal stability depends on the land cover (rocks, sand, vegetation, crops...) Except in specific areas such as deserts, the longer the intervale between the acquisitions, the lower the coherence. Temporal evolutions and baseline decorrelation make the use of large time series more difficult for distributed targets than for coherent targets. To overcome this difficulty, a different approach has been proposed based on the use of Small BAseline Subsets (SBAS) [BER 02]. Differential interferograms obtained from pairs of images with the most appropriate spatial and temporal baselines are selected in the graph of all possible interferograms. Various procedures (Least Mean Squares, Singular Value Decomposition...) are used to derive the deformation signal from the connected phase values [LOP 09]. This approach is less restrictive than the PS approach in terms of scatterers stability over the whole data set. It can be applied to study changes over natural areas where the number of PS is too small for the previous approach.

An "intermediate" approach has also been proposed by Hooper et al. [HOO 08] with a selection of « reliable » pixels not only based on the amplitude, but also on the interferometric phase stability (in space and time). As phase unwrapping is performed in 3-D (space-time), assumptions on the temporal displacement behavior are not necessary and can be replaced by a simple temporal smoothing procedure, which is more adapted to the needs of applications which do not provide strong prior information.

5.3.3.3. *Fringe pattern evolution*

The interest and the difficulty of change analysis in series of SAR interferograms are illustrated in Figure 5.12. The 4 differential interferograms acquired between October 1995 and April 1996 by ERS tandem mission, show the 1-day displacements of several glaciers of the Chamonix Mont-Blanc test site (cf. Figure 5.10). One can observe very similar fringe patterns at the different dates over the Mer-de-Glace and Argentière glaciers for instance. These fringes reveal the areas where the glaciers accelerate or slow down and the repeatability of the patterns tends to confirm that atmospheric perturbations can be neglected at the scale of these glaciers in one day interferograms acquired during the cold season.

To detect fringe evolutions, it is interesting to compute the wrapped phase difference between those differential interferograms as illustrated in Figure 5.12–e,–f. On the three glaciers located in the upper part, one can observe fringe patterns which are mainly due to DEM errors over those glaciers. This hypothesis has been confirmed by measuring the correlation of the phase difference with the baseline differences. On the

./../ChangeDetection/mercier/figs/phase_octobre95.eps

./../ChangeDetection/mercier/figs/phase_decembre95.e

(a) 22/23 Oct.95, $B_\perp = -107$m

(b) 31 Dec.95/1 Jan.96, $B_\perp = 208$m

./../ChangeDetection/mercier/figs/phase_mars96.eps

./../ChangeDetection/mercier/figs/phase_avril96.eps

(c) 10/11 Mar.96, $B_\perp = 9$m

(d) 14/15 Apr.96, $B_\perp = 93$m

./../ChangeDetection/mercier/figs/phase_Avril_Octobre_IDAN.eps

./../ChangeDetection/mercier/figs/phase_Avril_Mars_I

(e) Difference 14/15 Apr.96 - 22/23 Oct.95

(f) Difference 14/15 Apr.96 - 10/11 Mar.96

Figure 5.12. *ERS-1/2 tandem differential interferograms over Chamonix Mont-blanc area (1024×1024 pixels). (a)-(d) : Phase filtered by amplitude driven adaptive neighborhood, from autumn 1995 to spring 1996 ; (e)-(f) : evolution of the fringe pattern over the studied glaciers (cf. Fig.5.10).*

other glaciers, in the areas where the coherence is high enough in both interferograms, the phase difference is almost constant between March and April 1996 interferograms, which reveals that the displacement fields did not change. The April 1996 - October 1995 difference shows more significant variations (more than one fringe over Argentière glacier for instance) probably due to the evolution of the displacement fields between autumn and spring. The next generation of SAR satellites launched in 2006-2007 (TerraSAR-X with 11 days repeat cycle and COSMO-SkyMed with 4-day intervals with the full constellation) will provide new multi-temporal data sets and allow further research to investigate the potential of X-band high resolution images for such application.

[ABD 08] ABDELLATIF B., LECERF R., MERCIER G., HUBERT-MOY L., « Low resolution time series analysis with erroneous data », *IEEE Trans. Geosci. Remote Sensing*, vol. 46, n°7, juillet 2008.

[ADA 95] ADAMS J., SABOL D., KAPOS V., FILHO A., ROBERS D., SMITH M., GILLESPIE A., « Classification of multi-spectral images based on fractions of end members : applications to land-cover change in the Brazilian Amazon », *Remote Sensing of Environment*, vol. 12, p. 137–154, 1995.

[ALB 08] ALBOODY A., SEDES F., INGLADA J., « Post-classification and spatial reasoning : new approach to change detection for updating GIS database », *3rd Int. Conf. on Information and Communication Technologies : From Theory to Applications (ICTTA'08)*, Damascus, Syria, p. 1–7, 7-11 April 2008.

[Ami 04] AMITHA PERERA A., HOOGS A., « Bayesian object-level change detection in grayscale imagery », *Proc. of the 17th Int. Conf. on Pattern Recognition (ICPR'04)*, vol. 1, p. 71–75, août 2004.

[ARM 02] ARMENAKIS C., CYR I., PAPANIKOLAOU E., « Change detection methods for the revision of topographic databases », *Symposium on Geospatial theory, processing and applications, ISPRS Commission IV*, Ottawa, Canada, p. 1–7, 9-12 July 2002.

[AVR 02] AVRAHAM G., SHOSHANY M., NETANYAHU N., « Spectral and spatial parameterization of multi-date satellite images for change detection of linear features », *IEEE Trans. Geosci. Remote Sensing*, vol. 3, n°6, p. 1848–1850, juin 2002.

[BAR 93] BARET F., JACQUEMOUD S., HANOCQ J., « About the soil line concept in remote sensing », *Advances in Space Research*, vol. 13, p. 281–284, 1993.

[BAZ 05] BAZI Y., BRUZZONE L., MELGANI F., « An unsupervised approach based on the generalized Gaussian model to automatic change detection in multitemporal SAR images », *IEEE Trans. Geosci. Remote Sensing*, vol. 43, n°4, p. 874–887, avril 2005.

[BEA 04] BEAUCHEMIN M., FUNG K., « Investigation of multiscale product for change detection in difference images », *Proc of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, vol. 6, p. 3853–3856, septembre 2004.

[BER 02] BERARDINO P., FORNARO G., LANARI R., SANSOSTI E., « A new algorithm for surface deformation monitoring based on Small Baseline Differential SAR Interferograms », *IEEE Trans. Geosci. and Remote Sensing*, vol. 40, p. 2375–2382, 2002.

[Bo- 02] Bo-Cai G., Ping Y., Wei H., Rong-Rong L., Wiscombe W., « An algorithm using visible and 1.38-/spl mu/m channels to retrieve cirrus cloud reflectances from aircraft and satellite data », *IEEE Trans. Geosci. Remote Sensing*, vol. 40, n°8, p. 1659–1668, 2002.

[BON 08] Bontemps S., Bogaert P., Titeux N., Defourny P., « An object-based change detection method accounting for temporal dependences in time series with medium to coarse spatial resolution », *Remote Sensing of Environment*, vol. 112, p. 3181–3191, 2008.

[BOV 07] Bovolo F., Bruzzone L., « A novel theoretical framework for unsupervised change detection based on CVA in polar domain », *IEEE Trans. Geosci. Remote Sensing*, vol. 45, n°1, p. 218–236, 2007.

[BRU 99] Bruzzone L., Fernàndez Prieto D., « An MRF approach to unsupervised change detection », *Proc of the IEEE International Conference on Image Processing (ICIP)*, vol. 1, p. 143–147, 1999.

[BRU 00] Bruzzone L., Fernàndez Prieto D., « Automatic analysis of the difference for unsupervised change detection », *IEEE Trans. Geosci. Remote Sensing*, vol. 38, p. 1171–1182, mai 2000.

[BRU 02] Bruzzone L., Fernàndez Prieto D., « An adaptive semiparametric and context-based approach to unsupervised change detection in multitemporal remote sensing images », *IEEE Trans. Image Processing*, vol. 11, p. 452–466, avril 2002.

[BRU 04] Bruzzone L., Marconcini M., Wegmuller U., Wiesmann A., « An advanced system for the automatic classification of multitemporal SAR images », *IEEE Trans. Geosci. Remote Sensing*, vol. 42, n°6, p. 1321–1334, juin 2004.

[BUJ 04] Bujor F., Trouvé E., Valet E., Nicolas J.-M., Rudant J., « Application of log-cumulants to the detection of spatiotemporal discontinuities in multitemporal SAR images », *IEEE Trans. Geosci. Remote Sensing*, vol. 42, n°10, p. 2073–2084, octobre 2004.

[BYR 98] Byrne G., Crapper P., Mayo K., « Monitoring land-cover change by principal component analysis of multitemporal Landsat data », *Remote Sensing of Environment*, vol. 10, p. 175–184, 198.

[CAI 93] Caillol H., Hillion A., Pieczynski W., « Fuzzy random fields and unsupervised image segmentation », *IEEE Trans. Geosci. Remote Sensing*, vol. 31, n°4, p. 801–810, 1993.

[CAI 97] Caillol H., Pieczynski W., Hillion A., « Estimation of fuzzy Gaussian mixture and unsupervised statistical image segmentation », *IEEE Trans. Image Processing*, vol. 6, n°3, p. 425–440, 1997.

[Cam 04] Camacho-De Coca F., García-Haro F., Gilabert M., Meliá J., « Vegetation cover seasonal changes assessment from TM imagery in a semi-arid landscape », *Int. Journal of Remote Sensing*, vol. 25, p. 3451–3476, 2004.

[CAR 06] Carincotte C., Derrode S., Bourennane S., « Unsupervised change detection on SAR images using fuzzy hidden Markov chains », *IEEE Trans. Geosci. Remote Sensing*, vol. 44, n°2, p. 432–441, 2006.

[CHE 03] CHEN P., SRINIVASAN R., FEDOSEJEVS G., KINIRY J., « Evaluating different NDVI composite techniques using NOAA-14 AVHRR data », *Int. Journal of Remote Sensing*, vol. 24, n°17, p. 3403–3412, Taylor & Francis, 2003.

[CHE 05] CHEN F., ZHAO Z., PENG L., YAN D., « Clouds and cloud shadows removal from high resolution remote sensing images », *Proc of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, vol. 6, p. 4256–4259, 2005.

[CHE 07] CHESNEL A.-L., BINET R., WALD L., « Object oriented assessment of damage due to natural disaster using very high resolution images », *Proc of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, p. 3736–3739, juillet 2007.

[CHU 04] CHUN F., JIAN-WEN M., QIN D., XUE C., « An improved method for cloud removal in ASTER data change detection », *Proc of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, vol. 5, p. 3387–3389, 2004.

[CHU 05] CHUVIECO E., VENTURA G., PILAR MARTIN M., GÉOMEZ I., « Assessment of multitemporal compositing techniques of MODIS and AVHRR images for burned land mapping », *Remote Sensing of Environment*, vol. 94, n°4, p. 450–462, 2005.

[CIH 94] CIHLAR J., MANAK D., VOISIN N., « AVHRR bidirectional reflectance effects and compositing », *Remote Sensing of Environment*, vol. 48, n°1, p. 77–88, 1994.

[COP 04] COPPIN P., JONCKHEERE I., NACKAERTS K., MUYS B., LAMBIN E., « Digital change detection methods in ecosystem monitoring : a review », *Int. Journal of Remote Sensing*, vol. 25, n°9, p. 1565–1596, May 2004.

[COT 05] COTTRELL M., LETRÉMY P., « Missing values : processing whith Kohonen algorithm », *Int. Symp. on Applied Stochastic Models and Data Analysis (ASMDA)*, Brest, France, p. 17–20, 2005.

[D'A 04] D'ADDABBO A., SATALINO G., PASQUARIELLO G., BLONDA P., « Three different unsupervised methods for change detection : an application », *Proc of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, vol. 3, p. 1980–1983, septembre 2004.

[DEE 98] DEER P., Digital change detection in remotely sensed imagery using fuzzy set theory, PhD thesis, University of Adelaïde, Australia, Department of Geography and Computer Science, 1998.

[DEE 02] DEER P., EKLUND P., « Values for the fuzzy C-means classifier in change detection for remote sensing », *Int. Conf. on Information Processing and Management of Uncertainty*, p. 187–194, 2002.

[DEF 94] DEFRIES R., TOWNSHEND J., « NDVI-derived land cover classifications at a global scale », *Int. Journal of Remote Sensing*, vol. 15, n°17, p. 3567–3586, 1994.

[DEF 95] DEFRIES R., HANSEN M., TOWNSHEND J., « Global discrimination of land cover types from metrics derived from AVHRR pathfinder data », *Remote Sensing of Environment*, vol. 54, n°3, p. 209–222, 1995.

[DEL 97] DELIGNON Y., GARELLO R., HILLION A., « Statistical modelling of ocean SAR images », *IEE Proc. on Radar, Sonar and Navig.*, vol. 44, n°66, p. 348–354, 1997.

[DER 03] Derrode S., Mercier G., Pieczynski W., « Unsupervised change detection in SAR images using a multicomponent HMC model », *Proc. of the 2nd Int. MultiTemp Workshop*, Ispra (Italy), 16-18 July 2003.

[DES 04] Desclee B., Bogaert P., Defourny P., « Object-based method for automatic forest change detection », *Proc of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, vol. 5, p. 3383–3386, septembre , 20-24 2004.

[DES 06] Desclée B., Bogaert P., Defourny P., « Forest change detection by statistical object-based method », *Remote Sensing of Environment*, vol. 102, p. 1–11, 2006.

[DES 07] Desolneux A., Moisan L., Morel J., *From gestalt theory to image analysis - a probabilistic approach*, Springer, 2007.

[Dre 93] Dreshler-Fischer L., Drewniok C., « A knowledge based approach to the detection and interpretation of changes in aerial images », *Proc of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, vol. 1, p. 159–161, 1993.

[DVO 83] Dvorchenko V. N., « Bounds on (deterministic) correlation functions with applications to registration », *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. PAMI-5, p. 206–213, février 1983.

[EKL 95] Eklundh L. R., « Noise estimation in NOAA AVHRR maximum-value composite NDVI images », *Int. Journal of Remote Sensing*, vol. 16, n°15, p. 2955–2962, Taylor and Francis, 1995.

[FER 01] Ferretti A., Prati C., Rocca F., « Permanent scatterers in SAR interferometry », *IEEE Trans. Geosci. and Remote Sensing*, vol. 39, n°1, p. 8-20, 2001.

[FES 02] Fessant F., Midenet S., « Self-organising map for data imputation and correction in surveys », *Neural Computing & Applications*, vol. 10, n°4, p. 300–310, 2002.

[FOO 99] Foody G., Boyd D., « Detection of partial land cover change associated with the migration of inter-class transitionnal zone », *Int. J. of Remote Sensing*, vol. 14, p. 2723–2740, 1999.

[FRA 05] Fraser R. H., Abuelgasimb A., Latifovic R., « A method for detecting large-scale forest cover change using coarse spatial resolution imagery », *Remote Sensing of Environment*, vol. 95, p. 414–427, 2005.

[FRO 04] Froger J.-L., Fukushima Y., Briole P., Staudacher T., Souriot T., Villeneuce N., « The deformation field of the August 2003 eruption at Piton de la Fournaise, Reunion Island, mapped by ASAR interrferometry », *Geophys. Res. Lett.*, vol. 31, 2004.

[FUN 87] Fung T., Ledrew E., « Application of principal components analysis to change detection », *Photogrammetric Engineering & Remote Sensing*, vol. 53, p. 1649–1658, 1987.

[FUN 90] Fung T., « An assessment of TM imagery for land-cover change detection », *IEEE Trans. Geosci. Remote Sensing*, vol. 28, n°4, p. 681–684, 1990.

[GIT 02] Gitelson A., Stark R., Grits U., Rundquist D., Kaufman Y., Derry D., « Vegetation and soil lines in visible spectral space : a concept and technique for remote estimation of vegetation fraction », *Int. Journal of Remote Sensing*, vol. 23, n°13, p. 2537–2562, 2002.

[GUO 06] GUORUI M., HAIGANG S., PINGXIANG L., QIANQING Q., « A kernel change detection algorithm in remote sense imagery », *Proc of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, juillet 2006.

[HIL 00] HILGER K., NIELSEN A., « Targeting input data for change detection studies by suppression of undesired spectra », *Seminar on Remote Sensing and Image Analysis Techniques for Revision of Topographic Databases*, Copenhagen, Denmark, February, 29 2000.

[HOB 90] HOBBS R., « Remote sensing of spatial and temporal dynamics of vegetation », *Remote Sensing of Biosphere Functioning. Ecological Studies 79.*, p. 203–219, Springer-Verlag New York, Inc., New York., 1990.

[HOL 86] HOLBEN B. N., « Characteristics of maximum-value composite images from temporal AVHRR data », *Int. Journal of Remote Sensing*, vol. 7, n°11, p. 1405–1419, 1986.

[HOO 08] HOOPER A., « A multi-temporal InSAR method incorporating both persistent scatterer and small baseline approaches », *Geophysical Research Letters*, vol. 35, 2008.

[HUE 02] HUETE A., DIDAN K., MIURA T., RODRIGUEZ E., GAO X., FERREIRA L., « Overview of the radiometric and biophysical performance of the MODIS vegetation indices », *Remote Sensing of Environment*, vol. 83, n°1-2, p. 195–213, 2002.

[HWA 04] HWANG Y., KIM J.-S., KWEON I., « Change detection using a statistical model of the noise in color images », *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS'04)*, vol. 3, p. 2713–2718, octobre 2004.

[IM 05] IM J., JANSEN J. R., « A change detection model based on neighborhood correlation image analysis and decision tree classification », *Remote Sensing of Environment*, vol. 99, p. 326–340, 2005.

[ING 03a] INGLADA J., « Change detection on SAR images by using a parametric estimation of the Kullback-Leibler divergence », *Proc of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Toulouse, France, July, 21-25 2003.

[ING 03b] INGLADA J., FAVARD J.-C., YESOU H., CLANDILLON S., BESTAULT C., « Lava flow mapping during the Nyiragongo January, 2002 eruption over the city of Goma (D.R. Congo) in the frame of the international charter space and major disasters », *Proc of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, vol. 3, p. 1540–1542, juillet , 21–25, 2003.

[ING 04] INGLADA J., GIROS A., « On the possibility of automatic multisensor image registration », *IEEE Trans. Geosci. Remote Sensing*, vol. 42, n°10, p. 2104–2120, octobre 2004.

[ING 07a] INGLADA J., MERCIER G., « A new statistical similarity measure for change detection in multitemporal SAR images and its extension to multiscale change analysis », *IEEE Trans. Geosci. Remote Sensing*, vol. 45, n°5, p. 1432–1446, mai 2007.

[ING 07b] INGLADA J., MICHEL J., « Spatial reasoning and multiscale segmentation for object recognition in HR optical remote sensing images », *Proc of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, p. 4798–4801, juillet 2007.

[ITA 05] ITAMI F., WATANABE E., NISHIHARA A., « A change detection method for image sequences based on physical models », *IEICE Trans. Fundam. Electron. Commun. Comput. Sci.*, vol. E88-A(8), p. 2100–2107, 2005.

[JAK 01] JAKUBAUSKAS M., LEGATES D., KASTENS J., « Harmonic analysis of time-series AVHRR NDVI data », *Photogrammetric Engineering and Remote Sensing*, vol. 67, n°4, p. 461–470, 2001.

[JOH 94] JOHNSON N., KOTZ S., *Distribution in statistics : Continuous univariate distributions, Vol. 1 and 2*, John Wiley and Sons, New York, 1994.

[KAS 02] KASETKASEM T., VARSHNEY P., « An image change detection algorithm based on Markov random Field models », *IEEE Trans. Geosci. Remote Sensing*, vol. 40, p. 1815–1823, août 2002.

[KET 07] KETTERLIN A., GANÇARSKI P., « Sequence similarity and multi-date image segmentation », *Proc. of the IEEE Int. Workshop on Analysis of Multi-temporal Remote Sensing Images (MultiTemp)*, juillet 2007.

[KOH 97] KOHONEN T., *Self-organizing maps*, vol. 30 de *Springer series in Informatioin Sciensces*, Springer, second edition édition, 1997.

[KUG 75] KUGLIN C., HINES D., « The phase correlation image alignment method », *Proc. IEEE Conf. Cybernetics Soc.*, p. 163–165, 1975.

[KUL 51] KULLBACK S., LEIBLER R., « On information and sufficiency », *Annals of Mathematical Statistics*, vol. 22, n°1, p. 79–86, mars 1951.

[LAM 97] LAMBIN E., EHRLICH D., « Land-cover changes in sub-saharan Africa (1982-1991) : Application of a change index based on remotely sensed surface temperature and vegetation indices at a continental scale », *Remote Sensing of Environment*, vol. 61, n°2, p. 181–200, 1997.

[LAS 06] LASAPONARA R., « Estimating interannual variations in vegetated areas of sardinia island using SPOT/VEGETATION NDVI temporal series », *IEEE Geosci. Remote Sensing Letters*, vol. 3, n°4, p. 481–483, octobre 2006.

[LEC 05] LECERF R., CORPETTI T., HUBERT-MOY L., DUBREUIL V., « Monitoring land use and land cover changes in oceanic and fragmented landscapes with reconstructed MODIS time series », *Int. Workshop on the Analysis of Multi-Temporal Remote Sensing Images*, p. 195–199, 2005.

[LI 01] LI L., LEUNG M., « Robust change detection by fusing intensity and texture differences », *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'01)*, 2001.

[LI 04] LI J., NARAYANAN R., « Integrated spectral and spatial information mining in remote sensing imagery », *IEEE Trans. Geosci. Remote Sensing*, vol. 42, n°3, p. 673–686, mars 2004.

[LIN 99] LIN J., SAITO N., LEVINE R., Edgeworth approximation of the Kullback-Leibler distance towards problems in Image Analysis, Rapport, University of California, Davis, 1999, http ://www.math.ucdavis.edu/˜saito.

[LIU 05] LIU Q., SCLABASSI R., SUN M., « An application of MAP-MRF to change detection in image sequence based on mean field theory », *EURASIP Journal on Applied Signal Processing*, vol. 13, p. 1956–1968, 2005.

[LOM 02] LOMBARDO P., PELLIZZERI T., « Maximum likelihood signal processing techniques to detect a step pattern of change in multitemporal SAR images », *IEEE Trans. Geosci. Remote Sensing*, vol. 40, n°4, p. 853–870, avril 2002.

[LOP 09] LOPEZ-QUIROZ P., DOIN M.-P., TUPIN F., BRIOLE P., NICOLAS J.-M., « Time series analysis of Mexico city subsidence constrained by radar interferometry », *Journal of Applied Geophysics*, 2009.

[LOV 00] LOVELAND T., REED B., BROWN J., OHLEN D., ZHU Z., YANG L., MERCHANT J., « Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data », *Int. Journal of Remote Sensing*, vol. 21, n°6–7, p. 1303–1330, 2000.

[LU 04] LU D., MAUSEL P., BRONDÍZIO E., MORAN E., « Change detection techniques », *Int. Journal of Remote Sensing*, vol. 25, n°12, p. 2365–2407, juin 2004.

[MAI 08] MAITRE H., ET AL., *Processing of Synthetic Aperture Radar (SAR) Images*, Wiley, 2008.

[MCC 87] MCCULLAGH P., *Tensor methods in statistics*, Chapman and Hall, London, 1987.

[MEL 06] MELGANI F., « Contextual reconstruction of cloud-contaminated multitemporal multispectral images », *IEEE Trans. Geosci. Remote Sensing*, vol. 44, n°2, p. 442–455, février 2006.

[MER 06] MERCIER G., DERRODE S., PIECZYNSKI W., NICOLAS J.-M., JOANNIC-CHARDIN A., INGLADA J., « Copula-based stochastic kernels for abrupt change detection », *Proc of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2006.

[MIL 88] MILNE A., « Change direction analysis using landsat imagery : a review of methodology », *Int. Geoscience and Remote Sensing Symposium (IGARSS'88)*, p. 541–544, 1988.

[MOO 01] MOODY A., JOHNSON D. M., « Land-surface phenologies from AVHRR using the discrete Fourier transform », *Remote Sensing of Environment*, vol. 75, n°3, p. 305–323, 2001.

[MOO 05] MOODY E., KING M., PLATNICK S., SCHAAF C., FENG G., « Spatially complete global spectral surface albedos : value-added datasets derived from Terra MODIS land products », *IEEE Trans. Geosci. Remote Sensing*, vol. 43, n°1, p. 144–158, 2005.

[NEM 04] NEMMOUR H., CHIBANI Y., « Combining comparative and simultaneous analysis approaches based on fuzzy integration for change detection », *XXth ISPRS Congress, Geo-Imagery Bridging Continents*, Istanbul, Turkey, July, 12-23 2004.

[NIE 98] NIELSEN A., CONRADSEN K., SIMPSON J., « Multivariate alteration detection (MAD) and MAF post processing in multi-spectral bi-temporal image data : new approaches to change detection studies », *Remote Sensing of Environment*, vol. 64, p. 1–19, 1998.

[PAC 07] PACIFICI F., DEL FRATE F., SOLIMINI C., EMERY W., « An innovative neural-net method to detect temporal changes in high-resolution optical satellite imagery », *IEEE Trans. Geosci. Remote Sensing*, vol. 45, n°9, p. 2940–2952, septembre 2007.

[PED 03] PEDERSEN R., JONSSON S., ARNADOTTIR T., SIGMUNDSSON F., FEIGL K. L., « Fault slip distribution derived from joint inversion of InSAR and GPS measurements of

two June 2000 Mw 6.5 earthquakes in South Iceland », *Earth Planet. Sci. Lett.*, , n°213, p. 487–502, 2003.

[PLI 00] PLISNIER P., SERNEELS S., LAMBIN F., « Impact of ENSO on East African ecosystems : a multivariate analysis based on climate and remote sensing data », *Global Ecology & Biogeography*, vol. 9, p. 481–497, 2000.

[RAU 09] RAUCOULES D., BOURGINE B., DE MICHELE M., COZANNET G. L., CLOSSET L., BREMMER C., VELDKAMP H., TRAGHEIM D., BATESON L., CROSETTO M., AGUDO M., ENGDAHL M., « Validation and Intercomparison of Persistent Scatterers Interferometry : PSIC4 project results », *Journal of Applied Geophysics*, 2009.

[REE 94] REED B., BROWN J., VANDERZEE D., LOVELAND T., MERCHANT J., OHLEN D., « Measuring phenological variability from satellite imagery », *J. of Vegetation Science*, vol. 5, p. 703–714, 1994.

[REI 07] REICHSTEIN M., CIAIS P., PAPALE D., VALENTINI R., RUNNING S., VIOVY N., CRAMER W., GRANIER A., OGEE J., ALLARD V., AUBINET M., BERNHOFER C., BUCHMANN N., CARRARA A., GRUENWALD T., HEINESCH B., KNOHL A., LOUSTAU D., MANCA G., MATTEUCCI G., MIGLIETTA F., OURCIVAL J., PILEGAARD K., PUMPANEN J., RAMBAL S., SCHAPHOFF S., SEUFERT G., SOUSSANA J.-F., SANZ M.-J., VESALA T., HEIMANN M., « Reduction of ecosystem productivity and respiration during the European summer 2003 heatwave : a joint flux tower, remote sensing and modelling analysis », *Global Change Biology*, vol. 13, n°3, p. 634–651, 2007.

[RIG 93] RIGNOT E., VAN ZYL J., « Change detection techniques for ERS-1 SAR data », *IEEE Trans. Geosci. Remote Sensing*, vol. 31, n°4, p. 896–906, juillet 1993.

[ROB 98] ROBERTS D., GARDNER M., CHURCH R., USTIN S., SCHEER G., GREEN R., « Mapping chaparral in the Santa Monica mountains using multiple endmember spectral mixture models », *Remote Sensing of Environment*, vol. 65, p. 267–279, 1998.

[ROB 07] ROBIN A., MOISAN L., LE HÉGARAT-MASCLE S., « An unsupervised approach for subpixelic land-cover change detection », *Proc. of the IEEE Int. Workshop on Analysis of Multi-temporal Remote Sensing Images (MultiTemp)*, juillet 2007.

[ROE 00] ROERINK G., MENENTI M., VERHOEF W., « Reconstructing cloud free NDVI composites using Fourier analysis of time series », *Int. Journal of Remote Sensing*, vol. 21, n°9, p. 1911–1917, 2000.

[RUA 02] RUAN S., MORETTI B., FADILI J., BLOYET D., « Fuzzy Markovian segmentation in application of magnetic resonance images », *Computer Vision and Image Understanding*, vol. 85, p. 54–69, 2002.

[SAK 05] SAKAMOTO T., YOKOZAWA M., TORITANI H., SHIBAYAMA M., ISHITSUKA N., OHNO H., « A crop phenology detection method using time-series MODIS data », *Remote Sensing of Environment*, vol. 96, n°3-4, p. 366–374, 2005.

[SAL 97] SALZENSTEIN F., PIECZYNSKI W., « Parameter estimation in hidden fuzzy Markov random fields and image segmentation », *Graphical Models and Image Processing*, vol. 59, n°4, p. 205–220, 1997.

[SAL 06] SALZENSTEIN F., COLLET C., « Fuzzy Markov random fields versus chains for multispectral image segmentation », *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 28, n°11, p. 1753–1767, 2006.

[SCH 03] SCHIAVON G., DEL FRATE F., SOLIMINI C., « High resolution multi-spectral analysis of urban areas with QuickBird imagery and synergy with ERS data », *Proc of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Toulouse, France, July, 21-25 2003.

[SER 06] SERPICO S. B., MOSER G., « Weight parameter optimization by the Ho-Kashyap algorithm in MRF models for supervised image classification », *IEEE Trans. Geosci. Remote Sensing*, vol. 44, n°12, p. 3695–3705, 2006.

[SIN 89] SINGH A., « Digital change detection techniques using remotely-sensed data », *Int. Journal of Remote Sensing*, vol. 10, p. 989–1003, 1989.

[SKA 92] SKARBEK W., « *Generalized Hilbert scan in image printing* », Chapitre Theoretical Foundations of Computer Vision, Akademie Verlag, Berlin, Germany, R. Klette and W. G. Kropetsh édition, 1992.

[STU 91] STUART A., ORD J., *Kendall's advanced theory of Statistics*, Edward Arnold, 5th édition, 1991.

[TAR 96] TARAYRE H., MASSONNET D., « Atmospheric propagation heterogeneities revealed by ERS-1 interferometry », *Geophysical Research Letters*, vol. 23, n°9, p. 989-992, 1996.

[TRO 98] TROUVÉ E., NICOLAS J. M., MAÎTRE H., « Improving Phase Unwrapping Techniques by the Use of Local Frequency », *IEEE Trans. Geosci. and Remote Sensing*, vol. 36, n°6, p. 1963-1972, novembre 1998.

[TRO 07] TROUVÉ E., VASILE G., GAY M., BOMBRUN L., GRUSSENMEYER P., LANDES T., NICOLAS J., BOLON P., PETILLOT I., JULEA A., VALET L., CHANUSSOT J., KOEHL M., « Combining Airborne Photographs and Spaceborne SAR Data to Monitor Temperate Glaciers. Potentials and Limits », *IEEE Trans. Geosci. and Remote Sensing*, vol. 45, n°4, p. 905–924, avril 2007.

[ULA 86] ULABY F. T., KOUYATE F., BRISCO B., LEE WILLIAMS T. H., « Textural information in SAR images », *IEEE Trans. Geosci. Remote Sensing*, vol. 24, n°2, p. 235–245, mars 1986.

[UTO 07] UTO K., KOSUGI Y., « Self-organizing property of nonlinear mapping for change detection », *Proc of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Barcelona, Spain, July, 23-28 2007.

[VAS 06] VASILE G., TROUVÉ E., LEE J.-S., BUZULOIU V., « Intensity-Driven-Adaptive-Neighborhood Technique for Polarimetric and Interferometric SAR Parameters Estimation », *IEEE Trans. Geosci. and Remote Sensing*, vol. 44, n°6, p. 1609-1621, 2006.

[VEL 02] VELLOSO M., DE SOUZA F., SIMOES M., « Improved radiometric normalization for land cover change detection : an automated relative correction with artificial neural network », *Proc of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, vol. 6, p. 3435–3437, juin 2002.

[VIV 92] VIVOY N., ARINO O., BELWARD A., « The Best Index Slop Extraction (BISE) : A method for reducing noise in NDVI time-series », *Int. Journal of Remote Sensing*, vol. 13, n°8, p. 1585–1590, 1992.

[WAL 04] WALTER V., « Object-based classification of remote sensing data for change detection », *ISPRS Journal of Photogrammetry & Remote Sensing*, vol. 58, p. 225–238, janvier 2004.

[WAR 81] WARD K., « Compound representation of high resolution sea clutter », *Electronics Letters*, vol. 17, p. 561–565, 1981.

[WHI 97] WHITE M., THOMTON P., RUNNING S., « A continental phenology model for monitoring vegetation responses to interannual climatic variability », *Global biogeochemical cycles*, vol. 11, n°2, p. 217–234, 1997.

[WIE 97] WIEMKER R., « An iterative spectral-spatial Bayesian labelling approach for unsupervised robust change detection on remotely sensed multispectral imagery », *Proc. of the 7th Int. Conf. on Computer Analysis of Images and Patterns (CAIP'97)*, p. 263–271, 1997.

[XIA 03] XIANG G., HUETE A., DIDAN K., « Multisensor comparisons and validation of MODIS vegetation indices at the semiarid Jornada experimental range », *IEEE Trans. Geosci. Remote Sensing*, vol. 41, n°10, p. 2368–2381, octobre 2003.

[XUE 04] XUE C., XIAOWEN L., JIANWEN M., « Urban change detection based on self-organizing feature map neural network », *Proc of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, vol. 5, p. 3428–3431, 2004.

[YOO 03] YOON G., BO YUN Y., PARK J., « Change vector analysis : detecting of areas associated with foood using Landsat TM », *Proc of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, vol. 5, p. 3386–3388, 2003.

[ZEB 92] ZEBKER H., VILLASENOR J., « Decorrelation in Interferometric Radar Echoes », *IEEE Trans. Geosci. and Remote Sensing*, vol. 30, n°5, p. 950–959, septembre 1992.

[ZHA 04] ZHANG L., LIAO M., WANG Y., LU L., WANG Y., « Robust approach to the MAD change detection method », *Remote Sensing for Environmental Monitoring, GIS Applications, and Geology IV*, vol. 5574, p. 184–193, octobre 2004.

[ZHI 06] ZHIGAO Y., QIANQING Q., QIFENG Z., « Change detection in high spatial resolution images based on support vector machines », *Proc of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, juillet 2006.

Chapitre 6

Bayesian Approach to Linear Spectral Mixture Analysis

In several application domains such as remote sensing and analytical chemistry, it is frequently needed to process multivariate data sets obtained by spectral measurements on multicomponent materials. The processing of the resulting data aims at identifying the different components of the materials and assessing their abundances. Assuming the observed data as mixture spectra, these goals can be formulated as a blind source separation problem [JUT 91, COM 91], where, in the case of optical spectroscopy, the linear instantaneous mixture model holds according to the Beer-Lamber law [PFE 51] (also termed as the Beer's law or the Beer-Lambert-Bouguer law [RIC 94]). The component spectra are identified as the source signals and the abundances as the mixing coefficients. The spectral mixture analysis problem is also encountered in multivariate image processing [CHA 07, MOU 08] where the sources correspond to endmember spectra and the mixing coefficients are linked to the abundance maps in the imaged scene. In this chapter, we present a Bayesian method [MOU 06, **?**] allowing to process a set of measured multivariate data with the aim to estimate the component spectra and the abundances by using Bayesian estimation and Markov chain Monte Carlo methods.

6.1. The Spectral Mixing Model

According to the linear mixing model, the spectral mixture data are interpreted as a weighted sum of the unknown component spectra. The mixing model assumes that

Chapitre rédigé par S. Moussaoui, D. Brie, C. Carteret.

m measured spectra $\left\{ X_{(i,k)}, k = 1, ..., n \right\}_{i=1}^{m}$ are linear combinations of p unknown pure component spectra $\left\{ S_{(j,k)}, k = 1, ..., n \right\}_{j=1}^{p}$. This mixing model is expressed as

$$X_{(i,k)} = \sum_{j=1}^{p} A_{(i,j)} \, S_{(j,k)} \, + \, E_{(i,k)}, \qquad (6.1)$$

where $i = 1, \ldots, m$ and $j = 1, \ldots, p$, respectively, index the measured spectra and the unknown pure component spectra and the index k corresponds to the spectral variable $\{\lambda_k, k = 1, ..., n\}$. Each mixing coefficient $A_{(i,j)}$ is proportional to the concentration of the j-th pure component in the i-th mixture. The additive noise terms $\left\{ E_{(i,k)}, k = 1, ..., n \right\}_{i=1}^{m}$ represent the measurement errors and model imperfections. The main constraint in this mixing model model is the non-negativity of both the source signals and the mixing coefficients :

$$A_{(i,j)} \geqslant 0, \quad \forall i, j \quad \text{and} \quad S_{(j,k)} \geqslant 0, \quad \forall j, k \qquad (6.2)$$

Using matrix notations this model is written as

$$\boldsymbol{X} = \boldsymbol{A} \, \boldsymbol{S} + \boldsymbol{E}, \qquad (6.3)$$

where the row vectors of the $(m \times n)$ data matrix $\boldsymbol{X}$ contain the m measured spectra, $\boldsymbol{A}$ is the $(m \times p)$ mixing matrix, with its column vectors representing the mixing coefficient of each pure component. $\boldsymbol{S}$ is the $(p \times n)$ matrix, with its row vectors containing the p pure component spectra and $\boldsymbol{E}$ is the $(m \times n)$ additive noise matrix.

By assuming a known number of components, the mixture analysis goal is to jointly estimate the pure component spectra and the mixing coefficient. For a qualitative analysis, pure spectra are necessary to identify the chemical composition of the substances and the mixing coefficients are used, in a quantitative analysis, to evaluate the concentrations of each component. In the chemometrics community, this problem is termed by *curve resolution* [LAW 71], *mixture analysis* [WIN 90] and *factor analysis* [MAL 00] while in signal and image processing field this problem is called *blind source separation* [JUT 91]. In the sequel, the problem is referred to as spectral mixture analysis and we implicitly assume a linear mixing.

This problem can be solved using Bayesian estimation theory, which consists in assigning prior distributions to account for any available knowledge on component spectra and mixing coefficients. The probabilistic formulation of the problem enables the use of stochastic simulation tools to draw samples from the posterior distribution which are used to evaluate various type of estimators. Section 6.2 of this chapter gives a short overview on existing spectral mixture analysis methods which are mainly presented regarding the underlying assumptions and estimation algorithms, thus allowing to point out some of their limitations that may handled by the Bayesian estimation and Markov Chain Monte Carlo (MCMC) methods [ROB 99, CHE 00]. The probabilistic

formulation of the problem is detailed and a discussion on its relation with the *positive matrix factorization* (PMF) method [PAA 94] is given. The Bayesian method is illustrated in section 6.4 and the behavior of the method is discussed regarding to the noise level, the source correlation and the presence of a background.

6.2. Overview on Linear Spectral Mixture Analysis

In this section a short overview of spectral mixture analysis methods is given. A more complete state of the art may be found in [JIA 04]. A mixture analysis algorithm is defined firstly, by the assumptions made on the pure component spectra and their abundance profiles which lead to an objective function, and secondly by the optimization algorithm which is used to optimize this objective function. The main constraint in spectral mixture data analysis is the non-negativity of both the pure component spectra and the abundances. However, only accounting for this constraint does not lead to unique solution [LAW 71]. Thus, additional constraints or assumptions are required to select a particular solution. According to the way these constraints and assumptions are introduced to the model, the available mixture analysis methods can be classified into two main categories. The first one, termed as algebraic methods, is based on Lawton-Sylvestre approach [LAW 71], termed by *multivariate curve resolution*, while the second is based on a *non-negative least squares estimation* [TAU 93a, LEE 99] .

6.2.1. *Algebraic Methods*

Methods of this approach firstly decompose the data matrix X, using principal component analysis (PCA), and then finds a linear transformation T that transforms the principal components V and their weight matrix U into non-negative estimates of pure spectra and mixing coefficients. However, since only accounting for non-negativity does not ensure the uniqueness of the solution, this approach leads to a set of admissible solutions [OHT 73, SAS 83, BOR 85, HEN 90]. In order to reduce this set [SAS 84, SAS 87, SAS 89] suggest to add further constraints, in addition to the non-negativity, and propose to search a linear transformation by minimizing a compound criterion, in which a first part penalizes negative estimates of the pure spectra and mixing coefficients, and the second part uses an entropic cost function to make the estimated spectra smoother and mutually independent. The optimization of the whole objective function is performed by using the simplex method. However, in spite of the theoretical uniqueness of the solution by considering the additional *a priori* constraints, this method may converge to local or false minimums. This is due to the fact that the convexity of the criterion is not proven and the global criterion shape is highly depending on the regularization parameters that are specified manually. Recently, this method have been revisited in [YIN 98, WID 03] where a simulated annealing optimization algorithm has been used, but this algorithm may also converge to

a local minimum. In the signal processing community, a similar approach to Lawton-Sylvestre method, based on statistical independence of pure spectra and called *independent component analysis* [JUT 91, COM 94, HYV 01, CIC 02], is used for the analysis of mixtures. To get the mutually independent signals, the transformation matrix is reduced to a unitary rotation matrix. Assuming the statistical independence, ensures the uniqueness of the solution, but it does not guarantee its non-negativity. To handle this problem, the *non-negative independent component analysis* (NNICA) [PLU 03], finds a transformation matrix by optimizing an objective function that enforces the non-negativity of the estimated pure spectra. However, this method assumes orthogonal pure spectra and does not account explicitly the mixing coefficient non-negativity.

6.2.2. *Non-negative Least Squares Estimation*

This approach performs constrained least squares estimation and there are mainly two methods : *alternating least squares* (ALS) [TAU 93b, TAU 93a] and *non-negative matrix factorization* (NMF) [LEE 99]. ALS performs a decomposition of the mixture by minimizing at each iteration the least squares criterion under the non-negativity constraint [LAW 74, BRO 97]. The NMF method minimizes the least squares criterion by using a gradient descent algorithm over this objective function and updates iteratively concentrations and pure spectra using a particular multiplicative learning rule that ensures the estimates to remain non-negative. The key point is that the solution using only non-negativity is not unique, therefore the results provided by ALS and NMF methods depend on their initializations. In practice, the NMF method is randomly initialized [SAJ 03], while the ALS method is initialized by the results obtained with a non-constrained decomposition method such as principal component analysis (PCA), factor analysis algorithms [GEM 84b, MAE 88, SAN 97], or using pure variable detection methods such as *simple-to-use interactive self modeling mixture analysis* (SIMPLISMA) [WIN 91] and orthogonal projection approach (OPA) [SAN 96]. More recently ICA methods are also used [NUZ 98] as an initialization method. Similarly to the algebraic methods, additional constraints such as closure, unimodality, selectivity may be added to reduce the set of admissible solutions [De 97, BRO 98, TAU 00]. These additional constraints are accounted through either a penalized least squares estimation or a constrained estimation. In the penalized least square methods, a regularization criterion is added to the weighted mean squares criterion in order to select a particular solution fulfilling some assumptions defined *a priori*. This is the basis of methods such as *positive matrix factorization* (PMF) [PAA 94, PAA 97] or *non-negative sparse coding* (NNSC) [HOY 04]. Unlike the constrained least squares methods, the PMF approach leads to an unconstrained optimization problem and ensures the uniqueness of the solution for a fixed set of regularization parameters. Coming from the fact that penalized least squares estimation methods can be formulated in a Bayesian framework [DEM 89], we propose a fully Bayesian approach to address the inverse problem of linear spectral mixture analysis.

6.3. The Bayesian Separation Approach

Spectral mixture analysis is an ill-posed inverse problem since the solution is not unique. Therefore, to get a preferable solution one has to use any available knowledge or additional assumptions. Therefore, this inverse problem can be successfully addressed using Bayesian estimation theory by encoding the available information through the assignment of *a priori* distributions on the pure component spectra and concentration profiles. The formulation of source separation and factor analysis using Bayesian estimation theory has been suggested recently in [ROB 98, KNU 99, MOH 99, ROW 03] and used in signal and image processing problems [SNO 06, FÉV 06]. However, its application to the separation of spectral mixture data has only received a few attention [OCH 99, MIS 01, MOU 06]. Since only accounting for the non-negativity constraint, spectral mixture analysis still remains ill-posed, one has to use additional assumptions on the component spectra and abundances to get a unique solution. Indeed, we introduce an assumption on their statistical distributions which aims at encoding the non-negativity and the pure spectra sparsity. Before going further, let us recall the Bayesian approach to mixture analysis.

6.3.1. *Bayesian Approach*

The main idea of the Bayesian approach to mixture analysis is to encode the available knowledge on the pure spectra and concentration profiles through the assignment of prior distributions $p(\boldsymbol{S})$ and $p(\boldsymbol{A})$, which according to Bayes' theorem and using the likelihood $p(\boldsymbol{X}|\boldsymbol{S}, \boldsymbol{A})$ leads to the *a posteriori* density

$$p(\boldsymbol{S}, \boldsymbol{A}|\boldsymbol{X}) = p(\boldsymbol{X}|\boldsymbol{S}, \boldsymbol{A}) \times p(\boldsymbol{S}) \times p(\boldsymbol{A}) \div p(\boldsymbol{X}), \qquad (6.4)$$

where the independence between $\boldsymbol{A}$ and $\boldsymbol{S}$ is assumed. Since $p(\boldsymbol{X})$ is a normalization constant, one can write

$$p(\boldsymbol{S}, \boldsymbol{A}|\boldsymbol{X}) \propto p(\boldsymbol{X}|\boldsymbol{S}, \boldsymbol{A}) \times p(\boldsymbol{A}) \times p(\boldsymbol{A}). \qquad (6.5)$$

This *a posteriori* density combines explicitly the *a priori* knowledge and assumptions on the pure component spectra and concentrations with the information coming from the measured data.

To get estimates of pure component spectra and concentrations, various Bayesian estimators [ROB 01] can be applied

– the *joint maximum a posteriori* (JMAP) estimates are obtained by

$$\left(\hat{\boldsymbol{S}}_{JMAP}, \hat{\boldsymbol{A}}_{JMAP}\right) = \arg \max_{\boldsymbol{S}, \boldsymbol{A}} p(\boldsymbol{S}, \boldsymbol{A}|\boldsymbol{X}). \qquad (6.6)$$

This maximisation is equivalent to a joint minimisation of the corresponding criterion, defined by

$$J(\boldsymbol{S}, \boldsymbol{A}) = -\log p(\boldsymbol{S}, \boldsymbol{A}|\boldsymbol{X}), \qquad (6.7)$$

and depends not only on the measured data but also on the prior distributions of pure spectra and concentration profiles

$$J(\boldsymbol{S}, \boldsymbol{A}) = -\log p(\boldsymbol{X}|\boldsymbol{S}, \boldsymbol{A}) - \log p(\boldsymbol{S}) - \log p(\boldsymbol{A}). \qquad (6.8)$$

Each prior model leads to a particular regularization criterion. This aspect can be seen as an implicit penalty function synthesis through a probabilistic formulation of the prior knowledge. This formulation is known as a Bayesian interpretation of penalized least squares estimation methods.

– the *marginal maximum a posteriori* (MMAP) estimates are obtained by integrating (marginalizing) with respect to $\boldsymbol{S}$ or $\boldsymbol{A}$ and maximizing the resulting posterior marginal distribution $p(\boldsymbol{A}|\boldsymbol{X})$ or $p(\boldsymbol{S}|\boldsymbol{X})$.

– the *marginal posterior mean* (MPM) estimates are obtained from the mean of the marginal posterior distributions $p(\boldsymbol{A}|\boldsymbol{X})$ and $p(\boldsymbol{S}|\boldsymbol{X})$. In practice, the marginalization is generally performed using MCMC methods and the posterior means are calculated from the simulated Markov chains.

In the following, the proposed Bayesian inference for the analysis of spectral mixtures and its related estimation algorithm are presented. The mathematical computation details can be found in [MOU 06].

6.3.2. *Noise Distribution and Likelihood*

Each noise sequence $\left\{ E_{(i,k)}, k = 1, ..., n \right\}$ is assumed Gaussian, zero mean, independent and identically distributed with variance σ_i^2. Moreover, all the noise sequences are assumed mutually independent. Thus,

$$p\left(\boldsymbol{E}|\boldsymbol{\theta}_1\right) = \prod_{i=1}^{m} \prod_{k=1}^{n} \mathcal{N}\left(E_{(i,k)}; 0, \sigma_i^2\right), \qquad (6.9)$$

where $\boldsymbol{\theta}_1 = \left[\sigma_1^2, \ldots, \sigma_m^2\right]^T$ and $\mathcal{N}(x; \mu, \nu)$ denotes the normal distribution for the variable x with mean μ and variance ν

$$\mathcal{N}(x; \mu, \nu) = \frac{1}{\sqrt{2\pi\nu}} \exp\left[-\frac{1}{2\nu}(x - \mu)^2\right]. \qquad (6.10)$$

Using this noise distribution and the mixing model, the expression of the likelihood is obtained as

$$p\left(\boldsymbol{X}|\boldsymbol{S}, \boldsymbol{A}, \boldsymbol{\theta}_1\right) \propto \prod_{i=1}^{m} \prod_{k=1}^{n} \left(\frac{1}{\sigma_i}\right)^n \exp\left[-\frac{1}{2\sigma_i^2}\left(X_{(i,k)} - [\boldsymbol{A}\boldsymbol{S}]_{(i,k)}\right)^2\right]. \qquad (6.11)$$

6.3.3. *Prior Distributions of Pure Spectra and Mixing Coefficients*

The pure component spectra are considered mutually independent and identically distributed and each pure spectrum $\{S_{(j,k)}, k = 1, ..., n\}$ is assumed distributed as Gamma distribution of parameters (α_j, β_j). These parameters are considered constant for each spectrum over all the spectral band but may differ from one pure spectrum to another. Thus, the prior distribution of the pure spectra is

$$p(\boldsymbol{S}|\boldsymbol{\theta}_2) = \prod_{j=1}^{p} \prod_{k=1}^{n} \mathcal{G}(S_{(j,k)}; \alpha_j, \beta_j), \qquad (6.12)$$

where $\mathcal{G}(z; \alpha, \beta)$ represents the Gamma density for the variable z with parameters α and β. The vector $\boldsymbol{\theta}_2 = [\alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_p]^T$ contains the parameters of all the prior distributions.

The Gamma density is expressed as

$$\mathcal{G}(z; \alpha, \beta) = \begin{cases} \dfrac{\beta^{\alpha}}{\Gamma(\alpha)} \, x^{\alpha-1} \, \exp\left[-\beta x\right] & \text{for } x \geq 0, \\ 0 & \text{for } x < 0, \end{cases} \qquad (6.13)$$

where $\Gamma(\alpha)$ is the Gamma function. The choice of this distribution is motivated by the fact that :

– it takes into account explicitly the non-negativity since $p(z < 0) = 0$;

– due to its two parameters α and β, this density can be used to fit various distributions and has the Khi-2, exponential distributions as particular cases [STU 94]. Figure 6.1, illustrates three typical shapes of the Gamma density. Case (a) is well adapted to encode sparsity (concentration of the distribution in the neighborhood of zero), while case (b) corresponds, for example, to the case of a sparse pure spectrum superimposed on a background, typically consisting in the tail of a wide peak. Case (c) shows that the Gamma distribution can also lead to a noninformative positive prior.

Each j-th column of the mixing matrix is also assumed distributed as a Gamma density of parameters (γ_j, δ_j). These parameters are considered constant for each concentration profile but may differ from one profile to another. The prior distribution of mixing coefficients is then given as

$$p(\boldsymbol{A}|\boldsymbol{\theta}_3) = \prod_{i=1}^{m} \prod_{j=1}^{p} \mathcal{G}(A_{(i,j)}; \gamma_j, \delta_j), \qquad (6.14)$$

where $\boldsymbol{\theta}_3 = [\gamma_1, \ldots, \gamma_p, \delta_1, \ldots, \delta_p]^T$.

Figure 6.1. *Illustrations of the typical Gamma distribution shapes according to the values of its parameters α and β.*

6.3.4. *Posterior Density and Resulting Criterion*

Noting $\boldsymbol{\theta} = [\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3]$ and according to equation (6.5) one can deduce the expression of the posterior law

$$p(\boldsymbol{S}, \boldsymbol{A}|\boldsymbol{X}, \boldsymbol{\theta}) \propto \prod_{k=1}^{n}\prod_{i=1}^{m} \mathcal{N}(E_{(i,k)}; 0, \sigma_i^2)$$
$$\times \prod_{k=1}^{n}\prod_{j=1}^{p} \mathcal{G}(S_{(j,k)}; \alpha_j, \beta_j) \times \prod_{i=1}^{m}\prod_{j=1}^{p} \mathcal{G}(A_{(i,j)}; \gamma_j, \delta_j). \quad (6.15)$$

The corresponding criterion takes the form

$$J_1(\boldsymbol{S}, \boldsymbol{A}|\boldsymbol{\theta}) = Q(\boldsymbol{S}, \boldsymbol{A}|\boldsymbol{\theta}_1) + R_S(\boldsymbol{S}|\boldsymbol{\theta}_2) + R_A(\boldsymbol{A}|\boldsymbol{\theta}_3), \quad (6.16)$$

where the terms Q, R_S, and R_A are given by

$$
\begin{cases}
Q(\boldsymbol{S}, \boldsymbol{A}|\boldsymbol{\theta}_1) = \displaystyle\sum_{i=1}^{m}\sum_{k=1}^{n}\frac{1}{2\sigma_i^2}\left(X_{(i,k)} - [\boldsymbol{AS}]_{(i,k)}\right)^2, \\[2ex]
R_S(\boldsymbol{S}|\boldsymbol{\theta}_2) = \displaystyle\sum_{j=1}^{p}\sum_{k=1}^{n}(1-\alpha_j)\log S_{(j,k)} + \sum_{j=1}^{p}\sum_{k=1}^{n}\beta_j\, S_{(j,k)}, \\[2ex]
R_A(\boldsymbol{A}|\boldsymbol{\theta}_3) = \displaystyle\sum_{j=1}^{p}\sum_{i=1}^{m}(1-\gamma_j)\log A_{(i,j)} + \sum_{j=1}^{p}\sum_{i=1}^{m}\delta_j\, A_{(i,j)}.
\end{cases} \tag{6.17}
$$

The first term Q can be seen as a data fitting measure, while the last two terms are regularization criteria that prevent negative values of $\boldsymbol{A}$ and $\boldsymbol{S}$.

6.3.5. *Bayesian Formulation of the PMF Criterion*

In the PMF method [PAA 97], the minimized criterion is given by

$$
\begin{aligned}
J_2(\boldsymbol{S}, \boldsymbol{A}|\boldsymbol{\theta}) = \sum_{i=1}^{m}\sum_{k=1}^{n}\frac{1}{2\sigma_{(i,k)}^2}&\left(X_{(i,k)} - [\boldsymbol{AS}]_{(i,k)}\right)^2 \\
&- \alpha\sum_{j=1}^{p}\sum_{k=1}^{n}\log S_{(j,k)} + \beta\sum_{j=1}^{p}\sum_{k=1}^{n}S_{(j,k)}^2 \\
&- \gamma\sum_{j=1}^{p}\sum_{i=1}^{m}\log A_{(i,j)} + \delta\sum_{j=1}^{p}\sum_{i=1}^{m}A_{(i,j)}^2. \quad (6.18)
\end{aligned}
$$

where $(\alpha, \beta, \gamma, \delta)$ are positive regularization parameters. This regularized least squares method uses on pure spectra and concentrations an *a priori* cost function of the form

$$
f(z) = -\alpha\log z + \beta z^2, \text{ with } z \geq 0 \tag{6.19}
$$

and corresponds to assign to the variable z a prior distribution defined as

$$
p(z|\alpha, \beta) = K(\alpha, \beta)\exp\left[-f(z)\right] = K(\alpha, \beta)\,z^\alpha\,\exp\left[-\beta z^2\right] \quad \text{with } z \geq 0, \tag{6.20}
$$

where $K(\alpha, \beta) = 2\beta^{(\alpha+1)/2}/\Gamma((\alpha+1)/2)$. It can be noted that this distribution corresponds to assuming a Gamma prior distribution on z^2 with parameters $((\alpha+1)/2, \beta)$. Figure 6.2 shows three typical shapes of this distribution. One can particularly note a similarity with the Gamma distribution profiles shown in Figure 6.1.

To make a comparison between the PMF method and our proposed approach, the following discussion can be pointed out :

Figure 6.2. *Illustrations of the* a priori *distribution associated to the PMF method.*

(i) The case $\alpha = 0$, corresponds to assigning an *a priori* truncated Gaussian distribution. Contrarily to the Gamma distribution, when $\alpha > 0$ this distribution is not appropriate to encode the sparsity since its mode is not located at zero.

(ii) In the PMF method, the values of the hyperparameters should be fixed *a priori* and are considered equal for all the pure spectra and for all the concentrations, which is nothing but assuming that all the pure spectra and concentration profiles have the same distribution. In the proposed inference these parameters may differ for different components and due to the probabilistic formulation, all the hyperparameters are estimated jointly with the pure component spectra and concentration profiles.

(iii) Due to the non-convexity of the two criteria J_1 and J_2 when the hyperparameter estimation is added, deterministic optimization methods such as gradient or Newton-Raphson algorithms needs either a good staring point or multiple random initializations in order to reach the global minimum.

To handle this optimization problems, the hyperparameters are included in the Bayesian model and stochastic simulation tools are used to calculate the Bayesian estimators.

6.3.6. *Inference of the hyperparameters*

In practice, the parameters of the Gamma distributions as well as the noise variances are not known *a priori*. So, for an unsupervised processing, they must also considered in the inference. Thus, by using Bayes' theorem and assigning appropriate *a priori* distributions to these hyperparameters, the whole *a posteriori* distribution, including the hyperparameters, is expressed as

$$p(\boldsymbol{S}, \boldsymbol{A}, \boldsymbol{\theta}|\boldsymbol{X}) \propto p(\boldsymbol{S}, \boldsymbol{A}|\boldsymbol{X}, \boldsymbol{\theta}) \times p(\boldsymbol{\theta}). \qquad (6.21)$$

The joint estimation of the pure spectra, the mixing coefficients and the hyperparameters is then performed from this *a posteriori* distribution. However, this inference also needs the definition of appropriate priors for the hyperparameters, that is choosing *a priori* distributions allowing to set the range of admissible values of the hyperparameters, which is much easier than specifying a specific value of each parameter.

6.3.6.1. *Priors on the noise variances*

Each noise variance is positive, therefore an appropriate distribution is to assign an *a priori* Gamma distribution to $1/\sigma_i^2$ with parameters $(2, \epsilon)$ where ϵ is set to a small value (10^{-3}). This distribution is chosen because it encodes the positivity and leads to simpler computations of the *a posteriori* distribution $p(\boldsymbol{\theta}_1|\boldsymbol{S}, \boldsymbol{A}, \boldsymbol{X})$ [ROB 01].

6.3.6.2. *Priors on the Gamma distribution parameters*

The parameters of the Gamma distributions associated to the pure spectra and concentration profiles are assumed as *a priori* Gamma distributed with parameters $(2, \epsilon)$ where ϵ is set to a small value (10^{-3}). Similarly to the noise variances, this distribution is chosen since it encodes the positivity of the parameters and leads to simpler computations of the *a posteriori* distributions $p(\boldsymbol{\theta}_2|\boldsymbol{S})$ and $p(\boldsymbol{\theta}_3|\boldsymbol{A})$. Note also that these prior distributions are not very informative (quasi-uniform shape) since their mode is $1/\epsilon$ and their variance is equal to $2/\epsilon^2$.

6.3.7. *Estimation via Markov Chain Monte Carlo Methods*

Markov Chain Monte Carlo (MCMC) is a class of stochastic simulation tools for generating random variables from univariate or multivariate probability distribution functions [HAS 70, GEM 84a]. These methods are extensively documented in the statistical literature (see the books [GIL 99, ROB 99, CHE 00] and the references therein). The main objective of MCMC methods is to simulate a stationary ergodic Markov chain whose samples asymptotically follow the posterior density $p(\boldsymbol{S}, \boldsymbol{A}, \boldsymbol{\theta}|\boldsymbol{X})$.

Estimates of the pure component spectra and the mixing coefficients are then calculated from the simulated Markov chains. In the sequel, the main steps of the MCMC algorithm for sampling the posterior distribution (6.21) are given.

6.3.7.1. *Posterior Density Simulation using MCMC methods*

To simulate $p\left(\boldsymbol{S}, \boldsymbol{A}, \boldsymbol{\theta} | \boldsymbol{X}\right)$, the Gibbs sampler [GEM 84a] is used. At each new iteration r of this algorithm, the main steps consist in simulating

1. the pure spectra $\boldsymbol{S}^{(r+1)}$ from

$$p\left(\boldsymbol{S} | \boldsymbol{X}, \boldsymbol{A}^{(r)}, \boldsymbol{\theta}^{(r)}\right) \propto p\left(\boldsymbol{X} | \boldsymbol{S}, \boldsymbol{A}^{(r)}, \boldsymbol{\theta}^{(r)}\right) \times p\left(\boldsymbol{S} | \boldsymbol{\theta}^{(r)}\right);$$

2. the mixing coefficients $\boldsymbol{A}^{(r+1)}$ from

$$p\left(\boldsymbol{A} | \boldsymbol{X}, \boldsymbol{S}^{(r+1)}, \boldsymbol{\theta}^{(r)}\right) \propto p\left(\boldsymbol{X} | \boldsymbol{S}^{(r+1)}, \boldsymbol{A}, \boldsymbol{\theta}^{(r)}\right) \times p\left(\boldsymbol{A} | \boldsymbol{\theta}^{(r)}\right);$$

3. the noise variances $\boldsymbol{\theta}_1^{(r+1)}$ from

$$p\left(\boldsymbol{\theta}_1 | \boldsymbol{X}, \boldsymbol{S}^{(r+1)}, \boldsymbol{A}^{(r+1)}\right) \propto p\left(\boldsymbol{X} | \boldsymbol{S}^{(r+1)}, \boldsymbol{A}^{(r+1)}, \boldsymbol{\theta}_1\right) \times p\left(\boldsymbol{\theta}_1\right);$$

4. the source hyperparameters $\boldsymbol{\theta}_2^{(r+1)}$ from

$$p\left(\boldsymbol{\theta}_2 | \boldsymbol{S}^{(r+1)}\right) \propto p\left(\boldsymbol{S}^{(r+1)} | \boldsymbol{\theta}_2\right) \times p\left(\boldsymbol{\theta}_2\right);$$

5. the mixing coefficient hyperparameters $\boldsymbol{\theta}_3^{(r+1)}$ from

$$p\left(\boldsymbol{\theta}_3 | \boldsymbol{A}^{(r+1)}\right) \propto p\left(\boldsymbol{A}^{(r+1)} | \boldsymbol{\theta}_3\right) \times p\left(\boldsymbol{\theta}_3\right).$$

All the stochastic simulation steps including the expressions of the conditional posterior distributions and their simulation techniques are detailed in [MOU 06], where this method is termed as *Bayesian Positive Source Separation* (BPSS).

6.3.7.2. *Estimation using the Marginal Posterior Mean Estimator*

The Gibbs sampler is iterated in such a way to simulate M samples $\left\{\boldsymbol{S}^{(r)}, \boldsymbol{A}^{(r)}, \boldsymbol{\theta}^{(r)}\right\}_{r=1}^{M}$ from the posterior distribution (6.21). The first L samples corresponding to the burn-in of the Markov chains are discarded and therefore the Monte Carlo approximation for the minimum mean squares estimator (MMSE) is achieved by

$$\hat{\boldsymbol{z}} = \frac{1}{M - L} \sum_{r=L+1}^{M} \boldsymbol{z}^{(r)}, \tag{6.22}$$

where $\boldsymbol{z} \in \{\boldsymbol{S}, \boldsymbol{A}, \boldsymbol{\theta}\}$.

6.4. Illustration of the Bayesian Approach using Synthetic Data Sets

6.4.1. *Mixture Data Synthesis*

The mixture data are obtained by constructing analytically three spectral signals of $n = 1000$ data points. Each pure spectrum is obtained as the superposition of 15 Gaussian and Lorentzian [1] shapes with randomly simulated location, amplitude and width parameters. The concentration profiles are obtained from a sequential reaction $A + B \rightarrow C \rightarrow D + E$, where the constructed spectra are assumed, respectively, as those of the components A, C, D. The spectra of reactant B and E are supposed non-observable. The evolution profiles are calculated for $m = 10$ measures with randomly chosen constant rates. The retained mixing coefficients are chosen in such a way that there is no zero concentration value in the measured data. Figures 6.3 and 6.4 show the simulated synthetic pure component spectra and their concentration profiles.

6.4.2. *Performance Evaluation and Noise Level Index*

The noise level in each i-th mixture is defined as the ratio between the standard deviation of the i-th noise sequence and the root mean square of the i-th noise free mixture. It is expressed in percent. The reconstruction quality of each pure spectrum or mixing coefficient profile is assessed using a measure of dissimilarity between their recovered and their reference values. The dissimilarity is noted *diss* and defined by

$$diss(z_j, \hat{z}_j) = \sqrt{1 - corr(z_j, \hat{z}_j)^2} \quad \geq 0, \tag{6.23}$$

where $z \in \{S, A\}$ and $corr(z_j, \hat{z}_j)$ is the correlation coefficient between z_j and its estimate $\hat{z}_j$. The overall estimation performance is assessed using the *global dissimilarity*, defined by

$$gdiss = \frac{1}{p} \sum_{j=1}^{p} diss(z_j, \hat{z}_j). \tag{6.24}$$

The estimation is better for lower values of these performance indexes. The results are also evaluated qualitatively by a visual comparison of the reference and recovered spectra and mixing coefficients. Note that the use of the dissimilarity measure is possible only when the reference spectra are available, in an evaluation framework of the methods. Otherwise, a criterion such as the lack of fit can be used to assess how the estimated pure spectra and mixing coefficients can reproduce the observed data. However, it does not measure how well the spectra and concentration profiles are estimated.

---

1. The Lorentzian function and the Cauchy distribution correspond to the same function, whose inverse Fourier transform is the Laplace distribution. The term ?Cauchy distribution ? is usually used in statistics while the term ?Lorentzian ? is used in physics.
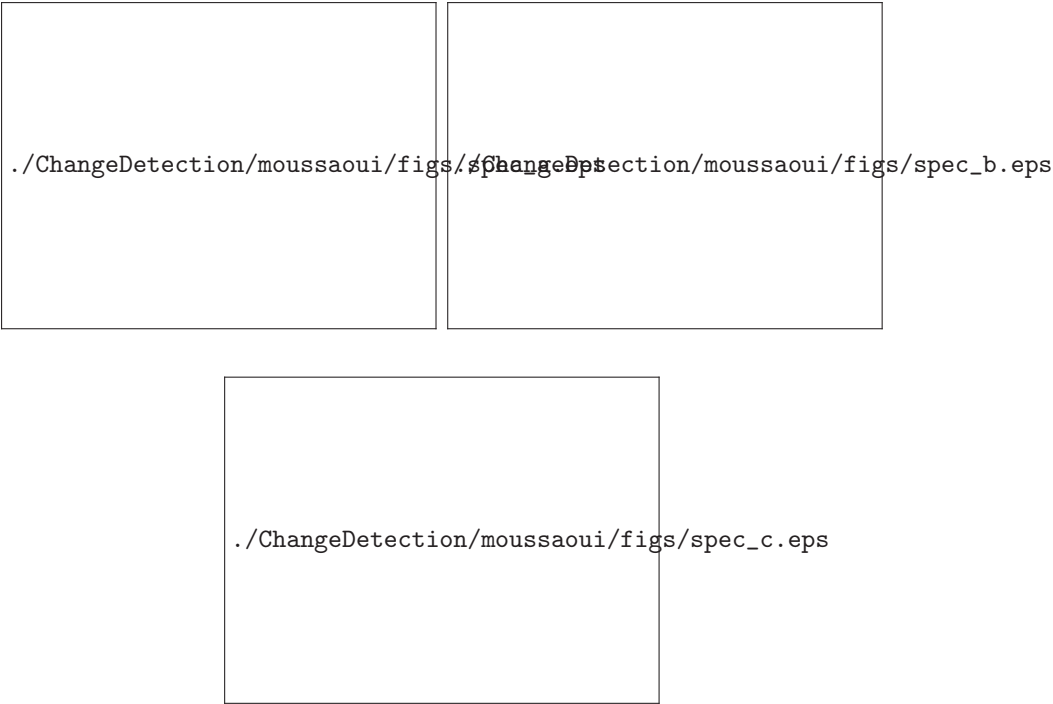
Figure 6.3. *Simulated pure spectra.*

Figure 6.4. *(a) Simulated concentration profiles during all the reaction time and (b) retained mixing coefficients for an observation time interval* $[4, 30\,min]$ *with* $m = 10$ *mixtures.*

6.4.3. *Admissible solutions*

A key point in mixture analysis is rotational ambiguity since the resolution using only non-negativity leads to a set of admissible solutions, which in the noise free case are a linear transformation of the true solution [LAW 71, BOR 85]. A challenging task is to state the admissible solutions and to define further information or assumptions allowing to get a solution where the estimated pure spectra and mixing coefficients are as close as possible to the actual ones [DON 03, MOU 05]. To give an idea about the range of admissible solutions of the synthetic mixture data presented in 6.4.1, the algorithm of [SAS 83] is used to obtain a set of feasible linear transformations of the actual pure spectra and mixing coefficients yielding transformed pure spectra and mixing coefficients which also are non-negative. Assume pure spectra summing to unity and consider the transformation matrix given as

$$
\boldsymbol{T} = \begin{bmatrix} (1 - t_1 - t_2) & t_1 & t_2 \\ t_3 & (1 - t_3 - t_4) & t_4 \\ t_5 & t_6 & (1 - t_5 - t_6) \end{bmatrix}, \qquad (6.25)
$$

which ensures to get transformed pure spectra summing to unity thus allowing therefore to handle the scale ambiguity.

Figure 6.5 shows the values of (t_1, t_2), (t_3, t_4) and (t_5, t_6) yielding a set of non-negativity constrained admissible solutions. It appears clearly that the rotational ambiguity is two-sided since there are transformation matrices with either negative or positive parameters. In that respect, using methods based only on non-negativity, such as ALS or NMF, yield for each initialization one particular solution among these admissible ones.

6.4.4. *Separation using the Bayesian approach*

The proposed algorithm is applied to the mixture data described in section 6.4.1. Random Gaussian noise is added to the mixture data in such a way to get a noise level of 5% in each mixture. To run the MCMC algorithm, the Markov chains associated to the pure component spectra, the mixing coefficients and the hyperparameters are randomly initialized to random positive values. The sampler is then run for 5000 iterations and the last 1000 simulated samples are retained to calculate the marginal posterior mean estimates. Figure 6.6 shows the evolution of the marginal posterior mean estimates of the parameters of pure spectra and mixing coefficients *a priori* distributions. Note particularly that the values of these parameters are different for the three components and the values of α_j are less than unity, which corresponds to assigning an *a priori* Gamma distribution of the first case illustrated in figure 6.1. The final estimates of pure component spectra and mixing coefficient profiles are reported in figure 6.7. It can be noted the high similarity between the recovered pure spectra and mixing coefficients and their actual values.

Figure 6.5. *Permitted regions of the rotational parameters. These regions has been obtained using the method of [SAS 83]. The algorithm has been run with 1000 random initializations.*

6.4.5. *Performance Evaluation*

6.4.5.1. *Computation Time*

The computation time at each iteration of the MCMC algorithm depends on the number of data points, n, in each spectrum, the number of observed mixtures, m, and the number of pure components, p. The overall computation time also depends on the number of iterations necessary to the algorithm convergence. In order to give an idea about the computation time of the algorithm, an experiment is performed by varying the three parameters (m, n, p) and evaluating the time duration of the analysis using a computer of an Intel Pentium 4 3.2 GHz processor and 2 GB RAM. Figure 6.8 shows that the running time of the algorithm increases roughly linearly with all the three parameters (m, n, p). For medium data sets, the analysis can be performed in a reasonable amount of time : for a mixture with dimensions $(m = 30, n = 1000, p = 4)$, the running time for 5000 iterations is about 12 minutes. However, for larger data sets and higher number of iterations, the algorithm needs longer time for the analysis.

Figure 6.6. *Evolution of the marginal posterior mean estimates of the parameters of the prior distributions associated to the pure spectra and mixing coefficients.*

6.4.5.2. *Effect of Noise Level*

The noise level is increased gradually from 1% to 40% and for each value of the noise level the estimation quality of the pure spectra and mixing coefficient is evaluated in terms of global dissimilarity. From figure 6.9, it can be noted that for low noise levels the pure spectra are reconstructed correctly. For a higher noise level, the pure spectra dissimilarity takes higher values since a noise components remains in the estimated spectra, however the dissimilarity of the mixing coefficients is less sensitive the noise level than the pure spectra. A possible approach to enhance the quality of the pure spectra reconstruction is to include a denoising step during the analysis by introducing a smoothness hypothesis.

6.4.5.3. *Effect of Spectra Correlation*

In order to discuss the effect of the mutual correlation of the pure spectra on the performances of the BPSS method, a mixture of two synthetic spectra is used. The correlation coefficient of these spectra is varying while the mixing coefficient are kept fixed. This simulation is performed by a first randomly simulated spectrum which is shifted in order to deduce the second spectrum. Therefore, the correlation coefficient

Figure 6.7. *(a–c) Estimated spectra (continuous curves) and comparison with the reference spectra (dotted curves) and (d) estimated mixing coefficients for 5% noise in the mixtures.*

Figure 6.8. *Average computation time per iteration with respect of the data matrix size and number of components.*

Figure 6.9. *Influence of the noise level on the estimation performances. A Monte Carlo experiment with 100 simulations is performed*
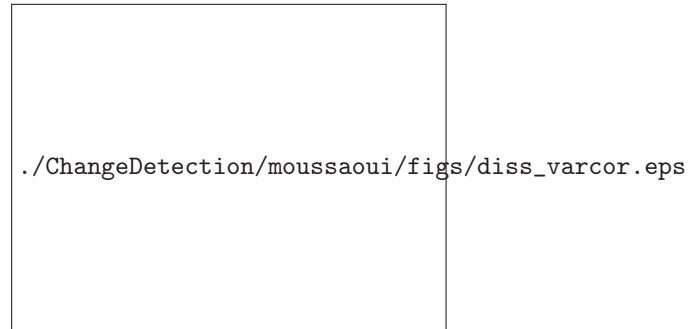
Figure 6.10. *Influence of the spectra correlation on the algorithm performances.*

between the two spectra will depend on the value of this shift. The BPSS algorithm is then applied for the analysis of the resulting mixture. The number of algorithm iterations is fixed to 10000 and estimation is performed using the last 1000 realization of the Markov chain. The results of this experience are shown in figure 6.10, where the performance index of the BPSS method is evaluated with respect to the correlation coefficient of the two spectra and compared with that of the NNICA method. The NNICA method is chosen because this method works under the constraint of orthogonal signals. The two methods reach their best performances as the mutual correlation is small. On the one hand, it can be noted that BPSS algorithm is less sensitive to the mutual correlation of the two spectra as compared to NNICA method. For example, for a correlation coefficient of 0.6 the performance index of the BPSS method remains acceptable. This result can be explained by the fact that, contrarily to the NNICA method, the BPSS algorithm do not optimize an independence measure between the spectra and do not impose strictly their orthogonality.

6.4.5.4. *Effect of a Background*

To discuss this effect, a mixture of three synthetic spectra where different exponential backgrounds are added to the pure spectra. Figure 6.11 illustrates the three simulated spectra for which backgrounds with unit amplitude are added. The performances of the analysis are discussed with respect to the background level. The result of this experience is shown in figure 6.11. When no background is added, the performance index gives an information of the adequation of the Gamma density to represent the pure spectra distribution. By increasing the background level, the pure spectra will be dominated by the exponential background. In this case, the performance index indicates how the Gamma law can represent the background distribution. In the case of two backgrounds, these errors increases leading to a decrease of the performances. As a conclusion, this simulation shows that for low background amplitude, the Gamma prior model is adequate and leads to good performances but for high background this prior model may be inadequate. In that respect, an extension of the proposed approach is to consider a generalized or shifted Gamma distribution. Another alternative is to include a step of data pre-processing for background removal or to use an *a priori* model in which each pure spectrum may be considered as a superposition of a sparse spectrum with a slowly varying background.

6.5. Conclusion

A Bayesian method for spectral mixture analysis based on assigning Gamma distribution as an *a priori* model on pure component spectra and mixing coefficient was presented in this chapter. This model allows to encode both the sparsity and the non-negativity of these variables. The estimation is achieved using a Markov chain Monte Carlo methods which allows to simulate samples from the joint *a posteriori* distribution. The use of this stochastic estimation technique precludes the convergence to local minimums and yields a method working in an unsupervised framework. The only necessary parameters to tune is the number of components and the number of algorithm iterations. Some experiments with a synthetic data set have been performed to discuss the algorithm efficiency in terms of robustness against the noise level and computation time. It turned out that the method yields effective results even for high noise level in a reasonable time.

A fundamental question in source separation is the need to state further information, in addition to non-negativity, allowing to define the best solution among the admissible ones. By incorporating additional assumptions or constraints, like unimodality and full addditivity (closure) [BRO 98, **?**], the range of admissible solutions will be reduced to a smaller set. On the other hand, a penalized least squares or a Bayesian estimation method leads to one particular solution fulfilling at best the *a priori* distribution model.

Figure 6.11. *(a,b,c) Pure spectra for a background of unit amplitude in spectra A and B (d) evolution of the performance with respect to the background amplitude.*

In the proposed approach, the statistical independence and the sparsity of both the pure component spectra and the mixing coefficients are assumed. Although the Gamma distribution seems to be a good model to encode these information, but nothing guarantee that this prior model is adequate in all circumstances. In that respect, the BPSS method provides pure component spectra and mixing coefficients which are mutually statistically independent and whose statistical distribution is very close to a Gamma distribution. The performances of the BPSS method depend on the ability of the Gamma distribution to represent the actual pure component spectra and mixing coefficient distributions. If the pure component spectra are not Gamma distributed, the solution may be far from the true one and it would be advantageous to use another statistical model. For example, a model allowing to incorporate alternative information, such as smoothness of the component spectra or closure or unimodality of the mixing coefficients profiles.

[BOR 85] BORGEN O., KOWALSKI B., « An Extension of the Multivariate Component Resolution Method to three Component », *Analytica Chimica Acta*, vol. 174, p. 1–26, 1985.

[BRO 97] BRO R., DE JONG S., « A Fast Non-negativity Constrained Least Squares Algorithm », *Journal of Chemometrics*, vol. 11, p. 393–401, 1997.

[BRO 98] BRO R., SIDROPOULOS N., « Least squares algorithms under unimodality and non-negativity constraints », *Journal of Chemometrics*, vol. 12, p. 223–247, 1998.

[CHA 07] CHANG C.-I., *Hyperspectral Data Exploitation. Theory and Applications*, Wiley Interscience, 2007.

[CHE 00] CHEN M.-H., SHAO Q.-M., IBRAHIM J. G., *Monte Carlo Methods in Bayesian Computation*, Springer-Verlag, New York, 2000.

[CIC 02] CICHOCKI A., AMARI S., *Adaptive blind signal and image processing - Learning algorithms and applications*, John Wiley, 2002.

[COM 91] COMON P., JUTTEN C., HÉRAULT J., « Blind Separation of Sources, Part II : Problems Statement », *Signal Processsing*, vol. 24, p. 11-20, 1991.

[COM 94] COMON P., « Independent component analysis, a new concept ? », *Signal Processing,*, vol. 36, n°3, p. 287–314, 1994.

[De 97] DE JUAN A., Y.V.HEYDEN, TAULER R., MASSART D., « Assessment of New Constraints Applied to the Alternating Least Squares Method », *Analytica Chimica Acta*, vol. 346, p. 307–318, 1997.

[DEM 89] DEMOMENT G., « Image reconstruction and restoration : Overview of common estimation structures and problems », *IEEE Transactions on Acoustics Speech and Signal Processing*, vol. 37, n°12, p. 2024–2036, 1989.

[DON 03] DONOHO D., STODDEN V., « When Does Non-Negative Matrix Factorization Give a Correct Decomposition Into Parts ? », *Advances in Neural Information Processing Systems (NIPS'2003)*, Cambridge, United States, 2003.

[FÉV 06] FÉVOTTE C., GODSILL S., « A Bayesian Approach for Blind Separation of Sparse Sources », *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14, n°6, p. 2174–2188, novembre 2006.

[GEM 84a] GEMAN S., GEMAN D., « Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images », *IEEE Transactions on Pattern Analysis Machine Intelligence*, vol. 6, n°6, p. 721–741, 1984.

[GEM 84b] GEMPERLINE P., « A priori Estimates of the Elution Profiles of the Components in Overlapped Liquid Chromatography Peaks Using Target Factor Analysis », *Journal of Chemical Information and Computer Sciences*, vol. 24, p. 206–212, 1984.

[GIL 99] GILKS W., RICHARDSON S., SPIEGEHALTER D., *Markov Chain Monte Carlo in Practice*, Chapman & Hall, 1999.

[HAS 70] HASTINGS W., « Monte Carlo sampling methods using Markov chains and their applications », *Biometrika*, vol. 57, n°1, p. 97–109, 1970.

[HEN 90] HENRY R., KIM B., « Extension of Self-Modeling Curve Resolution to Mixtures of more than Three Components. Part I. Finding The Basic Feasible Region », *Chemometrics and Intelligent Laboratory Systems*, vol. 8, p. 205–216, 1990.

[HOY 04] HOYER P., « Non-negative Matrix Factorization with Sparseness Constraints », *Journal of Machine Learning Research*, vol. 5, p. 1457—1469, 2004.

[HYV 01] HYVÄRINEN A., KARHUNEN J., OJA E., *Independent Component Analysis*, John Wiley, New York, 2001.

[JIA 04] JIANG J.-H., LIANG Y., OZAKI Y., « Principles and methodologies in self-modeling curve resolution », *Chemometrics and Intelligent Laboratory Systems*, vol. 71, p. 1–12, 2004.

[JUT 91] JUTTEN C., HERAULT J., « Separation of Sources, Part I », *Signal Processing*, vol. 24, n°1, p. 1–10, 1991.

[KNU 99] KNUTH K., « A Bayesian approach to source separation », CARDOSO J.-F., JUTTEN C., LOUBATON P., Eds., *Proceedings of International Workshop on Independent Component Analysis and Signal Separation (ICA'99)*, Aussios, France, p. 283–288, 1999.

[LAW 71] LAWTON W., SYLVESTRE E., « Self-Modeling Curve Resolution », *Technometrics*, vol. 13, p. 617–633, 1971.

[LAW 74] LAWSON C., HANSON R., *Solving Least-Squares Problems*, Prentice-Hall, , 1974.

[LEE 99] LEE D., SEUNG H., « Learning the parts of objects by non–negative matrix factorization », *Nature*, vol. 401, p. 788–791, 1999.

[MAE 88] MAEDER M., ZILIAN A., « Evolving Factor Analysis, A New Multivariate Technique in Chromatography », *Chemometrics and Intelligent Laboratory Systems*, vol. 3, p. 205–213, 1988.

[MAL 00] MALINOWSKI E., *Factor Analysis in Chemistry*, John Willey & Sons, 3rd édition, 2000.

[MIS 01] MISKIN J., MACKAY D., « Ensemble Learning for blind source separation », *S. Roberts and R. Everson, editors, Independent Component Analysis : Principles and Practice*, Cambridge University Press, p. 209-233, 2001.

[MOH 99] MOHAMMAD-DJAFARI A., « A Bayesian approach to source separation », *American Institute of Physics (AIP) proceedings*, vol. 567, p. 221–244, 1999.

[MOU 05] MOUSSAOUI S., BRIE D., IDIER J., « Non-negative Source Separation : Range of Admissible Solutions and Conditions for the Uniqueness of the Solution », *proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'2005)*, Philadelphia, USA, March 2005.

[MOU 06] MOUSSAOUI S., BRIE D., MOHAMMAD-DJAFARI A., CARTERET C., « Separation of non-negative mixture of non-negative sources using a bayesian approach and MCMC sampling », *IEEE Trans. on Signal Processing*, vol. 54, n°11, p. 4133–4145, 2006.

[MOU 08] MOUSSAOUI S., HAUKSDOTTIR H., SCHMIDT F., JUTTEN C., CHANUSSOT J., BRIE D., DOUTE S., BENEDIKSSON J., « On the decomposition of Mars hyperspectral data by ICA and Bayesian positive source separation », *Neurocomputing*, vol. 10, p. 2194–2208, June 2008.

[NUZ 98] NUZILLARD D., BOURG S., NUZILLARD J.-M., « Model-free analysis of mixtures by NMR using blind source separation », *Journal of Magnetic Resonance*, vol. 133, p. 358–363, 1998.

[OCH 99] OCHS M. F., STOYANOVA R. S., ARIAS-MENDOZA F., BROWN T. R., « A new method for spectral decomposition using a bilinear Bayesian approach », *Journal of Magnetic Resonance*, vol. 137, p. 161–176, 1999.

[OHT 73] OHTA N., « Estimating Absorption Bands of Component Dyes by Means of principal Component Analysis », *Analytical Chemistry*, vol. 45, n°3, p. 553–557, 1973.

[PAA 94] PAATERO P., TAPPER U., « Positive matrix factorization : a nonnegative factor model with optimal utilization of error estimates of data values », *Environmetrics*, vol. 5, p. 111–126, 1994.

[PAA 97] PAATERO P., « Least Squares Formulation of Robust non-negative factor analysis », *Chemomerics and Intelligent Laboratory Systems*, vol. 37, p. 23–35, 1997.

[PFE 51] PFEIFFER H., LIEBHAFSKY H., « The origins of Beer's law », *Journal of Chemical Education*, vol. 28, page123, 1951.

[PLU 03] PLUMBLEY M., « Algorithms for non–negative independent component analysis », *IEEE Transactions on Neural Networks*, vol. 14, n°3, p. 534–543, 2003.

[RIC 94] RICCI R., DITZLER M., NESTOR L., « Discovering the Beer-Lambert Law », *Journal of Chemical Education*, vol. 71, p. 983–985, November 1994.

[ROB 98] ROBERTS S., « Independent Component Analysis : Source Assessment and Separation, A Bayesian Approach », *IEE Proceedings on Vision, Image and Signal Processing*, vol. 145, n°3, p. 149–154, 1998.

[ROB 99] ROBERT C., *Monte Carlo Statistical Methods*, Springer-Verlag, 1999.

[ROB 01] ROBERT C., *The Bayesian Choice*, Springer-Verlag, 2nd édition, 2001.

[ROW 03] ROWE D., *Multivariate Bayesian Statistics : Models for Source Separation and Signal Unmixing*, CRC Press, Boca Raton, Florida, USA, 2003.

[SAJ 03] SAJDA P., DU S., BROWN T., PARRA L., STOYANOVA R., « Recovery of Constituent Spectra in 3D Chemical Shift Imaging using Non-Negative Matrix Factorization », *Proceedings of International Conference on Independent Component Analysis and Blind Signal Separation (ICA'2003)*, p. 71–76, 2003.

[SAN 96] SANCHEZ F. C., TOFT J., VAN DEN BOGAERT B., MASSART D. L., « Orthogonal projection approach applied to peak purity assessment », *Analytical Chemistry*, vol. 68, p. 79–85, 1996.

[SAN 97] SANCHEZ F. C., RUTAN S., GARCIA M. G., MASSART D., « Resolution of multicomponent overlapped peaks by the orthogonal projection approach, evolving factor analysis and window factor analysis », *Chemometrics and Intelligent Laboratory Systems*, vol. 36, p. 153–164, 1997.

[SAS 83] SASAKI K., KAWATA S., MINAMI S., « Constrained Nonlinear Method for Estimating Component Spectra from Multicomponent Mixtures », *Applied Optics*, vol. 22, n°22, p. 3599–3606, 1983.

[SAS 84] Sasaki K., Kawata S., Minami S., « Estimation of Component Spectral Curves from Unknown Mixture Data », *Applied Optics*, vol. 23, n°12, p. 1955–1959, 1984.

[SAS 87] Sasaki K., Kawata S., Minami S., « Component Analysis of Spatial and Spectral Patterns in Multispectral Images. I. Basics », *Journal of the Optical Society of America. A*, vol. 4, n°11, p. 2101–2106, 1987.

[SAS 89] Sasaki K., Kawata S., Minami S., « Component Analysis of Spatial and Spectral Patterns in Multispectral Images. II. Entropy Minimization », *Journal of the Optical Society of America. A*, vol. 6, n°1, p. 73–79, 1989.

[SNO 06] Snoussi H., Idier J., « Bayesian Blind Separation of Generalized Hyperbolic Processes in Noisy and Underdeterminate Mixtures », *IEEE Transactions on Signal Processing*, vol. 54, n°9, p. 3257 - 3269, September 2006.

[STU 94] Stuart A., Ord J., *Kendall's Advanced Theory of Statistics*, vol. 1 : Distribution Theory, John Wiley & Sons, 6th édition, 1994.

[TAU 93a] Tauler R., Izquierdo-Ridorsa A., Casassas E., « Simultaneous analysis of several spectroscopic titrations with self-modeling curve resolution », *Chemometrics and Intelligent Laboratory Systems*, vol. 18, p. 293–300, 1993.

[TAU 93b] Tauler R., Kowalski B., Fleming S., « Multivariate Curve Resolution Applied to Spectral Data from Multiple Runs of an Industrial Process », *Analytical Chemistry*, vol. 65, p. 2040–2047, 1993.

[TAU 00] Tauler R., « Calculation of maximum and minimum band bounderies of feasible solutions for species profiles obtained by multivariate curve resolution », *Journal of Chemometrics*, vol. 15, p. 627–646, 2000.

[WID 03] Widjaja E., Garland M., « Pure component spectral reconstruction from mixture data using SVD, global entropy minimization, and simulated annealing. Numerical investigations of admissible objective functions using a synthetic 7-species data set », *Journal of Computational Chemistry*, vol. 23, n°9, p. 911-919, 2003.

[WIN 90] Windig W., « Mixture Analysis of Spectral Data by Multivariate Methods », *Chemometrics and Intelligent Laboratory Systems*, vol. 9, p. 7-30, 1990.

[WIN 91] Windig W., Guilment J., « Interactive Self-Modeling Mixture Analysis », *Analytical Chemistry*, vol. 63, p. 1425–1432, 1991.

[YIN 98] Yingzhi Z., Garland M., « An improved algorithm for estimating pure component spectra in exploratory chemometric studies based on entropy minimization », *Analytica Chimica Acta*, vol. 359, p. 303–310, 1998.

Chapitre 7

Detection and tracking of emission rays in radioastronomy

7.1. Problem Statement

7.1.1. *Multispectral Data Cube Kinematic Segmentation in Radioastronomy*

Galaxies are astronomical object gathering many gas clouds (or gas structures) in motion : due to Universe expansion (also known as *redshift*) they are generally moving away from the Earth. Also, the gas structures are moving separately because of the specific movements within the galaxy. All these moves are very interesting for the astronomers since they allow them to understand the formation and evolution of galaxies. They can be observed through the spectrum of the gas which is shifted due to the Universe expansion and the specific movements within the galaxy : this phenomenon is known as Doppler effect. Then, the kinematic zones in the galaxy are characterized by their own spectrum.

Radio interferometers like the VLA (USA), Westerbork (Netherlands), ATCA (Australia) or Plateau de Bure (France) provide line observations which are 3D data cubes : two dimensions correspond to the astronomical coordinates and the thid dimendion is the frequency (or velocity channels). Within these data cubes each image pixel contains an atomic or molecular line spectrum shifted with respect to the moves of the gas. The observations are made around a single line (*e.g.* a CO or HI line) which is shifted according to the radial velocity of the observed gas. However, the complexity of the 3D structure contained in these data cubes increases with the sensitivity of the

Chapitre rédigé par V. MAZET, F. FLITTI, Ch. COLLET.

observations. Moreover, the amount of data to manipulate when spatial and spectral resolution of the sampling grids increase bring specific difficulties to manipulate huge data set (see chapter 13 part 5 of this book). In the radio domain, telescope sensitivities are increasing tremendously with the upgrade of existing instruments and the building of new telescopes (EVLA, ALMA, SKA). This will lead to an enormous increase of available 3D data, which will be more and more complex.

The standard method for the study of data cubes is the visual inspection of the channel maps and the creation of moment maps after clipping the spectra at a level of 3σ where σ is the rms (residual mean square) noise in one channel : the 0-order moment is the integrated intensity, the first order moment the velocity field and the second order moment the distribution of the velocity dispersion (figure 7.1). As long as the 3D intensity distribution is not too complex, these maps give a fair impression of the 3D information contained in the cube. However, when the 3D structure becomes complex, the visual inspection of the velocity channels becomes difficult and important information is lost in the moment maps, since they are produced by integrating the spectra and thus do not reflect the individual line profiles. The methods proposed in this chapter are an attempt to provide a segmentation map containing additional information on line profiles. Therefore, it will be possible to focus the inspection of the channel maps on kinematically interesting regions.

To illustrate our proposal, we have chosen a target which is not too complex : NGC 4254, a spiral galaxy located in the Virgo cluster. The Hɪ 21cm observations were made with the VLA [PHO 93]. The central velocity of the data cube is $v =$

(1)   (2)
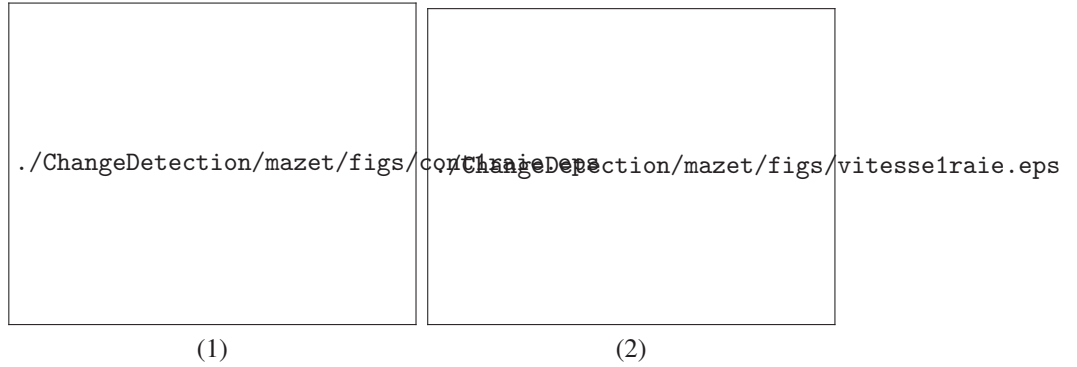
Figure 7.1. *NGC 4254 [PHO 93]. Left panel : Hɪ emission distribution (moment 0). The x-axis corresponds to right ascension, the y-axis to declination. The celestial north is at the top of the image and the celestial east at the left. The intensity is given in mJy. Right panel : Hɪ velocity field (moment 1). The velocities are given in channel maps.*

2408 km s^{-1} at channel number 22 and the velocity resolution is 10 km s^{-1} per channel. For simplicity, we keep in this chapter pixel numbers for the coordinate axis and channel numbers for the velocity axis. The rms noise of one channel map is $\sigma = 0.43$ mJy[1], so the cube was clipped at an intensity level of $3\sigma = 1.29$ mJy. Figure 7.1 shows the maps of the first two moments integrated over the whole data cube. The map of the HI emission distribution (moment 0, left panel) shows an inclined gas disk with a prominent one-armed spiral to the west. The velocity field (moment 1, right panel) is that of a rotating disk with perturbations to the north-east and to the north. The distribution of the velocity dispersion (moment 2), which is not shown, is almost uniform with a small maximum in the galaxy center. The subsequent work was done on the same data cube clipped at 3σ.

7.1.2. *Outline of the Proposed Approaches*

The goal is to help the astronomer in handling complex 3D data by providing an unsupervised tool which differentiate the kinematic zones by segmenting the hyperspectral image in homogeneous spectral areas. In other words, it consists in estimating the shift and the intensity of each line in each pixel (that is, each spectrum) in the image. Because several gas structures imply several line in each spectrum, we call in the sequel each line occurrence as *component*. In consequence, the cube is an image where each pixel is a spectrum containing one or several components whose position (shift), amplitude (intensity) and width have to be estimated. From this estimation, a second processing has to made a segmentation of the image in homogeneous kinematic zones.

The segmentation process on such a data cube requires a non trivial modeling step based on Bayesian inference for both methods. Due to the curse of dimensionality (Hughes phenomenon), Bayesian segmentation can sometimes be carried out on reduced data (*e.g.*, Principal or Independent Component Analysis [HYV 01], Projection Pursuit [HUB 85] etc.).

In a first approach (section 7.2), the spectrum of each pixel is fitted by a mixture model which is a combination of a fixed number of Gaussian functions (section 7.2.1). Then we select the six most representative mean values (channel number of the maximum) and variances (widths) (section 7.2.2). In this way the data cube, which had initially 42 channels, is reduced to 6 effective bands to form a basis. In a second step the weights of the Gaussians with fixed mean and variance are determined by fitting again the observed spectra. The segmentation task, which is carried out on the weights, then consists of clustering the pixels into different classes (or labels) according to similar behaviors defined by a chosen criterion (section 7.2.3). In this way we obtain a

---

1. The jansky (abbreviated Jy) is a unit of radio flux density (the rate of flow of radio waves) ; 1 Jy = 10^{-26} watts/m^2/Hz.

segmentation map containing spatially homogeneous classes of pixels with a similar spectrum. We present the application of this method to the data cube of NGC 4254 in section 7.2.4.

The second approach (section 7.3) consists in decomposing each pixel (*i.e.* each spectrum) into components whose number, position, amplitude and widths are independently estimated. For that purpose, a Monte Carlo Markov Chain (MCMC) algorithm is used (section 7.3.2). Then, a second step consists in "tracking" the estimated components, from one pixel to another, in order to gather similar components into a same class. Finally, the classes provide an estimation of the 3D spectral homogeneous structures (section 7.3.3).

Before detailing the approaches, let clarify their common features and differences. The two proposed approaches work in two steps : the first one is dedicated to the decomposition of each spectrum and the second one aims at classifying (or segmenting) the spectrum with respect to this decomposition. The decomposition step considers each spectrum as a sum of Gaussian components whose amplitude, position and width are estimated. Nevertheless, the first approach sets the component number to six in each spectrum while the second approach estimates the number for each spectrum. Besides, the first approach decomposes the spectra on a unique basis (that is, the six components have the same position and with in each spectrum), this is not the case with the second approach. This better estimation has to be tempered by the fact that the optimisation methods in each approach are different and can be time-consuming in the second one. Also, the classification methods in the second steps are different.

7.2. Multiband Segmentation by Reduction Approach

This approach is intended to be complementary to traditional methods for the study of 3D data cubes, *i.e.,* the eye inspection of the channel maps, the creation of moment maps, and the creation of position-velocity plots. Because spectroscopic line data are complex, a segmentation map coupled with the moment maps can be used for a first inspection of the cube. Moreover, masks corresponding to different classes in the segmentation map can be produced to isolate kinematically interesting regions.

7.2.1. *Gaussian Mixture Model and Parameter Estimation*

Let a multispectral image with N channels defined on a regular lattice S of size $H \times W$ pixels. This image can be manipulated as $D = H \times W$ spectra y_s, $s \in S$, each one of N spectral bands. Each spectrum is modeled by a P-component Gaussian mixture $\hat{y}_s(j)$:

$$\hat{y}_s(j) = \sum_{k=1}^{P} \frac{a_{sk}}{\sqrt{2\pi w_{sk}}} \exp\left(\frac{-1}{2w_{sk}}(j - c_{sk})^2\right) \tag{7.1}$$

where a_{sk} stands for the weight, w_{sk} the variance, c_{sk} the mean associated with the kth Gaussian component of the sth spectrum, and j is the index of the frequency channel. The observed spectra of each pixel are fitted by a mixture of Gaussian functions using an Expectation-Maximization (EM) algorithm, which is an iterative method for maximum likelihood estimation.

Usually, the EM algorithm works on realizations of an unknown probability density function (pdf) [**?**]. In our case we assume that the spectrum already represents the pdf. Thus, we adapt the algorithm to fit directly the spectrum. To satisfy the assumptions required by the EM adaptation, each spectrum y_s is considered to be normalized such that $y_s(j) \geq 0$, $\forall j$ and $\sum_{j=1}^{N} y_s(j) = 1$. Within the iterative process the contribution of the kth component of the Gaussian mixture $\mathcal{R}^{[i]}(k, s, j)$ is calculated as follows :

$$\mathcal{R}^{[i]}(k, s, j) = \frac{\frac{a_{sk}^{[i]}}{\sqrt{2\pi w_{sk}^{[i]}}} \exp\left(\frac{-1}{2w_{sk}^{[i]}}(j - c_{sk}^{[i]})^2\right)}{\sum_{k=1}^{P} \frac{a_{sk}^{[i]}}{\sqrt{2\pi w_{sk}^{[i]}}} \exp\left(\frac{-1}{2w_{sk}^{[i]}}(j - c_{sk}^{[i]})^2\right)} \ , \tag{7.2}$$

where i represents the iteration step. These contributions are then used to calculate the parameters in the next iterative step according to :

$$a_{sk}^{[i+1]} = \frac{\sum_{j=1}^{N} y_s(j)\mathcal{R}^{[i]}(k, s, j)}{\sum_{j=1}^{N} y_s(j)} \ , \tag{7.3}$$

$$c_{sk}^{[i+1]} = \frac{\sum_{j=1}^{N} j y_s(j)\mathcal{R}^{[i]}(k, s, j)}{\sum_{j=1}^{N} y_s(j)\mathcal{R}^{[i]}(k, s, j)} \ , \tag{7.4}$$

$$w_{sk}^{[i+1]} = \frac{\sum_{j=1}^{N} (j - c_{sk}^{[i+1]})^2 y_s(j)\mathcal{R}^{[i]}(k, s, j)}{\sum_{j=1}^{N} y_s(j)\mathcal{R}^{[i]}(k, s, j)} \ . \tag{7.5}$$

The weights a_{sk} are assumed to be all positive. After convergence, one obtains the parameters a_{sk}, w_{sk}, and c_{sk} to estimate $\hat{y}_s(j)$ of (equation 7.1).

7.2.2. *Spectral Basis Selection*

The modeling using equation 7.1 needs 3 parameters for each Gaussian component and leads to $3 \times P$ parameters instead of N channels for each pixel. The Markovian classifier allows a maximum number of ten input images due to the curse of dimensionality (Hughes phenomenon) : beyond this point, the parameter estimation becomes instable and inaccurate due to a lack of observed data compared with their high dimensionality. Nevertheless, with $P = 6$ we already obtain 18 reduced images (three parameters set $\{a_{sk}, w_{sk}, c_{sk}\}_{k=1,\cdots,6}$ for each $s \in S$). We therefore choose to select

the most representative w_{sk} and c_{sk} among all Gaussian mixtures described in section 7.2.1, which are then retained as a decomposition basis for all the pixels. For this parameter selection we used a clustering algorithm (K-means [DUD 01]) on the set of vectors $\{c_{sk}, w_{sk}\}_{s \in S}$.

In this way, equation 7.1 becomes :

$$\hat{\hat{y}}_s(j) = \sum_{k=1}^{P} \frac{\tilde{a}_{sk}}{\sqrt{2\pi w_k}} \exp\left(\frac{-1}{2w_k}(j - c_k)^2\right) \tag{7.6}$$

where c_k and w_k do not depend anymore on the location s. This is equivalent to the projection of the spectra on the subspace generated by the following basis :

$$\left\{ \frac{1}{\sqrt{2\pi w_k}} \exp\left(\frac{-1}{2w_k}(j - c_k)^2\right) \right\}_{k=1,\cdots,P}.$$

We choose the Levenberg-Marquardt algorithm [WAL 97] (widely used in astronomy) to determine efficiently the weights $\tilde{a}_{sk}$ with the selected basis. Thus a multicomponent image z with P components of size $H \times W$ is obtained to feed the Markovian classifier. The Markovian assumption takes into account the neighborhoods when classifying each pixel, allowing us to regularize the solution in a statistical framework [Gem 84, Pro 04]. Each pixel s is now represented by a P vector we note $z_s = (\tilde{a}_{sk})_{k=1,\cdots,P}$.

7.2.3. *Bayesian Classifier fed by Reduced Data*

From the observed image z to the segmented image x, the algorithm can be decomposed into three main steps [Pro 04] :

1) Initialization step : the aim is to provide a first estimation of the parameters (K-means algorithm) ;

2) Segmentation step : the restoration is then achieved using the Maximum Posterior Mode (MPM) segmentation rule [Pro 04] ;

3) Parameter estimation step : this step is realized using the Iterated Conditional Estimation algorithm [PIE 92].

For each observation z_s, one associates a hidden state $x_s = \omega_i$, where ω_i belongs to the label set $\Omega = \{\omega_i\}_{i=1,\ldots,K}$. The segmentation map x is obtained using a joint probability $P(x, z)$ and a chosen cost function. In hierarchical Markovian modeling, one assumes an in-scale dependence between hidden states (figure 7.2) to model the spatial correlation of the observations.

Figure 7.2. *Example of a dependency graph corresponding to a quadtree structure on a 4×4 lattice. Black circles represent labels and white circles represent multicomponent observations $\{z_s = (\tilde{a}_{sk})_{k=1,\cdots,P}\}_{s\in S}$. Each node t has a unique parent t^- and four children t^+.*

Let the quadtree $G = (T, L)$ be a graph composed of a set T of nodes and a set L of edges. A hidden state will be associated with each node, as illustrated in figure 7.2. Each node t, apart from the root r, has a unique predecessor, its "parent" t^- which leads ultimately to the root. Each node t, apart from the terminal ones, the "leaves", has four "children" t^+. The set of nodes T can be divided into "scales", $T = T^0 \cup T^1 \ldots \cup T^R$, according to the path length from each node to the root. Thus, $T^R = \{r\}$, T^n involves 4^{R-n} sites, and T^0 is the finest scale formed by the leaves ($T^0 = S$).

We consider a labeling process[2] x which assigns a class label x_t to each node of $G : x = \{x^n\}_{n=0}^R$ with $x^n = \{x_t, t \in T^n\}$ where x_t takes its values in the set Ω. The hidden process, *i.e.* the class label x, is assumed to be Markovian in scale. In this way each label set at level n only depends on the upper levels : $P\left(x^n \left| x^k, k > n\right.\right) = P\left(x^n \left| x^{n+1}\right.\right)$. Moreover, the probabilities of inter-scale transitions can be factorized in the following way [LAF 00b, Lue 93] : $P\left(x^n \left| x^{n+1}\right.\right) = \prod_{t\in T^n} P(x_t \left| x_{t^-}\right.)$, where

---

2. Note that, for clarity, x stands for the random process and its realization.

t^- indicates the parent of site s, as illustrated in figure 7.2. The likelihood of the observations z being in a state x at the bottom of the quadtree ($T^0 = S$) is expressed as the following product (assuming conditional independence) : $P(z|x) = \prod_{s \in S} P(z_s|x_s)$, where $\forall s \in S$, $P(z_s|x_s = \omega_i) \triangleq P_i^n(z_s)$ represents the likelihood of the data z_s. A multidimensional Gaussian pdf model is used to derive the latter expression.

Under these assumptions the joint distribution $P(x, z)$ follows a Gibbs law whose expression is given by : $P(x, z) = P(x_r) \prod_{t \neq r} P(x_t|x_{t^-}) \prod_{t \in T^0} P(z_s|x_s)$ [LAF 00b].

The expression of $P(x, z)$ allows us to calculate exactly and efficiently $P(x_t = \omega_i|z)$ for all nodes $t \in T$. The segmentation label map at the bottom of the quadtree is finally given by : $\hat{x}_s = \arg\max_{\omega_i \in \Omega} P(x_s = \omega_i|z)$. This expression assumes the model parameters (parameters of the Gaussian pdfs and inter-scale transition probabilities) are known. Such an estimation is obtained in an unsupervised way using the

./ChangeDetection/mazet/figs/methode.eps

Figure 7.3. *Data cube processing method summary : (a) adapted EM algorithm to fit all spectra using equation 7.1. (b) Selection of the P most relevant Gaussians obtained in (a) using a K-means algorithm. (c) Projection on the basis chosen in (b) using a Levenberg-Marquardt algorithm. (d) The reduced images feed a Markovian classifier leading to the final segmentation map.*

Iterative Conditional Estimation (ICE) algorithm [Pie 94]. In practice, one alternates model parameter estimation and segmentation until convergence [Pro 04].

To summarize, one feeds a Markovian classifier with a multicomponent image $\{z_s = (\tilde{a}_{sk})_{k=1,\cdots,P}\}_{s \in S}$ (figure 7.3). The output of this classifier is a segmentation map $\{\hat{x}_s\}_{s \in S}$ which contains spatially homogeneous classes of pixels with similar spectral behavior [F.F 05].

7.2.4. *Results on NGC 4254 Cube*

We have applied the dimensionality reduction scheme on the NGC 4254 cube, which contains $N = 42$ bands of size $H = W = 512$ pixels. As explained in section 7.2.2 we fixed the number of Gaussian components to $P = 6$. It turned out that intensity in all velocities corresponding to channel numbers 10 to 35 are equally present in the data cube and that the widths of the corresponding lines do not vary significantly. Therefore the selected means are almost equidistant while variances have similar values (means are given in channel number) :

$$\begin{pmatrix} c_k \\ w_k \end{pmatrix} \in \left\{ \begin{pmatrix} 12.3 \\ 2.25 \end{pmatrix}, \begin{pmatrix} 16.0 \\ 2.25 \end{pmatrix}, \begin{pmatrix} 20.3 \\ 2.56 \end{pmatrix}, \begin{pmatrix} 25.0 \\ 2.25 \end{pmatrix}, \begin{pmatrix} 29.2 \\ 1.96 \end{pmatrix}, \begin{pmatrix} 33.0 \\ 1.44 \end{pmatrix} \right\}.$$

In this way the initial 42 velocity channels are reduced to 6 effective channels. The basis selection algorithm (section 7.2.2) ensures that we do not lose important information.

Figure 7.4 shows the multivariate image (6 images in inverse video of the weights of the 6 Gaussians with fixed variances and mean values) obtained with our reduction technique. Due to the fact that the mean values of the basis are approximately equidistant, the multivariate images look like binned channel maps. These reduced images feed the Markovian classifier, allowing us to isolate regions of similar spectra. The most adapted number of classes depends on the complexity of the data cube. Selecting an overly small number of classes results in loss of information, selecting an overly large number leads to classes without physical meaning, which are sometimes made up of dispersed patches on the image. It is beyond the scope of this chapter to discuss how to determine in general the most appropriate number of classes. However, the expert may have an expectation for the number of classes of interest. One then may explore different classification solutions with a variable number of classes around this expected value. In our case we found that 7 classes are sufficient to describe the main features of the NGC 4254 data cube.

The final segmentation map of 7 classes together with the average observed and model spectra for each class is shown in figure 7.5. In general, the observed spectra are well fitted by a weighted combination of our Gaussian basis functions. The fact that the average observed spectra for all classes are single peaked shows that we actually

./ChangeDetection/mazet/figs/EM6Gclip_1.eps   ./ChangeDetection/mazet/figs/EM6Gclip_2.eps

(1)   (2)

./ChangeDetection/mazet/figs/EM6Gclip_3.eps   ./ChangeDetection/mazet/figs/EM6Gclip_4.eps

(3)   (4)

./ChangeDetection/mazet/figs/EM6Gclip_5.eps   ./ChangeDetection/mazet/figs/EM6Gclip_6.eps

(5)   (6)
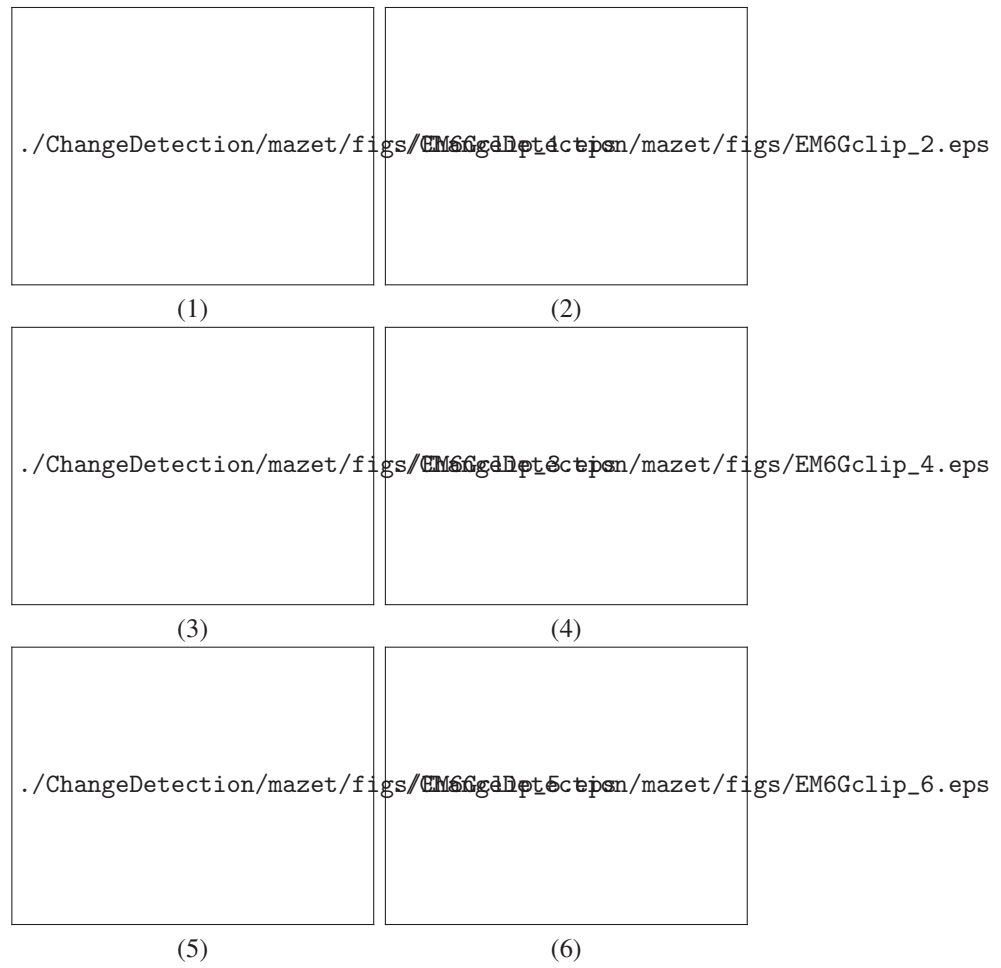
Figure 7.4. *Maps of the weights of the 6 Gaussians with fixed mean values and variances. In the case of NGC 4254, this corresponds to a reduction of the original 42 channels to 6 effective channels. The x and y axis are the astronomical coordinates.*

obtain classes of distinct spectral line profiles with this first method. The fine structure of the model line profiles is due to the limited number of Gaussians of the basis. We consider that the estimation is acceptable since the widths of the fine structure of the average model spectra are always smaller than the widths of the average observed spectrum. Finally, the important information is the average observed line profile.

The comparison between the segmentation map and the velocity field (right panel of figure 7.1) shows that the segmentation is mainly done according to velocity, *i.e.*,
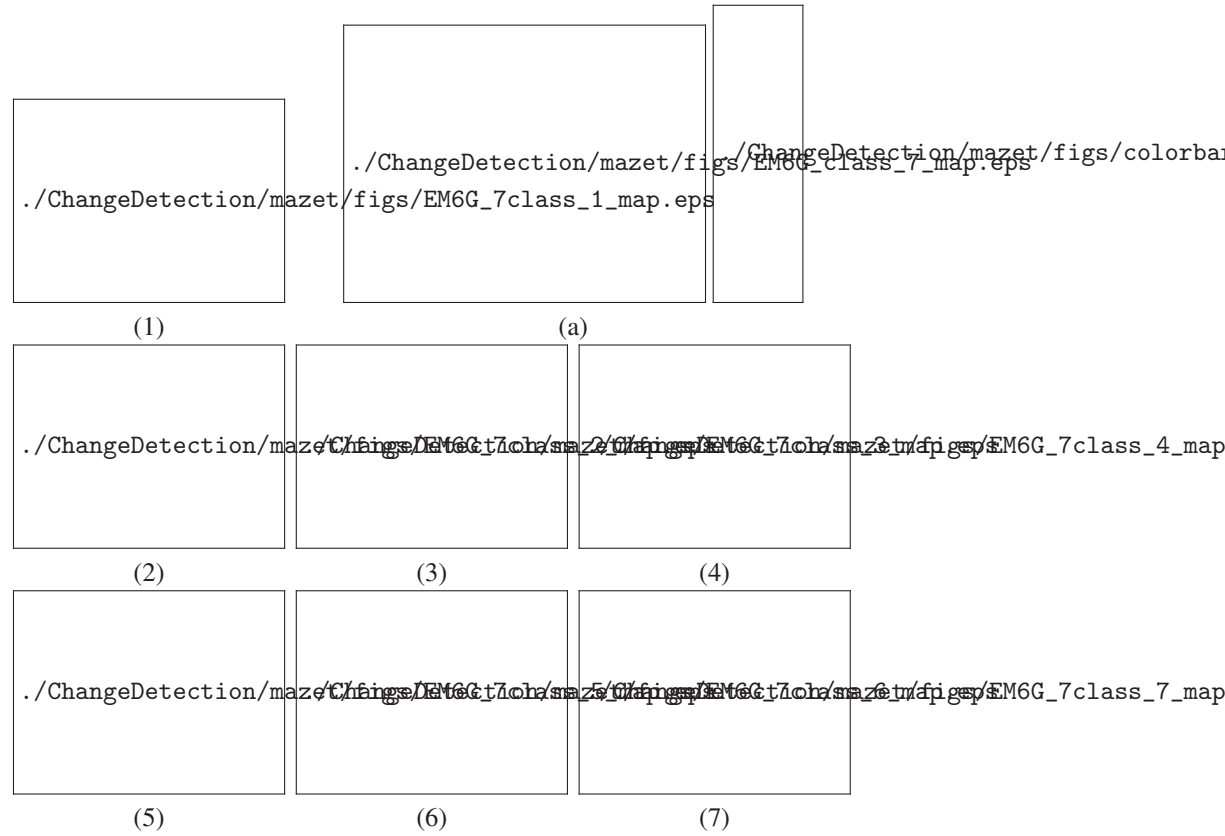
Figure 7.5. *Results of segmentation of the maps shown in figure 7.4 : (a) segmentation map. The x and y axis are the astronomical coordinates. (1-7) average spectra for each class (solid : observed spectrum, dashed : model spectrum). The velocity channels are represented on the x axis, the intensity on the y axis (in Jy). The peaks of the line profiles are indicated by an arrow.*

the position of the peak in the 1D spectrum. This is due to the nature of the basis (almost equidistant in velocity ; see section 7.2.4). Only the peak of the averaged line profile of class 4 lies within the full width half maximum (FWHM) of the averaged line profile of class 5. Here the segmentation is also based on the peak value of the line profile which is much smaller for the line profiles of class 4 than for those of class 5. Thus for a sufficiently large number of classes (equal to or larger than the number of basis Gaussians) the segmentation is also based on the peak intensity of the line profiles. Two perturbations of the velocity field in the northern part of the galaxy can be identified (right panel of figure 7.5) :

1) a departure from the symmetric velocity gradient near pixel (70,50) where the red/yellow region extends into the green region ;

2) the blue region in the north of the galaxy.

Both regions are recognized as distinct classes (1 and 4) in figure 7.5. We conclude with this first approach that our data processing method is operating successfully.

For this relatively simple data cube, the segmentation map (figure 7.5) does not contain additional information to the velocity field (figure 7.1 right panel). This segmentation map might be used in a first step to produce masks allowing to isolate the kinematically interesting regions in the north of the galaxy (class 1 and 4). Nevertheless, this data cube corresponds to a simple case and the kinematic behavior of galaxies is often more complex. Indeed, the kinematic map could exhibit inhomogeneous areas (multiple emission components within the gas structures, *e.g.* double line profile). So, the next section presents a more advanced processing scheme to deal with such challenge.

7.2.5. *Double Line Profiles*

We have seen that the proposed segmentation method produces a discrete velocity field and masks the regions of asymmetries in the velocity field. The principal result is thus the labeling of different regions of interest. All these regions can already be approximately separated by eye on the moment maps (figure 7.7). However, a feature that cannot be detected easily by the moment maps is double line profiles. This information is lost by the velocity averaging process when the moment maps are produced. In order to investigate whether our segmentation method is able to detect regions of double line profiles, we added an artificial line to the spectra in a circular region north of the galaxy center (with its center position (60, 35), a radius of 10 pixels and a center velocity at channel number 28). The intensity of the artificial line is maximum at the center and falls off radially (we used a Gaussian profile with a width of 5 pixels). Figure 7.9 shows the moment maps of this new data cube. The additional lines produce a local maximum in the HI emission map (left panel) and a pronounced asymmetry in the velocity field (right panel).

The basis for this modified data cube is :

$$\begin{pmatrix} c_k \\ w_k \end{pmatrix} \in \left\{ \begin{pmatrix} 12.3 \\ 2.25 \end{pmatrix}, \begin{pmatrix} 16.0 \\ 2.25 \end{pmatrix}, \begin{pmatrix} 20.2 \\ 2.56 \end{pmatrix}, \begin{pmatrix} 24.8 \\ 2.56 \end{pmatrix}, \begin{pmatrix} 29.0 \\ 2.25 \end{pmatrix}, \begin{pmatrix} 33.0 \\ 1.44 \end{pmatrix} \right\}.$$

It is almost identical to that of the original data cube. Figure 7.8 shows the multivariate image (6 images in inverse video of the weights of the 6 Gaussians with fixed variances and mean values) obtained with our reduction technique on the modified

(1)   (a)

(2)   (3)   (4)

(5)   (6)   (7)
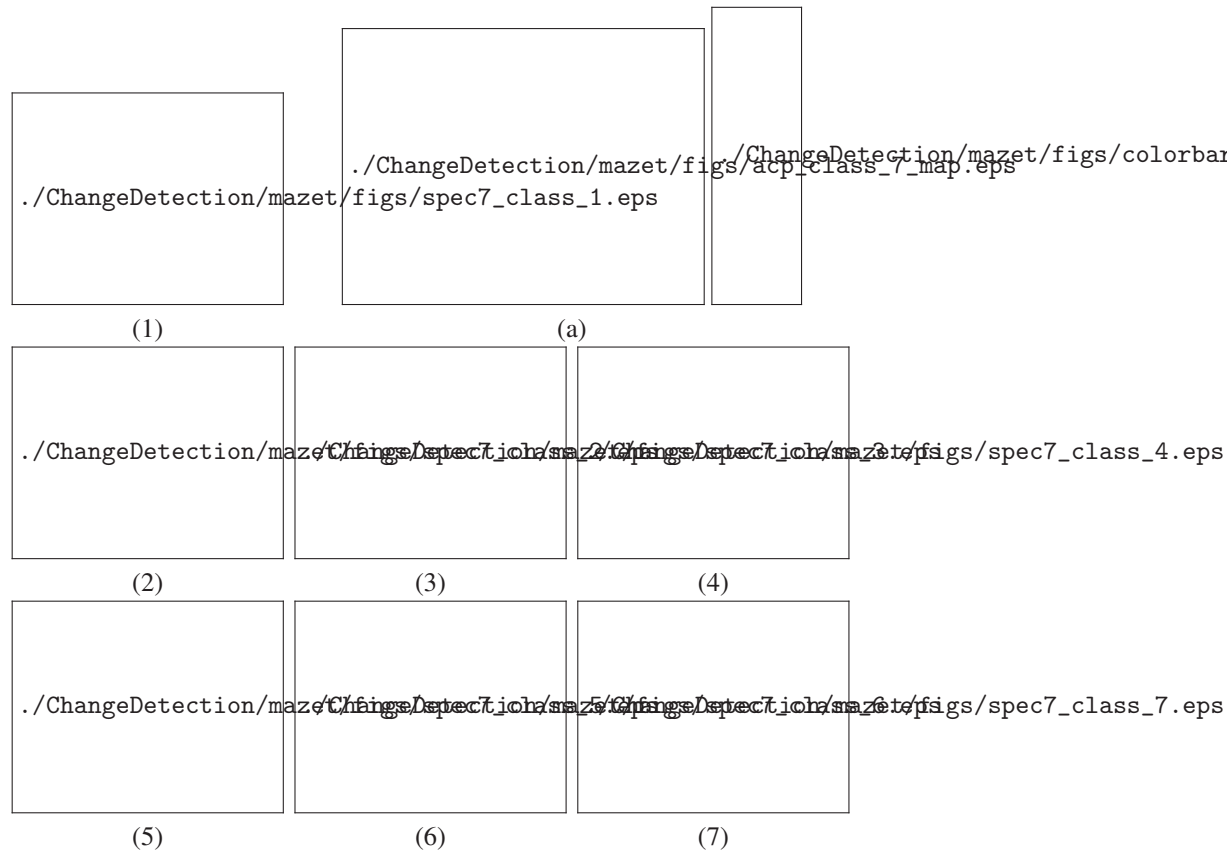
Figure 7.6. *Principal Component Analysis (PCA) of NGC 4254. Only the images corresponding to the six largest eigenvalues are kept. (a) segmentation map. The x and y axis are the astronomical coordinates. (1-7) average spectra for each class (solid : observed spectrum, dashed : model spectrum). The velocity channels are represented on the x axis, the intensity on the y axis (in Jy).*

./ChangeDetection/mazet/figs/Cont2raie.eps   ./ChangeDetection/mazet/figs/vitesse2raie.eps

(1)   (2)
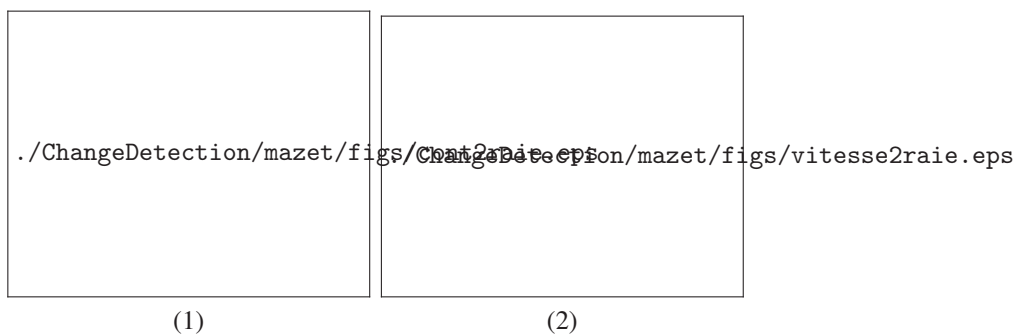
Figure 7.7. *NGC 4254 data cube after addition of a region with an artificial line creating a double line profile (center at (60, 35)). Left panel : H*I *emission distribution (moment 0). The x-axis corresponds to right ascension, the y-axis to declination. The intensity is given in mJy. Right panel : H*I *velocity field (moment 1). The velocities are given in channel maps.*

data cube. The final segmentation map of this modified data cube together with the average observed and model spectra is presented in figure 7.9. Again, in the inner part of the galaxy a discrete velocity field is produced. The northern perturbed part of the velocity field is also recognized (class 1). The region of double line profiles appears as a new separate class (5). The averaged observed and fitted line profiles clearly show a double peak with a component of class 1 (central channel number : 17) and the new, artificially added component (central channel number : 28). This double line profile is clearly distinct from the fine structure of the modeled lines, which is always contained within the width of the observed line profile. Thus, our method is able to detect regions of double line profiles. The 42 channel maps of the original cube can then be inspected in detail using the mask of the double line region.

The double line profile could have been detected by the inspection of the channel maps by eye and/or by making multiple position–velocity plots. However, without a prior knowledge of the location of the double line profile this can be quite long and painful, especially if the cube is more complex than our test cube. Despite the loss of spectral resolution (the cube is reduced from 42 channels to 6 effective channels), our method is still able to detect and mask multiple line profiles. We think that this is very useful for an astronomer who has to handle complex 3D data cubes. The further investigation is simplified by the knowledge and possible masking of regions of multiple line profiles, which might be carried out with more sophisticated methods applied to the original data cube.

(1)

(2)

(3)

(4)

(5)

(6)
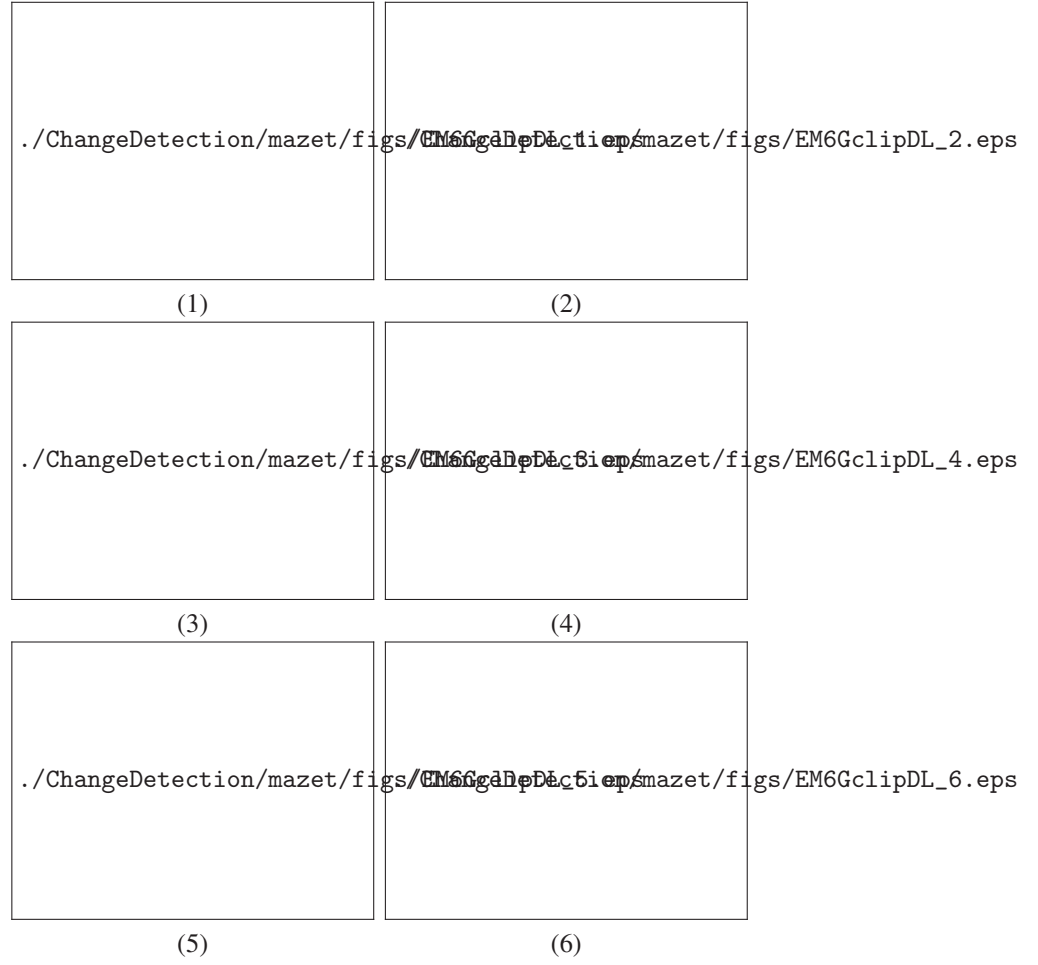
Figure 7.8. *Maps of the weights of the 6 Gaussians with fixed mean values and variances. A region of artificial line profiles creating a region of double line profiles is added (center at (60, 35)), which is clearly visible in map (1). The x and y axis are the astronomical coordinates.*

(1)

(a)

(4)

(2)

(3)

(5)

(6)

(7)
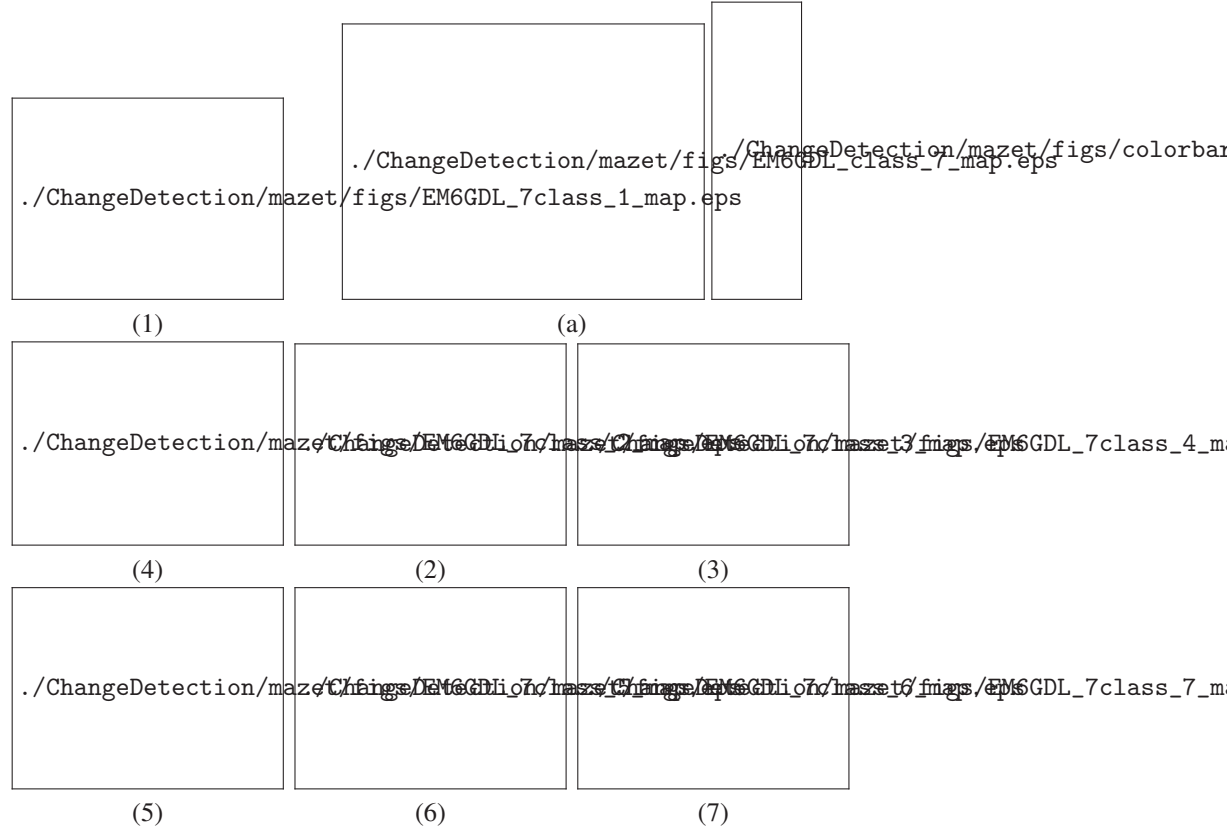
Figure 7.9. *Results of segmentation of the maps shown in figure 7.8 : (a) segmentation map. The x and y axis are the astronomical coordinates. (1-7) average spectra for each class (solid : observed spectrum, dashed : model spectrum). The velocity channels are represented on the x axis, the intensity on the y axis (in Jy). The peaks of the line profiles are indicated by an arrow.*

7.2.6. *Conclusion*

In this first method we propose to analyse data cubes by fitting for each observed spectrum a weighted combination of Gaussian components. The weight parameters of the Gaussians are the input of a Markovian segmentation algorithm. Due to the curse of dimensionality (Hughes phenomenon), the number of input parameters is limited to ten. In our approach, we chose to fix the number of Gaussian functions to 6 (the optimum number of classes depends on the complexity of the data cube) and to set

their mean values and variances to their 6 most representative values. The weights of these 6 Gaussians with fixed mean value and variance are determined by fitting the observed spectra on the selected basis. In this way the original number of channels of the data cube is reduced to 6 effective channels. A Markovian image segmentation is then done on the 6 maps of the weights of the Gaussians [F.F 05]. The final result is a segmentation map where regions of similar spectral line profiles are assembled into different classes.

7.3. Spectral Decomposition for Component Classification

7.3.1. *Introduction*

The approach presented in section 7.2 provides a map of the homogeneous kinematic zones in the galaxy, that is a segmentation showing the spatial regions with a similar kinematic behavior. However, it is unable to distinguish different gas structures, that is, two superimposed structures cannot be separated since each pixel of the image belongs to a single class. On the contrary, on structure may be split into different kinematic zones. This is clearly shown on figure 7.9 where the double line region is seen as a separate class, although the two lines are not distinguished. Besides, the great structure is split into different regions while it is a unique gas structure. In that respect, we have developed a second approach aiming at detecting the gas structures instead of the regions of same kinematic behavior. In this new approach, the classification is done by tracking the estimated components through the cube. This is not possible with the first decomposition algorithm because each pixel is decomposed into six components with identical positions and widths. Therefore, the second approach also have a new decomposition step in which the component positions and widths are estimated.

The first step of this approach (section 7.3.2) decomposes each pixel in a sum of Gaussian components whose number, positions, amplitudes and widths have to be estimated. These variables may differ between two pixels ; in other words, the decomposition is not set on a unique basis. This is a major difference with the previous approach which allows to track the components. The second step (section 7.3.3) classifies each estimated component with the assumption that two components in two neighbbouring pixels are considered in the same class if their parameters are similar.

We keep the same notations as in section 7.2 : let a multispectral image with N channels defined on a regular lattice S of size $H \times W$ pixels. Each of the $D = H \times W$ pixels $s \in S$ of the image is a spectrum y_s of length N.

7.3.2. *Spectrum Decomposition*

Clearly, the decomposition of a spectrum is an ill-posed inverse problem [IDI 08] ; this is reinfored by the non-linearity and the non-constant dimension model (see equation (7.7)). Thus, the problem is set in the Bayesian framework using a hierarchical model which allows to set priors on the hyperparameters instead of fixing them arbitrarily [GIL 99, ROB 02]. In that respect, there is a substantial gain in modeling the components by parametric functions since it reduces the number of unknowns, *i.e.* the solution space. Nevertheless, we do not consider any spatial prior in the decomposition step : this information will be used in the classification step (section 7.3.3). Besides, we are interesting in this step to decompose the whole cube : this allows the hyperparameters to be the same in each spectrum. For example, a single noise variance is considered allover the spectra ; this is a realistic assumption because the noise is stationary in radioastronomy.

The proposed model yields a complex posterior distribution which is expected to have many local minima. To handle this problem, we propose to use Monte Carlo Markov chain (MCMC) methods [GIL 99, ROB 02] to simulate the posterior. In fact, we use the so-called trans-dimensional MCMC methods [GRE 95, RIC 97] because the dimension of the problem is not constant. Indeed, the component number is unknown and has to be estimated, so the variable number is itself unknown and may change : the problem is said to be a non-constant dimension model. At last, to reduce the computation time, the spectra containing only noise are not decomposed (indeed, the cosmological background does not carry any information about the gas kinematic). To do that, a mask is computed on the multispectral image before decomposition to discard the spectra containing only noise. The mask is determined by clipping the cube at a level of 3σ (σ being the rms noise). A morphological dilation is then applied to connect close regions in the mask (a disk of diameter 7 pixels is chosen for structuring element).

Bayesian Model

The problem of decomposing a radio spectrum is typically an object extraction problem which consists in decomposing the data as a noisy sum of elementary "patterns"[3] modeling the components. These ones are modeled by a known parametric function f with unknown parameters (*e.g.* the pattern amplitude or position).

---

3. It is noteworthy that the problem also can be modelled as a marked point process [MØ 04]. In this case, a Poisson point process would be convenient because the components are supposed to be independent each other (there is no repulsive or attraction behavior). However, we think that this framework is too complex compared to the proposed model.

Each spectrum y_s ($s \in S$) is modeled as a noisy sum of P_s components :

$$y_s = \sum_{k=1}^{P_s} a_{sk} f(c_{sk}, w_{sk}) + e_s, \qquad (7.7)$$

where f is a vector function of length N and e_s is a $N \times 1$ vector modeling the noise. We stress that, contrary to the previous approach, the spectrum does not represent a probability density function. Furthermore, we suppose that the component number P_s is itself unknown and has to be estimated. This is a major difference with the first approach yielding a fine decomposition of the spectra (see figure 7.11).

The usual model in radioastronomy assumes that the components are Gaussian with position c_{sk}, amplitude a_{sk} and size w_{sk}. So, the vector function f for component $k \in \{1, \ldots, P_s\}$ in pixel $s \in S$ at frequency channel $j \in \{1, \ldots, N\}$ equals :

$$f_j(c_{sk}, w_{sk}) = \exp\left(-\frac{(j - c_{sk})^2}{2w_{sk}^2}\right).$$

For simplicity, the expression of a Gaussian function was multiplied by $\sqrt{2\pi w_{sk}^2}$ so that a_{sk} corresponds to the maximum (the amplitude) of the component. In addition, we have $\forall s, k, a_{sk} \geq 0$ because the components are supposed to be non-negative.

This model assumes that the components are perfect Gaussians while real lines may sometimes be asymmetric. In that case, an asymmetric component will be estimated by one large pattern and another (or many other) smaller. The solution has no physical sense, but the estimated structure seems to be not affected. Thus, adding a "skewness parameter" in the expression of f might be an interesting solution, but we think this would lead to a more complex model without improving significantly the results compared to the actual results.

Equation (7.7) also reads in matrix form :

$$y_s = F_s a_s + e_s \qquad (7.8)$$

where F_s is a $N \times P_s$ matrix and a_s is a $P_s \times 1$ vector :

$$F_s = \begin{pmatrix} f_1(c_{s1}, w_{s1}) & \cdots & f_1(c_{sP_s}, w_{sP_s}) \\ \vdots & & \vdots \\ f_N(c_{s1}, w_{s1}) & \cdots & f_N(c_{sP_s}, w_{sP_s}) \end{pmatrix} \quad \text{and} \quad a_s = \begin{pmatrix} a_{s1} \\ \vdots \\ a_{sP_s} \end{pmatrix}.$$

Prior Distributions

To get an integrable posterior distribution, we choose to set only proper priors (*i.e.* integrable priors) on the unknowns and the hyperparameters. Some priors are

conjugate so as to get conditional posterior distributions easy to simulate [ROB 02]. Especially, we choose conjugate priors for variances r_a and r_e (see further) because we do not have any prior information on these parameters. We also try to use common priors for which simulation algorithms are available [ROB 02].

The following priors stand for each $s \in S$ and $k \in \{1, \ldots, P_s\}$.

- The prior model is specified by supposing that P_s is drawn from a Poisson distribution [GRE 95] :

$$p(P_s) = e^{-\lambda} \frac{\lambda^{P_s}}{P_s!} \mathbb{1}_{\mathbb{N}}(P_s)$$

where λ can be interpreted as the expected number of components in the spectra (we choose $\lambda = 1$) and $\mathbb{1}_E(x)$ is the indicator function (that is, $\mathbb{1}_E(x) = 1$ if $x \in E$, 0 otherwise) ;

- the noise e_s is supposed to be white, zero-mean Gaussian, independent and identically distributed with variance r_e :

$$e_s \sim \mathcal{N}(0, r_e);$$

- we do not have any information about the component locations, we can only suppose that they are spread over the spectra. So, component locations c_{sk} are supposed uniformly distributed on $[1; N]$:

$$c_{sk} \sim \mathcal{U}_{[1;N]}(c_{sk});$$

- component amplitudes a_{sk} are distributed according to a (conjugate) Gaussian distribution with variance r_a and truncated in zero to get positive amplitudes. We note in the sequel $a_{sk} \sim \mathcal{N}^+(0, r_a)$ where $\mathcal{N}^+(\mu, \sigma^2)$ stands for a Gaussian distribution with positive support defined as :

$$p(x|\mu, \sigma) = \sqrt{\frac{2}{\pi\sigma^2}} \left[1 + \mathrm{erf}\left(\frac{\mu}{\sqrt{2\sigma^2}} \right) \right]^{-1} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2} \right) \mathbb{1}_{[0;+\infty[}(x)$$

and erf is the error function ;

- we choose an inverse gamma prior $\mathcal{IG}(\alpha_w, \beta_w)$ for component width w_{sk} because this is a positive-support distribution whose parameters α_w and β_w can be set according to the approximate component width known a priori. Thus, we suppose that the component width is 5, but, because this value is very approximative, we also suppose a large variance (equals to 100), yielding $\alpha_w \approx 2$ and $\beta_w \approx 5$;

- the hyperparameter r_a is distributed according to a conjugate (inverse gamma) prior $\mathcal{IG}(\alpha_a, \beta_a)$. This prior aims at avoiding numerical errors but not at influencing the estimation, so it has to be the less informative[4]. We propose to set the mean to the approximate real component amplitude (say ν) which can be roughly estimated, and to assign a large value to the variance. This yields :

$$\alpha_a = 2 + \varepsilon \quad \text{and} \quad \beta_a = \nu + \varepsilon \quad \text{with} \quad \varepsilon \ll 1;$$

- a common choice for noise variance prior is Jeffreys distribution, which is unfortunately improper. Therefore, we adopt again an inverse gamma prior $\mathcal{IG}(\alpha_e, \beta_e)$ for r_e, whose parameters are both set close to zero ($\alpha_e = \beta_e = \zeta$, with $\zeta \ll 1$) which is in fact the limit case corresponding to Jeffreys prior.

The directed acyclic graph showing the dependencies between the variables is displayed in figure 7.10. Squares represent fixed or observed quantities while circles represent the unknowns, and dotted (respectively plain) arrows represent the determinist (respectively probabilistic) dependencies.
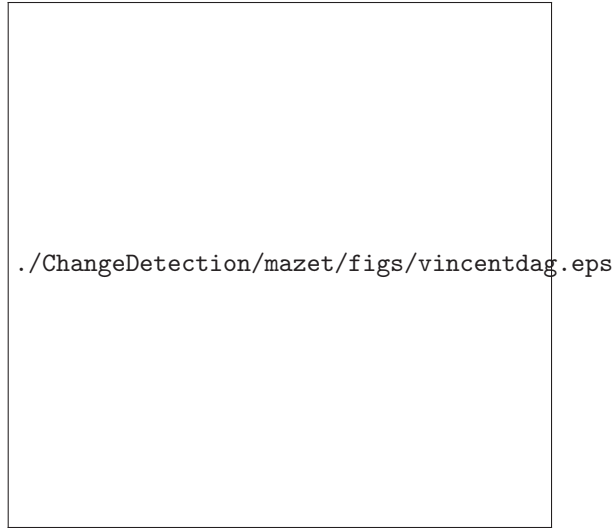
Figure 7.10. *Directed acyclic graph specific to the model.*

------

4. A uniform distribution, which is often used as a non-informative prior and corresponds to $\alpha_a = \beta_a = 0$ cannot be used because it is improper (*i.e.* non integrable).

Posterior Distributions

The posterior density integrability cannot been checked mathematically because of its complexity, but, as the priors are proper, a sufficient condition to prove the posterior integrability is fulfilled. The conditional posterior distributions of each unknown is obtained thanks to the prior defined above.

$$c_{sk}|\cdots \propto \exp\left(-\frac{1}{2r_e}\|y_s - F_s a_s\|^2\right) \mathbb{1}_{[1,N]}(c_{sk}),$$

$$a_{sk}|\cdots \sim \mathcal{N}^+(\mu_{sk}, \rho_{sk}),$$

$$w_{sk}|\cdots \propto \exp\left(-\frac{1}{2r_e}\|y_s - F_s a_s\|^2 - \frac{\beta_w}{w_{sk}}\right)\frac{1}{w_{sk}^{\alpha_w+1}}\mathbb{1}_{[0;+\infty[}(w_{sk}),$$

$$r_a|\cdots \sim \mathcal{IG}\left(\frac{L}{2} + \alpha_a, \frac{1}{2}\sum_s\sum_{k=1}^{P_s}a_{sk}^2 + \beta_w\right),$$

$$r_e|\cdots \sim \mathcal{IG}\left(\frac{NM}{2} + \alpha_e, \frac{1}{2}\sum_s\|y_s - F_s a_s\|^2 + \beta_e\right)$$

where $x|\cdots$ means x conditionally to y and the other variables and :

$$\mu_{sk} = \frac{\rho_{sk}}{r_e}z_{sk}^T F_{sk}, \qquad \rho_{sk} = \frac{r_a r_e}{r_e + r_a F_{sk}^T F_{sk}}, \qquad z_{sk} = y_s - F_s a_s + F_{sk}a_{sk}$$

and N is the spectrum length, M is the number of processed pixels, L denotes the total component number (*i.e.* $L = \sum_s P_s$) and F_{sk} corresponds to the kth column of matrix F_s.

The conditional posterior expressions for c_{sk} and w_{sk} as well as the hyperparameters are straightforward considering the global posterior. On the contrary, the conditional posterior for w_{sk} is not easy to obtain ; the detail of the computation can be found in [MAZ 05a, Appendix B].

Estimation

Because the posterior is a high dimensional complex density, usual optimisation techniques fail to provide a satisfactory solution. So, we propose to use MCMC methods to achieve the estimation problem [GIL 99, ROB 02] : they are efficient techniques for drawing samples X from a distribution π by generating a sequence of realizations $\{X^{(i)}\}$ through a Markov chain having π as its stationary distribution. Then, Monte Carlo integration approximates an estimator $E[\psi(X)]$ by :

$$E[\psi(X)] \approx \frac{1}{I}\sum_{i=1}^{I}\psi(X^{(i)}).$$

Besides, we need a MCMC method which is able to handle a variable dimension model. Indeed, the component number is unknown and has to be estimated : this is a condition to provide a fine decomposition for each spectrum. Some MCMC methods dealing with models of variable dimension have been proposed during the last decade : they are known as trans-dimensional MCMC methods. Among them, the reversible jump MCMC algorithm [GRE 95, RIC 97] appears to be popular, fast and flexible [CAP 03]. At each iteration of this algorithm, a move which can either change the model dimension or generate a random variable is randomly performed. Following recommendations of Cappé *et al.* [CAP 03], we propose these moves :

 – B_s ($s \in S$) "birth in s" : a component is created in pixel s ;
 – D_s ($s \in S$) "death in s" : a component is deleted in pixel s ;
 – U_s ($s \in S$) "update in s" : variables c_s, a_s and w_s are updated ;
 – H "hyperparameter update" : hyperparameters r_a and r_e are updated.

Only the moves B_s and D_s may change the model dimension. Let b_s, d_s, u_s and h be respectively the probabilities for moves B_s, D_s, U_s and H, they are chosen so that :

$$b_s = \frac{\gamma}{M+1} \min\left\{1, \frac{p(P_s+1)}{p(P_s)}\right\} \qquad d_s = \frac{\gamma}{M+1} \min\left\{1, \frac{p(P_s-1)}{p(P_s)}\right\}$$

$$u_s = \frac{1}{M+1} - b_s - d_s \qquad\qquad h = \frac{1}{M+1}$$

with γ such that $b_s + d_s \leq 0.9$ (we choose $\gamma = 0.45$) and $d_s = 0$ if $P_s = 0$.

We now consider the sampling method for generating random variables following the posterior distributions. Many methods available in litterature are used for sampling positive normal [MAZ 05b] and inverse gamma distributions [ROB 02, DEV 86]. Otherwise, c_{sk} and w_{sk} are sampled using a random-walk Metropolis-Hastings algorithm [ROB 02, GIL 99]. To improve the algorithm speed, they are sampled jointly to avoid calling the likelihood twice because it is costly in computing time. The proposal distribution is a (separable) truncated Gaussian centered on the current value :

$$\tilde{c}_{sk} \sim \mathcal{N}\left(c_{sk}^*, r_c\right) \mathbb{I}_{[1,N]}(\tilde{c}_{sk}), \qquad \tilde{w}_{sk} \sim \mathcal{N}^+\left(w_{sk}^*, r_w\right)$$

where $\tilde{\cdot}$ stands for the proposal and $\cdot^*$ denotes the current value. The algorithm efficiency depends on the scaling parameters r_c and r_w chosen by the user (generally with heuristics methods) ; Gelman et al. [GEL 96] recommend to look for obtaining an acceptance rate of 0.5. At last, the acceptance test is computed with the logarithm of the acceptance rate to avoid numerical errors.

Finally, the estimation is computed by picking in the generated Markov chain the sample which minimises the mean square error : this is a very simple estimation of the maximum a posteriori which does not need to save the Markov chains. Indeed, the number of unknowns, and as a result, the number of Markov chains to save, is prohibitive.

Example of Decomposition

To illustrate the relevance of this approach, the decomposition of spectrum $(222, 278)$ of NGC 4254 cube is shown in figure 7.11. The algorithm ran for $I = 5000$ iterations and the expected component number has been set to $\lambda = 1$ to favor a low number of component. The spectrum is denoted by the gray line and the decomposition by the black lines (dashed lines correspond to each estimated component and the plain line to their sum). The estimated component number is $\widehat{P} = 3$. We also plot the residual

./ChangeDetection/mazet/figs/vincentexample.eps
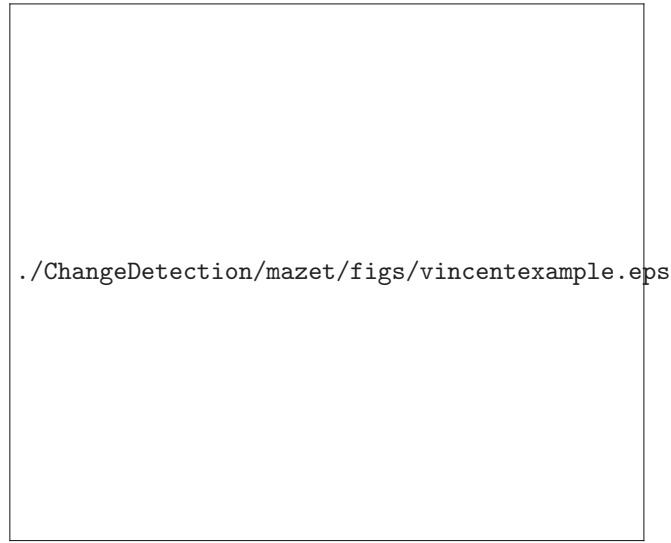
Figure 7.11. *Top : spectrum* $(222, 278)$ *in NGC 4254 cube (gray line) and its decomposition (black lines). Bottom : the corresponding residual.*

between the data and the estimation, showing that the decomposition is relevant.

7.3.3. *Classification*

The goal of this second approach is to detect the different gas structures in the galaxy. The decomposition method provides for each spectrum P_s components with parameter $x_{sk} = \{c_{sk}, a_{sk}, w_{sk}\}$. Thus, we aim in this step to assign a class (or index) $q_{sk} \in \mathbb{N}^*$ to each estimated component k in spectrum s. Each class represents one gas structure, so that components with the same class belong to the same gas structure.

First of all, the components whose amplitude is lower than a predefined threshold τ are vanished in the following procedure (this condition helps the astronomer to analyse the gas location with respect to the intensity). To perform the classification, we

assume that the component parameters exhibit weak variation between two neighbouring spectra, *i.e.* two components in two neighbouring spectra are considered in the same class if their parameters are close. The spatial dependency is introduced by defining a Gibbs field over the decomposed image [**?**] :

$$p(q|x) = \frac{1}{Z} \exp\left(-U(q|x)\right) = \frac{1}{Z} \exp\left(-\sum_{c \in \mathcal{C}} U_c(q|x)\right) \tag{7.9}$$

where Z is the partition function, $\mathcal{C}$ gathers the cliques of order 2 in a 4-connexity system and the potential function is defined as the total cost of the classification.

Let consider one component (s, k) located in spectrum $s \in S$ ($k \in \{1, \ldots, P_s\}$), and a neighbouring pixel $t \in S$. Then, the component (s, k) may be classified with a component (t, l) ($l \in \{1, \ldots, P_t\}$) if their parameters are similar. In this case, we define the cost of component (s, k) equals to a distance $D(x_{sk}, x_{tl})^2$ computed with the component parameters (we see further why we choose the square of the distance). On the contrary, if no component in spectrum t is close enough to component (s, k), we choose to set the cost of the component to a threshold T^2 which codes the weaker similarity allowed. Indeed, if the two components (s, k) and (t, l) are too different (that is $D(x_{sk}, x_{tl})^2 > T^2$), it would be less costly to let them in different classes. Finally, the total cost of the classification (*i.e.* the potential function) corresponds to the sum of the component costs.

Formally, these considerations read in the following manner. The potential function is defined as :

$$U_c(q|x) = \sum_{k=1}^{P_s} \varphi(x_{sk}, q_{sk}, x_t, q_t) \tag{7.10}$$

where s and t are the two spectra involved in the clique c, and $\varphi(x_{sk}, q_{sk}, x_t, q_t)$ represents the cost associated for the component (s, k) and defined as :

$$\varphi(x_{sk}, q_{sk}, x_t, q_t) = \begin{cases} D(x_{sk}, x_{tl})^2 & \text{if } \exists\, l \text{ such that } q_{sk} = q_{tl}, \\ T^2 & \text{otherwise.} \end{cases} \tag{7.11}$$

In some ways, $\varphi(x_{sk}, q_{sk}, x_t, q_t)$ can be seen as a truncated quadratic function which is known to be very appealing in the context of outliers detection [**?**].

We choose for the distance $D(x_{sk}, x_{tl})$ a normalized Euclidean distance :

$$D(x_{sk}, x_{tl}) = \sqrt{\left(\frac{c_{sk} - c_{tl}}{\delta_c}\right)^2 + \left(\frac{a_{sk} - a_{tl}}{\delta_a}\right)^2 + \left(\frac{w_{sk} - w_{tl}}{\delta_w}\right)^2}. \tag{7.12}$$

The distance is normalized because the three parameters have not the same unity. δ_c and δ_w are the normalizing factors in the frequency domain whereas δ_a is the one in

the intensity domain. We consider that two components are similar if their positions or widths do not differ for more than 1.2 wavelength channel, or if the difference between the amplitudes do not exceed 40% of the maximal amplitude. So, we set $\delta_c = \delta_w = 1.2$, $\delta_a = \max(a_{sk}, a_{s'k'}) \times 40\%$ and $T = 1$.

To resume, we look for :

$$\hat{q} = \arg\max_q p(q|x) \quad \Leftrightarrow \quad \hat{q} = \arg\min_q \sum_{c \in \mathcal{C}} \sum_{k=1}^{P_s} \varphi(x_{sk}, q_{sk}, x_t, q_t). \qquad (7.13)$$

We propose a greedy algorithm to perform the classification because it yields good results in an acceptable computation time (about 36 s on the cube considered in section 7.3.4 containing 9463 processed pixels). The algorithm is presented below. The main idea consists in tracking the components through the image by starting from an initial component and looking for the components with similar parameters spectrum by spectrum. These components are then classified in the same class, and the algorithm starts again until every estimated component is classified. Let z^* be the increasing index coding the class, and the set $\mathcal{L}$ gathers the estimated components to classify.

1. set $z^* = 0$

2. while it exists some components that are not yet classified :

    3. $z^* = z^* + 1$

    4. choose randomly a component (s, k)

    5. set $\mathcal{L} = \{(s, k)\}$

    6. while $\mathcal{L}$ is not empty :

        7. set (s, k) as the first element of $\mathcal{L}$

        8. set $q_{sk} = z^*$

        9. delete component (s, k) from $\mathcal{L}$

        10. among the 4 neighbouring pixels t of s, choose the components l that satisfy the following conditions :

            (C1) they are not yet classified

            (C2) they are similar to component (s, k) that is $D(x_{sk}, x_{tl})^2 < T^2$

            (C3) their amplitude is greater than τ

        11. Add (t, l) to $\mathcal{L}$

7.3.4. *Results on NGC 4254 Cube*

We applied the proposed approach on the NGC 4254 cube already presented in section 7.1. The algorithm ran for 5000 iterations and the variables are initialized by simulating them from the priors. The processing was carried out using Matlab on a double core (each 3.8 GHz) PC with 4 Gb of RAM and takes 337 minutes.

The greatest number of estimated components in a spectrum equals 5 (figure 7.12). The estimation is very satisfactory because the difference between the original and the estimated cubes is very small (figure 7.13); this is confirmed by watching every spectrum decomposition (not shown here).

./ChangeDetection/mazet/figs/vincentnbmotifs1.eps

Figure 7.12. *Map of the estimated component number.*

./ChangeDetection/mazet/figs/vincentimgrec1.eps

Figure 7.13. *Original (left) and estimated (right) moment 0 of the cube*
NGC 4254 (in inverse video). The mask border is depicted by the black line.

./ChangeDetection/mazet/figs/vincentimgclass1.eps
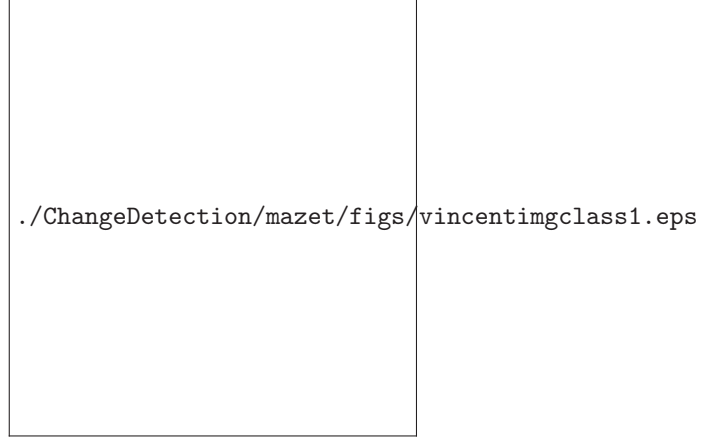
Figure 7.14. *Map of the two main classes of NGC 4254 (threshold t equals 0,01 times the maximum intensity).*

The estimated components are then classified into 8668 classes, but the majority are very small and, consequently, not significant. In fact, only two classes, gathering more than 100 components each one, are relevant : the large central structure and the "comma" shape in the bottom-left (figure 7.14). Thus, this new approach operates successfully since it is able to distinguish clearly the gas structures in the galaxy.

7.3.5. *Double Line Profiles*

We also want to assess the performances of the proposed approach in the case of a double line profile which cannot be detected easily by the moment maps (see figure 7.7). Again, the algorithm ran for 5000 iterations. As for the previous data cube, the decomposition is very good and the error between the original and estimated cubes is very small. The artificially added component appears clearly as a third relevant class (figure 7.15).

The analysis of the emission distribution (moment 0) and the velocity field (moment 1) of the three classes is also instructive (figure 7.16). Indeed, the velocity field of the large central structure shows a rotating disk. As well, the emission distribution of the artificial structure shows that the intensity of the artificial line is maximum at the center and falls off radially, while the velocity field is quite constant and equals 28.69. Therefore, the emission satisfactory since the artificial component is a Gaussian profile in intensity and has a center velocity at channel number 28.

Figure 7.15. *Map of the three main classes of NGC 4254 with an artificial component added (threshold t equals 0,01 times the maximum intensity).*

Figure 7.16. *Emission distributions (top) and velocity fields (bottom) of the three classes (images in inverse video). The images have the same color scale.*

7.4. Conclusion and Perspectives

We consider in this chapter the problem of the kinematic segmentation of gas structures in galaxies. Data are radioastronomical cubes (multispectral images) where each pixel is a spectrum whose component lines are shifted due to the Universe expansion (Doppler effect) and local velocities of the gas. The shifts are relevant of the formation and the behavior of the galaxy.

We propose two Bayesian methods to handle the problem. They are intented to help the astronomers to handle complex data cubes and to be complementary to the standard method of analysis. Both consists in decomposing the spectrum as a sum of Gaussian components so that the data are reduced and used in a second step when segmentation is done.

More specifically, the first approach provides a segmentation map showing the different kinematic regions. The decomposition is processed by first decomposing each spectrum in six Gaussians, then the more representative Gaussians are selected to form a basis from which the pixels are decomposed again. This yields reduced data which feeds a Markovian segmentation algorithm. Nevertheless, this first approach has two main limitations : on one hand the singularity of the basis on the whole spatial location is not always appropriate, on the other hand the classification depends on the number of classes which is empirically fixed.

The second approach tries to overcome these limits by estimating each gas structure. Also, the pixels are decomposed into a sum of Gaussians, but their number, position, amplitude and widths are estimated using the RJMCMC algorithm. In other words, the spectra are not decomposed on a unique basis : this allows to fit well complex profiles. Unfortunately, it remains impossible to keep the previous segmentation algorithm, because the decomposition results may vary on each spectrum (number of rays, parameter of each ray). Thus, a second step consists in "tracking" the estimated components to classify them. By this way, we are now able to separate different gas structure behavior (in a kinematic point of view). However, the second approach is unable to separate two superimposed structures if they physically cross because it exists at least one pixel where the components of the two structures are two similar. The ability to distinguished each structure is clearly an advantage compared to the first approach.

In conclusion, the two methods presented in this chapter are complementary and make the analysis easier than using only the moment maps. Future works will be dedicated to develop a third approach combining the two advantages of the proposed methods.

[A.B 00]   A.BIJAOUI, D.NUZILLARD, « Blind Source Separation of Multispectral Astronomical Images », *MPA/ESO/MPE Joint Astronomy Conf., Mining the Sky*, July 31 - August 4 2000, Garching, Germany.

[BAC 02] BACKER S. D., Unsupervised Pattern Recognition Dimmensionality reduction and Classfication, PhD thesis, University of Antwerp, 2002.

[Bes 77] BESAG J., « Statistical Analysis of Non-Lattice Data », *The Statistician*, vol. 24, p. 179-195, 1977.

[BIL 97] BILMES J., A Gentle Tutorial on the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models, Rapport n°ICSI-TR-97-021, University of Berkeley, http ://ssli.ee.washington.edu/people/bilmes/mypapers/em.pdf, 1997.

[CAP 03] CAPPÉ O., ROBERT C., RYDÉN T., « Reversible jump, birth-and-death and more general continuous time Markov chain Monte Carlo samplers », *Journal of the Royal Statistical Society : Series B*, vol. 65, n°3, p. 679-700, 2003.

[CAR 89] CARDOSO J. F., « Source separation using higher order moments », *Proc. ICASSP'89*, Glasgow, UK, p. 2109-2112, 1989.

[CAR 93] CARDOSO J. F., SOULOUMIAC A., « Blind beamforming for non Gaussian signals », *IEE Proceedings-F*, vol. 140(6), p. 362-370, 1993.

[CAR 02] CARDOSO J.-F., SNOUSSI H., DELABROUILLE J., PATANCHON G., « Blind separation of noisy Gaussian stationary sources. Application to cosmic microwave background imaging », *XI European Signal Processing Conference EUSIPCO, Toulouse*, p. 561-564, September 3-6 2002.

[COM 94] COMON P., « Independent Component Analysis a new concept ? », *Signal Processing*, p. 36 :287-314, 1994.

[Dem 86] DEMPSTER A., LAIRD N., RUBIN D., « Maximum Likelihood from Incomplete Data via the EM Algorithm », *Journal of the Royal Statistical Society*, vol. 39, p. 1-38, December 1986.

[DEV 86] DEVROYE L., *Non-Uniform Random Variate Generation*, Springer-Verlag, 1986.

[DUD 01] DUDA R. O., HART P. E., STORK D. G., *Pattern Classification*, Wiley, 2001.

[DYK 02] DYK D. V., « Hierarchical models, data augmentation and MCMC », *Statistical Challenges in Modern Astronomy III*, (Eds. G. J. Babu and E. D. Feigelson), Springer, New York, 41-56, 2002.

[F.F 05] F. FLITTI CH COLLET B. V., BONNAREL F., « Multiband segmentation of a spectroscopic line data cube : application to the HI data cube of the spiral galaxy NGC 4254 », *EURASIP journal on Applied Signal Processing, special issue on Applications of Signal Processing in Astrophysics and Cosmology*, vol. 15, p. 2546-2558, 2005.

[GEL 96] GELMAN A., ROBERTS G., GILKS W., « Efficient Metropolis jumping rules », BERNARDO J., BERGER J., DAWID A., SMITH A., Eds., *Bayesian Statistics 5*, Oxford University Press, p. 599-608, 1996.

[Gem 84] GEMAN S., GEMAN D., « Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images », *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 6, n°6, p. 721-741, November 1984.

[GIL 95] GILKS W., RICHARDSON S., SPIEGELHALTER D., *Markov Chain Monte Carlo in Practice*, Chapman & Hall/CRC, 1995.

[Gra 95] GRAFFIGNE C., HEITZ F., PÉREZ P., PRÊTEUX F., SIGELLE M., ZERUBIA J., « Hierarchical Markov random field models applied to image analysis : a review », *SPIE Neural Morphological and Stochastic Methods in Image and Signal Processing*, vol. 2568, San Diego, p. 2-17, 10-11 July 1995.

[GRE 95] GREEN P., « Reversible jump Markov chain Monte Carlo computation and Bayesian model determination », *Biometrika*, vol. 82, p. 711-732, 1995.

[HUB 85] HUBER P. J., « Projection Pursuit with Discussion », *The Annals of Statistics*, vol. 13, n°2, p. 435-525, 1985.

[HUG 68] HUGHES G., « On the mean accuracy of statistical pattern recognizers », *IEEE Trans. on Information Theory*, vol. 14(1), p. 55-63, 1968.

[HYV 99] HYVÄRINEN A., « Fast and robust fixed-point algorithms for Independent Component Analysis », *IEEE Trans. on Neural Networks*, vol. 10(3), p. 626-634, May 1999.

[HYV 01] HYVÄRINEN A., KARHUNEN J., OJA E., *Independent Component Analysis*, Wiley, 2001.

[IDI 08] IDIER J., Ed., *Bayesian Approach to Inverse Problems*, ISTE Ltd and John Wiley & Sons Inc, 2008.

[JAI] JAIN A., MURTY M., FLYNN P.

[JIM 95] JIMENEZ L., LANDGREBE D., « High Dimensional Feature Reduction via Projection Pursuit », *School of Electrical and Computer Engineering, Purdue University, West Lafayette in 47907-1285*, April 1995.

[Laf 00a] LAFERTÉ J.-M., PÉREZ P., HEITZ F., « Discrete Markov Image Modeling and Inference on the Quad-tree », *IEEE Trans. on Image Processing*, vol. 9, n°3, p. 390-404, March 2000.

[LAF 00b] LAFERTÉ J.-M., PÉREZ P., HEITZ F., « Discrete Markov Image Modeling and Inference on the Quad-tree », *IEEE Trans. on Image Processing*, vol. 9, n°3, p. 390-404, March 2000.

[LAN 99] LANDGREBE D., « *Information Processing for Remote sensing* », Chapitre Information Extraction Principles and Methods for Multispectral and Hyperspectral Image Data, World Scientific Publishing Co., Inc., 1999.

[LEN 01] LENNON M., MERCIER G., MOUCHOT M., HUBERT-MOY L., « Independent Component Analysis as a tool for the dimensionality reduction and the representation of hyperspectral images », *IGARSS 2003 : International Conference on Geoscience and Remote Sensing, Sydney, Australia*, July 2001.

[Lue 93] LUETTGEN M., Image Processing with Multiscale Stochastic Models, PhD thesis, MIT Laboratory of Information and Decision Systems, May 1993.

[MAZ 05a] MAZET V., Développement de méthodes de traitement de signaux spectroscopiques : estimation de la ligne de base et du spectre de raies, PhD thesis, Université Henri Poincaré, Nancy 1, décembre 2005.

[MAZ 05b] MAZET V., BRIE D., IDIER J., « Simulation of Positive Normal Variables using several Proposal Distributions », *IEEE Workshop Statistical Signal Processing 2005*, 2005.

[MCL 97] MCLACHLAN G. F., KRISHNAN T., *The EM algorithm and Extensions*, Wiley, 1997.

[Mig 99] MIGNOTTE M., COLLET C., PÉREZ P., BOUTHEMY P., « Bayesian Inference and Optimization Strategies for some Detection and Classification Problems in Sonar Imagery », *Non linear image processing - SPIE Conference 3646*, vol. 3646-02, San Jose, USA, p. 14-27, January 1999.

[MØ 04] MØLLER J., WAAGEPETERSEN R., *Statistical Inference and Simulation for Spatial Point Processes*, Chapman & Hall/CRC, 2004.

[NUZ 93] NUZILLARD D., BIJAOUI A., « Blind source separation and analysis of multispectral astronomical images », *Astronomy and Astrophysics*, vol. 174(1), p. 129-138, November 1993.

[Per 89] PERLEY R. A., SCHWAB F. R., BRIDLE A. H., « Synthesis imaging in radio astronomy », *ASP Conference Series*, vol. 6, San Francisco, CA, Astronomical Society of the Pacificx, 1989.

[PHO 93] PHOOKUN B., VOGEL S., MUNDY L., « NGC 4254 : a spiral galaxy with an m=1 mode and infalling gas », *Astrophysical Journal*, vol. 418, p. 113-122, November 1993.

[PIE 92] PIECZYNSKI W., « Statistical image segmentation », *Machine Graphics and Vision*, vol. 1, n°2, p. 261-268, 1992.

[Pie 94] PIECZYNSKI W., « Champs de Markov cachés et estimation conditionnelle itérative », *Traitement du signal*, vol. 11, n°2, p. 141-153, 1994.

[Pro 04] PROVOST J. N., COLLET C., ROSTAING P., PÉREZ P., BOUTHEMY P., « Hierarchical Markovian Segmentation of Multispectral Images for the Reconstruction of Water Depth Maps », *Computer Vision and Image Understanding*, vol. 93, n°2, p. 155-174, February 2004.

[REL 01] RELLIER G., DESCOMBES X., FALZON F., ZÉRUBIA J., Hyperspectral texture analysis using a Gauss Markov model, Rapport, INRIA Sofia-Antipolis, http ://www-sop.inria.fr/ariana/personnel/Guillaume.Rellier/demo/demo-e.pdf, 2001.

[RIC 97] RICHARDSON S., GREEN P., « On Bayesian analysis of mixtures with an unknown number of components », *Journal of the Royal Statistical Society : Series B*, vol. 59, n°4, p. 731-792, 1997.

[ROB 02] ROBERT C., CASELLA G., *Monte Carlo statistical methods*, Springer-Verlag, 2002.

[SAL 04] SALERNO E., TONAZZINI A., KURUOGLU E., BEDINI L., HERRANZ D., BACCIGALUPI C., « Source separation techniques applied to astrophysical maps », M.GH. NEGOITA R. H., L.C. JAIN E., Eds., *Knowledge-Based Intelligent Information and Engineering Systems VIII*, vol. 3 de *Lecture Notes in Artificial Intelligence 3215*,, Springer, p. 426-432, 2004.

[TAN 96] TANNER M. A., *Tools for statistical inference*, Springer-Verlag, 1996.

[Van 00]   VAN LIESHOUT M., *Markov Point Processes and Their Applications*, Imperial College Press, 2000.

[WAL 97]   WALTER E., PRONZATO L., *Identification of Parametric Models from Experimental Data*, Communications and Control Engineering Series, Springer, London, 1997.

Denoising and segmentation

Chapitre 8

Wavelet transform for the denoising of multivariate images

8.1. Introduction

An increasing attention is being paid to multispectral images for a great number of applications (medicine, agriculture, archeology, forestry, coastal management, remote sensing . . .) because many features of the underlying scene have unique spectral characteristics that become apparent in imagery when viewing combinations of its different components. Hence, in satellite imaging, a better analysis of the nature of the materials covering the surface of the earth is achieved [LAN 00]. Typically, multispectral imaging systems employ radiometers as acquisition instruments which operate in different spectral channels. Each one delivers a digital image in a small range of the visible or non visible wavelengths. As a result, the spectral components form a multicomponent image corresponding to a single sensed area. Usually, satellites have three to a dozen of radiometers. Multispectral sensors offer a valuable advantage over color aerial photographs, thanks to their ability to record reflected light in the near infrared domain. Near infrared is the most sensitive spectral domain used to map vegetation canopy properties [GUY 90]. There are several families of on-board multispectral radiometers in the different satellite systems. The first example is SPOT 3 which has two High Resolution Visible imaging systems (HRV1 and HRV2). Each HRV is designated to operate in two sensing modes : a 10 m resolution "Panchromatic" (P) mode over the range [0.5, 0.73] μm and a 20 m resolution multispectral mode. For the multispectral mode, the first channel is associated with the range [0.5, 0.59] μm, the second channel with the range [0.61, 0.78] μm and the third one with

Chapitre rédigé par C. CHAUX, A. BENAZZA-BENYAHIA, J.-C. PESQUET, L. DUVAL.

the range [0.79, 0.89] μm. The SPOT family provides a service continuity with the upgraded satellite SPOT 4 (launched on March 1998) and SPOT 5 (launched on May 2002). In addition to the former 3 channels, SPOT 4 and SPOT 5 imaging systems gather images in a fourth channel corresponding to a short wave infrared spectral range ([1.58, 1.75] μm). The fourth channel was introduced in order to allow early observations of plant growth.

Another well-known family of multispectral satellite imaging systems is the set of Thematic Mapper instruments with the launch of Landsat 1 in 1972. Since April 1999, Landsat 7 carries the Enhanced Thematic Mapper Plus (ETM+) sensors which are similar to the Thematic Mapper sensors with additional features. An ETM+ Landsat scene is formed by 7 spectral components at a 30 m spatial resolution (except in the thermal band with a spatial resolution of 60 m) and a panchromatic image with 15 m pixel resolution. Recently, commercial satellites like Ikonos and Quickbird have provided very high resolution images. For instance, Ikonos 4 (resp. Quickbird) collects data with a level of detail of 4 m (resp. 2.4 or 2.8 m) in 4 spectral ranges (blue, green, red, and near infrared).

Despite the dramatical technological advances in terms of spatial and spectral resolutions of the radiometers, data still suffer from several degradations. For instance, the sensor limited aperture, aberrations inherent to optical systems and mechanical vibrations create a blur effect in remote sensing images [JAI 89]. In optical remote sensing imagery, there are also many noise sources. Firstly, the number of photons received by each sensor during the obturation time may fluctuate around its average implying a photon noise. A thermal noise may be caused by the electronics of the recording and the communication channels during the data downlinking. Intermittent saturations of any detector in a radiometer may give rise to an impulsive noise whereas a structured periodic noise is generally caused by interferences between electronic components. Detector striping (resp. banding) are consequences of calibration differences among individual scanning detectors (resp. from scan-to-scan). Besides, component-to-component misregistration may occur : corresponding pixels in different components are not systematically associated with the same position on the ground.

As a result, it is mandatory to apply deblurring, denoising and geometric corrections to the degraded observations in order to fully exploit the information they contain. In this respect, it is used to distinguish between on-board and on-ground processing. Indeed, on-board procedures should simultaneously fulfill real-time constraints and low mass memory requirements. The involved acquisition bit-rates are high (especially for very high resolution missions) and hence, they complicate the software implementation of enhancement processing. This is the reason why ASIC hardware circuits are employed. Such on-board circuits enable very basic processing since they present a lower performance than ground-based ones. For instance, Landsat ETM+ raw data are corrected for scan line direction and band alignment only. No radiometric or geometric correction is applied. Consequently, most of the efforts for enhancing the data are performed after their reception at the terrestrial stations. In this context, denoising is a delicate task since it aims at attenuating the noise level while maintaining the significant

image features. Generally, the focus is put on additive Gaussian noise. In this respect, many works have been carried out concerning single-component images. The pionnering ones were based on linear spatial filters and nonlinear ones [JAI 89, PIT 90]. In parallel to these efforts, a gain in performance can be achieved by attenuating the noise in a transform domain in which the image representation yields a tractable statistical modelling. The seminal work of Donoho has shown the potentialities of the Wavelet Transform (WT) for reducing a Gaussian additive noise thanks to its sparsity and decorrelation properties [DON 93]. As a consequence, several wavelet-based image denoising methods were investigated.

The objective of this work is to give an overview of the most relevant on-ground wavelet-based noise reduction methods devoted to multicomponent images. Two approaches can be considered. The first one consists of independently applying any monochannel noise reduction method to each component. Although its principle is simple, this approach suffers from a serious drawback as it does not account for the cross-component dependences. This has motivated the development of an alternative approach in which the noisy components are jointly processed. In broad outline, it is also possible to classify all the denoising methods (whatever they are componentwise or multivariate ones) into non Bayesian and Bayesian methods. For the latter category, a prior distribution model is adopted for the unknown image.

This chapter is organized as follows. Notations and the observation model are presented in Section 8.2. Section 8.3 is a concise overview on wavelet transforms and filter banks. Componentwise and multichannel denoising methods are presented in Section 8.4 : a wide panel of approaches is tackled (wavelet-based, Bayesian estimation, . . .). Finally, some comparisons are drawn in Section 8.5 before concluding the chapter by Section 8.6.

8.2. Observation model

This section is devoted to the characterization of multichannel satellite images.

8.2.1. *Observed images*

The "clean" unknown multicomponent image $(s^{(1)}(\mathbf{k}), \ldots, s^{(B)}(\mathbf{k}))_{\mathbf{k} \in \mathbb{K}}$, where $\mathbb{K} \subset \mathbb{Z}^2$ is a set of spatial indices, consists of B images corresponding to B spectral bands captured by B sensors. The notation $s^{(b)}(\mathbf{k})$ thus designates the intensity value of the pixel at position $\mathbf{k}$ in the b-th image component as represented in Fig. 8.1.

To better figure out the multichannel context, Fig. 8.2 displays 6 components of a Landsat 7 image. As can be noticed, image components share common structures, some details being present in specific spectral bands only. This phenomenon can be explained by the fact that some sensors are, for example, better able to capture vegetation whereas others are calibrated for soils. This is one of multispectral image specific properties that should be taken into account in the design of processing methods.

$$s^{(1)} \qquad\qquad\qquad s^{(b)}(\mathbf{k}) \qquad\qquad\qquad s^{(B)}$$

./DenoisingSegmentation/chaux/figs/mod_pix.pstex
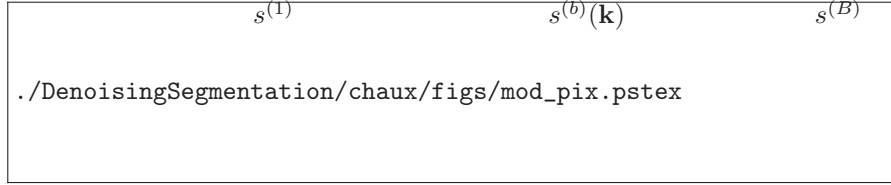
Figure 8.1. *In red, a pixel at spatial position* $\mathbf{k}$ *in the b-th image component.*

Unfortunately, the observed images denoted by $(r^{(1)}(\mathbf{k}))_{\mathbf{k}\in\mathbb{K}}$, ..., $(r^{(B)}(\mathbf{k}))_{\mathbf{k}\in\mathbb{K}}$ are subject to various degradations which are detailed in the next section.

8.2.2. *Degradation model*

The observed images are corrupted by noises coming from different sources [LAN 86] : atmospheric, sensor detector/preamplifier and quantization. In spite of the various statistical distributions of these noise sources, the global noise present in acquired data can be realistically modelled by an additive zero-mean spatially white Gaussian noise [COR 03, **?**] thus leading to the following model :

$$\forall b \in \{1, \ldots, B\}, \forall \mathbf{k} \in \mathbb{K}, \qquad r^{(b)}(\mathbf{k}) = s^{(b)}(\mathbf{k}) + n^{(b)}(\mathbf{k}). \qquad (8.1)$$

Following a multivariate approach, we define the unknown vector signal $\mathbf{s}$, the vector noise $\mathbf{n}$ and the vector observation $\mathbf{r}$, as

$$\forall \mathbf{k} \in \mathbb{K}, \quad \begin{cases} \mathbf{s}(\mathbf{k}) & \triangleq & [s^{(1)}(\mathbf{k}), \ldots, s^{(B)}(\mathbf{k})]^\top \\ \mathbf{n}(\mathbf{k}) & \triangleq & [n^{(1)}(\mathbf{k}), \ldots, n^{(B)}(\mathbf{k})]^\top \\ \mathbf{r}(\mathbf{k}) & \triangleq & [r^{(1)}(\mathbf{k}), \ldots, r^{(B)}(\mathbf{k})]^\top \end{cases}$$

and, consequently, Equation (8.1) can be reexpressed into a more concise form as

$$\forall \mathbf{k} \in \mathbb{K}, \qquad \mathbf{r}(\mathbf{k}) = \mathbf{s}(\mathbf{k}) + \mathbf{n}(\mathbf{k}) \qquad (8.2)$$

where $\mathbf{n}$ is an i.i.d. zero-mean Gaussian multivariate noise with covariance matrix $\mathbf{\Gamma}^{(\mathbf{n})} \in \mathbb{R}^{B \times B}$. This matrix can take different forms, three of which will catch our attention :

1) When the noise is uncorrelated from a component to another with the same variance σ^2 in each channel, the matrix takes the following form : $\mathbf{\Gamma}_1^{(\mathbf{n})} = \sigma^2 \ \mathbf{I}_B$, where $\mathbf{I}_B$ denotes the identity matrix of size $B \times B$.

2) When the noise is uncorrelated with various noise levels in the spectral bands, we have $\mathbf{\Gamma}_2^{(\mathbf{n})} = \mathrm{Diag}(\sigma_1^2, \ldots, \sigma_B^2)$.

./DenoisingSegmentation/chaux/figs/im_sat_trento1.eps   ./DenoisingSegmentation/chaux/figs/im_sat_trento2.eps

./DenoisingSegmentation/chaux/figs/im_sat_trento3.eps   ./DenoisingSegmentation/chaux/figs/im_sat_trento4.eps

./DenoisingSegmentation/chaux/figs/im_sat_trento5.eps   ./DenoisingSegmentation/chaux/figs/im_sat_trento7.eps

Figure 8.2. 6 *components of a Landsat* 7 *satellite image.*

3) Finally, a non-diagonal $\mathbf{\Gamma}_3^{(\mathbf{n})}$ matrix accounts for cross-channel correlations bet-

ween co-located noise samples. One can choose $\mathbf{\Gamma}_3^{(\mathbf{n})} = \sigma^2 \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \ddots & \rho \\ \vdots & \ddots & \ddots & \vdots \\ \rho & \cdots & \rho & 1 \end{pmatrix}$,

where $\rho \in (0, 1]$ is the correlation factor between two different noise components.

Our objective is thus to perform a multispectral image denoising under the considered assumptions. In this respect, we will see that the use of a multiscale linear transform such as the wavelet decomposition may be of great use.

8.3. An overview of discrete wavelets and multiscale transforms

8.3.1. *Historical remarks*

A discrete 1D signal r with location index k can be classically written as the following linear expansion :

$$\sum_{k \in \mathbb{Z}} r(k)\delta_k \,,$$

where δ represents the discrete (Kronecker) delta sequence located at 0. While this representation yields optimal sample location, corresponding to the canonical basis, it lacks in providing insights to the inherent signal structure which are beneficial to further processing. Due to the approximate linear nature of many physical processes, signal processing techniques have endeavored to employ a wealth of other suitable linear signal representations, such as the Fourier transform. Under some technical assumptions, the Fourier transform yields the following expansion of the signal :

$$\forall \nu \in [0, 1), \qquad R(\nu) = \sum_{k \in \mathbb{Z}} r(k) \exp(-\imath 2\pi k\nu).$$

Although widely used, the Fourier transform does not however allows us to enlighten the time behaviour of the signal. For signals possessing a certain regularity as well as singularities, localized representations called wavelets have generated a tremendous interest in the past 20 years. The story of wavelets is actually older, since it originated a century ago in a famous paper by A. Haar [HAA 10], who considered decompositions of functions into uniformly convergent series. It is generally considered that discrete wavelets, in their modern form, have emerged in the 1980's. Most of the related works have been nicely gathered in [HEI 06]. We follow here a derivation of wavelet representations based on filter banks, based on the pioneering work by Croisier *et al.* [CRO 76].

8.3.2. *2-band and M-band filter banks*

We consider square summable sequences $(h_m[k])_{k\in\mathbb{Z}}$ with $m \in \{0, \ldots, M-1\}$ representing the impulse responses of M filters. We often characterize each filter by its frequency response H_m, which is the Fourier transform of its impulse response. The basic building block for a 2-band filter bank based decomposition is illustrated in Fig. 8.3 : the digital signal $r(k)$ is decomposed into two frequency bands by a set of filters H_0 et H_1 followed by a decimation by a factor of 2, leading to a pair of coefficient sequences $(r_{1,0}(k))_{k\in\mathbb{Z}}$ and $(r_{1,1}(k))_{k\in\mathbb{Z}}$. The set of analysis filters H_0 and H_1 with its associated decimators is called an analysis filter bank. A reconstructed signal $\tilde{r}$ is obtained from $r_{1,0}$ and $r_{1,1}$ after a factor of 2 upsampling operator followed by filtering through synthesis filters $\widetilde{H}_0$ et $\widetilde{H}_1$ and summation.
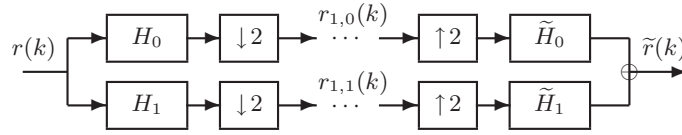
Figure 8.3. *Analysis/synthesis 2-band filter bank.*

The overall construction satisfies the Perfect Reconstruction property (PR) when the signals $\tilde{r}$ and r are equal (eventually up to an integer delay and a non-zero multiplicative factor, which can be incorporated in the filter coefficients). Such a property is verified for non trivial filter families (i.e. including filters with several delays) [SMI 84], whose properties are summarized for instance in [MEY 90, COH 92, DAU 92, MAL 98]. A traditional example is given by the Haar analysis filter bank with analysis filters H_0 and H_1 of length 2 :

$$\left(h_0[0], h_0[1]\right) = \frac{1}{\sqrt{2}}(1,\ 1)$$

$$\left(h_1[0], h_1[1]\right) = \frac{1}{\sqrt{2}}(-1,\ 1).$$

and synthesis filters $\widetilde{H}_0$ and $\widetilde{H}_1$ obtained by symmetry from H_0 and H_1 around the time origin.

Due to the relatively strong constraints imposed on the four filters H_0, H_1, $\widetilde{H}_0$ and $\widetilde{H}_1$ to satisfy the PR property, some authors have proposed a more general structure, named M-band filter banks, based on two sets of $M \geq 2$ analysis and synthesis filter banks, represented in Fig. 8.4.
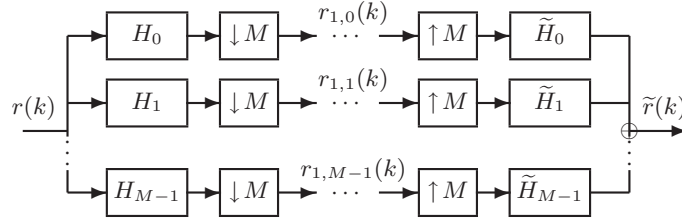
Figure 8.4. *Analysis/synthesis M-band filter bank.*

Similarly, the PR property may be obtained from an appropriate choice of analysis and synthesis filters [VAI 87], with improved flexibility in the design of the filters, since the 2-band case now represents a special instance of M-band filter banks. Moreover, the latter encompasses a large class of standard linear transforms, *e.g.* block transforms.

8.3.3. *Filter bank based multiresolution analysis*
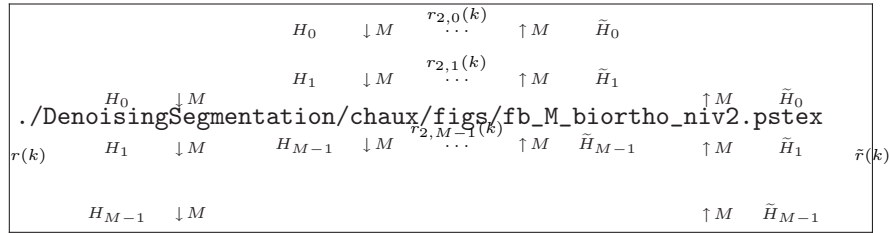
Figure 8.5. *2-level M-band wavelet analysis/synthesis wavelet decomposition.*

A multiresolution analysis of a signal consists of a decomposition where the signal is represented at different scales, allowing us to seize more easily its fine to coarse structures. It has been proved especially useful in signal recovery (denoising, deconvolution and reconstruction) as well as in data compression. A practical multiresolution analysis is obtained by cascading the basic analysis filter bank block. For the generic M-band filter bank case, assume that H_0 and H_{M-1} are a low-pass and a high-pass filter respectively, whereas $H_1, \ldots, H_{M-2}$ are band-pass filters. The low-pass filtering by H_0 followed by decimation yields a first subsampled approximation of the original signal, which may be further decomposed by the same filter bank, as represented in Fig. 8.5. The band-pass and high-pass branches yield subsampled versions of the signal details in different frequency bands, complementary to the low-pass approximation.

From the continuous-time viewpoint, such a multiresolution analysis can be studied in the space $L^2(\mathbb{R})$ of square integrable functions. Successive iterations of the basic M-band filter bank on the low-pass output is interpreted as approximations at resolution level j. The approximation spaces correspond to a decreasing sequence of nested subspaces $(V_j)_{j\in\mathbb{Z}}$ of $L^2(\mathbb{R})$, associated with one scaling function (or father wavelet) $\psi_0 \in L^2(\mathbb{R})$. The multiresolution analysis then corresponds to projections of the continuous-time signal onto subspaces $(W_{j,m})_{j\in\mathbb{Z},m\in\{1,\ldots,M-1\}}$, associated with $(M-1)$ mother wavelets $\psi_m \in L^2(\mathbb{R})$, $m \in \{1,\ldots,M-1\}$ [STE 93]. These functions are solutions of the following scaling equations :

$$\forall m \in \{0,\ldots,M-1\}, \qquad \frac{1}{\sqrt{M}}\psi_m\left(\frac{t}{M}\right) = \sum_{k=-\infty}^{\infty} h_m[k]\psi_0(t-k).$$

For $m \in \{0,\ldots,M-1\}$, $j \in \mathbb{Z}$ and $k \in \mathbb{Z}$, define the family of functions

$$\psi_{j,m,k}(t) = M^{-j/2}\psi_m(M^{-j}t - k).$$

Then, under orthogonality conditions, we can write :

$$r(t) = \sum_{m\in\{1,\ldots,M-1\}} \sum_{j\in\mathbb{Z}}\sum_{k\in\mathbb{Z}} r_{j,m}(k)\psi_{j,m,k}(t)$$

where

$$r_{j,m}(k) = \langle r, \psi_{j,m,k}\rangle$$

and $\langle \cdot, \cdot \rangle$ denotes the standard inner product of $L^2(\mathbb{R})$. The latter expansion is called an M-band wavelet decomposition of r onto the orthonormal wavelet basis $\{\psi_{j,m,k}, (j,k) \in \mathbb{Z}^2, m \in \{1,\ldots,M-1\}\}$. For more insight on the continuous-time wavelet decomposition, we refer to [MAL 98, FLA 98].

8.3.4. *2D extension*

For simplicity, we only consider separable two-dimensional wavelet transforms which constitute a direct extension of the 1D case. The image is processed in two steps : the filter bank is applied successively to the image rows and columns. Consequently, the obtained 2D wavelets are equal to the tensor product of the 1D wavelets and define a $L^2(\mathbb{R}^2)$ basis. Applying such a transform to multicomponent images consists of applying the 2D transform on each channel b giving rise to the following coefficients : $\forall b \in \{1,\ldots,B\}, \forall \mathbf{m} = (m_1,m_2) \in \{0,\ldots,M-1\}^2, \forall j \in \mathbb{Z}$ and $\forall \mathbf{k} = (k_1,k_2) \in \mathbb{Z}^2$,

$$r_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = \langle\langle r^{(b)}, \psi_{j,m_1,k_1}\psi_{j,m_2,k_2}\rangle\rangle$$

where $\langle\langle \cdot, \cdot \rangle\rangle$ denotes the standard inner product of $L^2(\mathbb{R}^2)$. The separable property of the transform allows us to obtain a directional analysis of images, separating the horizontal, vertical and "diagonal" directions.

8.3.5. *Other related representations*

Nevertheless, wavelets suffer from some drawbacks : the first one is a lack of shift invariance whose potential shift-variant edge artifacts are not desirable in applications like denoising. Another drawback of decompositions onto wavelet bases is that they provide a relatively rough directional analysis. Tools that improve the representation of geometric information like textures and edges, and preserve them during processing are thus required. Consequently, during the last decade, many authors proposed more sophisticated representation tools called frames having exact or approximate shift-invariance properties and/or better taking into account geometrical image features. One such frames is simply obtained by dropping the decimation step in the previous filter bank structures, so leading to an undecimated wavelet transform [COI 95, PES 96] which has a redundancy equal to the number J of considered resolution levels. Note that such overcomplete wavelet representations can be built by considering the union of M^J shifted wavelet bases. In this case, cycle spinning denoising techniques may be used which consist of estimating the signal in each basis and averaging the resulting M^J estimates. In order to reduce the computational cost of these decompositions or to better capture geometrical features, other frame representations have been designed. These decompositions provide a local, multiscale, directional analysis of images and they often have a limited redundancy [COI 92, CAN 06, DO 05, MAL 08, CHA 06].

8.3.6. *Related model*

An M-band orthonormal discrete wavelet decomposition over J resolution levels is performed to the observation field $r^{(b)}$ for each channel b. This decomposition produces $M^2 - 1$ wavelet subband sequences $r_{j,\mathbf{m}}^{(b)}$, $\mathbf{m} \in \{0, ..., M-1\}^2 \setminus \{(0,0)\}$, each of size $L_j \times L_j$, at every resolution level j and an additional approximation sequence $r_{J,\mathbf{0}}^{(b)}$ of size $L_J \times L_J$, at the coarsest resolution level J (to simplify our presentation, we consider square images).

On the one hand, the linearity of the Discrete Wavelet Transform (DWT) yields (see. Fig. 8.6) :

$$\forall \mathbf{k} \in \mathbb{K}_j, \quad \mathbf{r}_{j,\mathbf{m}}(\mathbf{k}) = \mathbf{s}_{j,\mathbf{m}}(\mathbf{k}) + \mathbf{n}_{j,\mathbf{m}}(\mathbf{k}) \tag{8.3}$$

where $\mathbb{K}_j = \{0, \dots, L_j - 1\}^2$ and

$$\mathbf{s}_{j,\mathbf{m}}(\mathbf{k}) \triangleq [s_{j,\mathbf{m}}^{(1)}(\mathbf{k}), \dots, s_{j,\mathbf{m}}^{(B)}(\mathbf{k})]^\top,$$

$$\mathbf{n}_{j,\mathbf{m}}(\mathbf{k}) \triangleq [n_{j,\mathbf{m}}^{(1)}(\mathbf{k}), \dots, n_{j,\mathbf{m}}^{(B)}(\mathbf{k})]^\top,$$

$$\mathbf{r}_{j,\mathbf{m}}(\mathbf{k}) \triangleq [r_{j,\mathbf{m}}^{(1)}(\mathbf{k}), \dots, r_{j,\mathbf{m}}^{(B)}(\mathbf{k})]^\top.$$

On the other hand, the orthonormality of the DWT preserves the spatial whiteness of $\mathbf{n}_{j,\mathbf{m}}$. More specifically, it is easily shown that the latter field is an i.i.d. $\mathcal{N}(\mathbf{0}, \mathbf{\Gamma^{(n)}})$ random vector process.

A final required assumption is that the random vectors $(\mathbf{s}_{j,\mathbf{m}}(\mathbf{k}))_{k \in \mathbb{K}}$ are identically distributed for any given value of $(j, \mathbf{m})$.

$$s^{(b)}_{j,\mathbf{m}} \qquad\qquad r^{(b)}_{j,\mathbf{m}}$$

$\mathcal{WT}$ $\qquad\qquad$ $\mathcal{WT}$

$s^{(b)}$ $\qquad\qquad$ $r^{(b)}$

$n^{(b)}$ /DenoisingSegmentation/chaux/figs/model_wavelets.pstex

iid $\mathcal{N}(\mathbf{0}, \sigma_b^2)$

$\mathcal{WT}$

$n^{(b)}_{j,\mathbf{m}}$
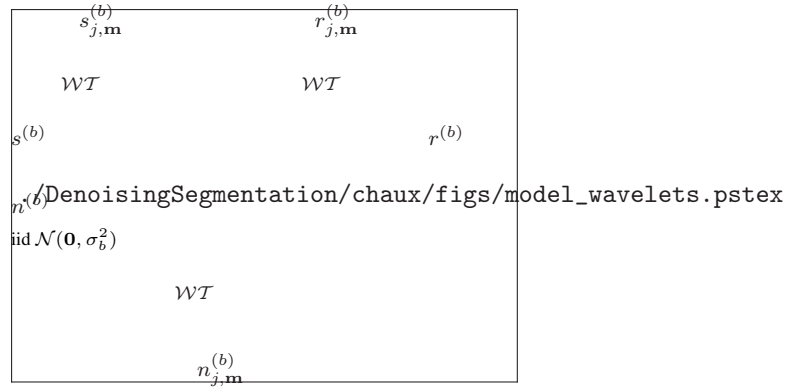
Figure 8.6. *Considered model in the wavelet transform domain.*

8.4. A survey of the most relevant univariate and multivariate denoising methods

Our objective is to build an estimator $\overset{\wedge}{\mathbf{s}}$ of the multichannel image $\mathbf{s}$ from the degraded observation $\mathbf{r}$. The estimating function is denoted by $\mathbf{f}$ and, we have thus $\overset{\wedge}{\mathbf{s}} = \mathbf{f}(\mathbf{r})$. In the present case, as shown in Section 8.2.2, we have to deal with Gaussian noise removal. This is a multivariate estimation problem since the original multichannel image is composed of $B \in \mathbb{N}^*$ components $s^{(b)}$ of size $L \times L$, with $b \in \{1, \ldots, B\}$. Different denoising techniques are presented below. We briefly describe Fourier domain methods and then we focus our attention on wavelet-based methods. But first and foremost, we present the general context we adopt for all the methods operating in the wavelet domain.

8.4.1. *Context in the wavelet domain*

In the wavelet domain, by using the notations defined in Section 8.3, the degradation model (8.2) becomes (8.3). Actually, we consider the more flexible situation where an observation sequence $(\overline{\mathbf{r}}^{(b)}_{j,\mathbf{m}}(\mathbf{k}))_{\mathbf{k} \in \mathbb{K}_j}$ of d-dimensional real-valued vectors with $b \in \{1, \ldots, B\}$ and $d > 1$, is given by

$$\forall \mathbf{k} \in \mathbb{K}_j, \qquad \overline{\mathbf{r}}^{(b)}_{j,\mathbf{m}}(\mathbf{k}) = \overline{\mathbf{s}}^{(b)}_{j,\mathbf{m}}(\mathbf{k}) + \overline{\mathbf{n}}^{(b)}_{j,\mathbf{m}}(\mathbf{k})$$

and $(\bar{\mathbf{n}}_{j,\mathbf{m}}^{(b)}(\mathbf{k}))_{\mathbf{k}\in\mathbb{K}_j}$ is a zero-mean spatially white Gaussian noise with covariance matrix $\boldsymbol{\Gamma}^{(\bar{\mathbf{n}}^{(b)})}$, which is assumed to be invertible. The three above vectors will be taken of the form :

$$\bar{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = \begin{bmatrix} r_{j,\mathbf{m}}^{(b)}(\mathbf{k}) \\ \tilde{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k}) \end{bmatrix}, \quad \bar{\mathbf{s}}_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = \begin{bmatrix} s_{j,\mathbf{m}}^{(b)}(\mathbf{k}) \\ \tilde{\mathbf{s}}_{j,\mathbf{m}}^{(b)}(\mathbf{k}) \end{bmatrix}, \quad \bar{\mathbf{n}}_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = \begin{bmatrix} n_{j,\mathbf{m}}^{(b)}(\mathbf{k}) \\ \tilde{\mathbf{n}}_{j,\mathbf{m}}^{(b)}(\mathbf{k}) \end{bmatrix}$$

where $\tilde{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k})$, $\tilde{\mathbf{s}}_{j,\mathbf{m}}^{(b)}(\mathbf{k})$ and $\tilde{\mathbf{n}}_{j,\mathbf{m}}^{(b)}(\mathbf{k})$ are random vectors of dimension $d-1$. These vectors may for example correspond to neighboring variables of the associated scalar variables $r_{j,\mathbf{m}}^{(b)}(\mathbf{k})$, $s_{j,\mathbf{m}}^{(b)}(\mathbf{k})$ and $n_{j,\mathbf{m}}^{(b)}(\mathbf{k})$. In this context, our objective is to estimate $s_{j,\mathbf{m}}^{(b)}(\mathbf{k})$ using the observation sequence $(\bar{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k}))_{\mathbf{k}\in\mathbb{K}_j}$. The vector $\bar{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k})$ is called the *Reference Observation Vector* (ROV) from which the following estimate is built :

$$\forall \mathbf{k} \in \mathbb{K}_j, \quad \hat{s}_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = f_{j,\mathbf{m}}^{(b)}\big(\bar{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k})\big).$$

Explicit choices of the ROV sequence $(\bar{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k}))_{\mathbf{k}\in\mathbb{K}_j}$ are detailed in the next paragraphs.

8.4.2. *Popular componentwise methods*

A first strategy for denoising a multichannel image is to perform a componentwise processing without taking into account any statistical dependence existing between the channels.

8.4.2.1. *Frequency domain*

A very popular method operating in the frequency domain is the Wiener filter [WIE 49]. This filter is designed so as to minimize the mean square error $\mathsf{E}[|\hat{s}^{(b)}(\mathbf{k}) - s^{(b)}(\mathbf{k})|^2]$, for every $b \in \{1,\dots,B\}$, under the assumption that $s^{(b)}$ is a wide-sense stationary random field. The frequency response of this filter reads

$$\forall \boldsymbol{\nu} \in [0,1)^2, \quad H^{(b)}(\boldsymbol{\nu}) = \frac{S_{s^{(b)}}(\boldsymbol{\nu})}{S_{s^{(b)}}(\boldsymbol{\nu}) + \sigma_b^2},$$

where $S_{s^{(b)}}$ denotes the power spectrum density of $s^{(b)}$ and σ_b is the standard deviation of the noise in channel b. One of the main drawbacks of this method is that it requires the a priori knowledge of the power spectrum density or an empirical estimation of it. Note that a multicomponent version of the Wiener filter has been derived in [ANG 91] by using the multi-input multi-output 2D filter minimizing $\mathsf{E}[\|\overset{\wedge}{\mathbf{s}}(\mathbf{k}) - \mathbf{s}(\mathbf{k})\|^2]$, [1] so taking into account the spectrum density matrix of $\mathbf{s}$.

---

1. $\|.\|$ denotes the classical Euclidean norm of $\mathbb{R}^B$.

However, it may appear more useful to solve the problem in the wavelet transform domain [ATK 03], where we can take advantage of a space-frequency representation of the images. Indeed, noise coefficients are usually distributed over small wavelet coefficients, whereas signal coefficients are concentrated on high magnitude ones. In addition, in [DON 94], Donoho and Jonhstone showed that a both simple and efficient approach for noise removal is available, through wavelet thresholding.

8.4.2.2. *Visushrink*

Visushrink [DON 93] is a componentwise method, which means that the ROV reduces to a scalar :

$$\bar{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = r_{j,\mathbf{m}}^{(b)}(\mathbf{k}).$$

Two kinds of thresholdings are usually employed :

– hard thresholding :

$$\forall \mathbf{k} \in \mathbb{K}_j, \quad f_{j,\mathbf{m}}^{(b)}\big(r_{j,\mathbf{m}}^{(b)}(\mathbf{k})\big) = \begin{cases} r_{j,\mathbf{m}}^{(b)}(\mathbf{k}) & \text{if } |r_{j,\mathbf{m}}^{(b)}(\mathbf{k})| > \chi^{(b)} \\ 0 & \text{otherwise.} \end{cases}$$

– soft thresholding :

$$\forall \mathbf{k} \in \mathbb{K}_j, \quad f_{j,\mathbf{m}}^{(b)}\big(r_{j,\mathbf{m}}^{(b)}(\mathbf{k})\big) = \text{sign}(r_{j,\mathbf{m}}^{(b)}(\mathbf{k})) \max\{|r_{j,\mathbf{m}}^{(b)}(\mathbf{k})| - \chi^{(b)}, 0\}$$

$$= \Big(\frac{|r_{j,\mathbf{m}}^{(b)}(\mathbf{k})| - \chi^{(b)}}{|r_{j,\mathbf{m}}^{(b)}(\mathbf{k})|}\Big)_+ r_{j,\mathbf{m}}^{(b)}(\mathbf{k}) \tag{8.4}$$

where $\text{sign}(\cdot)$ is the signum function.

These two shrinkage rules are illustrated in Fig. 8.7. The problem here is to find the best threshold value $\chi^{(b)} > 0$. In [DON 93], the authors have derived the so-called universal threshold $\chi^{(b)} = 2\sigma_b\sqrt{\log(L)}$ which relies on the fact that the maximum values of any set of L^2 independent random variables identically distributed as $\mathcal{N}(0, \sigma_b^2)$ are smaller than the proposed threshold $\chi^{(b)}$ with a high probability [MAL 98, p. 556].

8.4.3. *Extension to block-based method*

In order to take into account correlations between wavelet coefficients, some authors have proposed to apply a block shrinkage. More precisely, in [CAI 01], it is proposed to exploit the spatial dependences, which corresponds to the following choice of the ROV :

$$\bar{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = [r_{j,\mathbf{m}}^{(b)}(\mathbf{k}), r_{j,\mathbf{m}}^{(b)}(\mathbf{k} - \mathbf{k}_1), \dots, r_{j,\mathbf{m}}^{(b)}(\mathbf{k} - \mathbf{k}_{d-1})]^\top,$$

where $\mathbf{k}_1, \dots, \mathbf{k}_{d-1}$ allow us to define the neighborhood of interest for the pixel $\mathbf{k}$.
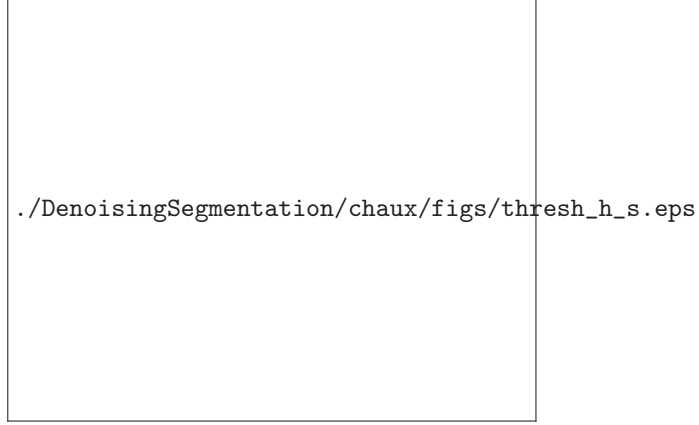
Figure 8.7. *Hard (continuous line) and soft (dashed line) thresholdings.*

The associated shrinkage rule named "NeighBlock" is given by

$$\overset{\curlywedge}{s}{}^{(b)}_{j,\mathbf{m}}(\mathbf{k}) = \left(\frac{\|\bar{\mathbf{r}}^{(b)}_{j,\mathbf{m}}(\mathbf{k})\|^2 - \bar{\chi}d\sigma_b^2}{\|\bar{\mathbf{r}}^{(b)}_{j,\mathbf{m}}(\mathbf{k})\|^2} \right)_+ r^{(b)}_{j,\mathbf{m}}(\mathbf{k}),$$

where $\bar{\chi} > 0$ and d is the number of components in the ROV.

In [ŞEN 02], interscale dependencies have been exploited by defining the following ROV :

$$\bar{\mathbf{r}}^{(b)}_{j,\mathbf{m}}(\mathbf{k}) = [r^{(b)}_{j,\mathbf{m}}(\mathbf{k}), r^{(b)}_{j+1,\mathbf{m}}(\lceil \frac{\mathbf{k}}{M} \rceil), \ldots, r^{(b)}_{J,\mathbf{m}}(\lceil \frac{\mathbf{k}}{M^{J-j}} \rceil)]^\top.$$

The associated estimator called "bivariate shrinkage" is defined by :

$$\overset{\curlywedge}{s}{}^{(b)}_{j,\mathbf{m}}(\mathbf{k}) = \left(\frac{\|\bar{\mathbf{r}}^{(b)}_{j,\mathbf{m}}(\mathbf{k})\| - \frac{\sqrt{3}\sigma_b^2}{\sigma_{s^{(b)}}}}{\|\bar{\mathbf{r}}^{(b)}_{j,\mathbf{m}}(\mathbf{k})\|} \right)_+ r^{(b)}_{j,\mathbf{m}}(\mathbf{k}).$$

where $\sigma_{s^{(b)}} > 0$. It can be derived by a Maximum A Posteriori (MAP) rule by considering as a prior model for the wavelet coefficients the non-Gaussian bivariate probability density function

$$p(s^{(b)}_{j,\mathbf{m}}(\mathbf{k}), s^{(b)}_{j+1,\mathbf{m}}(\lceil \frac{\mathbf{k}}{2} \rceil)) \propto \exp\left(- \frac{\sqrt{3}}{\sigma_{s^{(b)}}} \sqrt{\left|s^{(b)}_{j,\mathbf{m}}(\mathbf{k})\right|^2 + \left|s^{(b)}_{j+1,\mathbf{m}}(\lceil \frac{\mathbf{k}}{2} \rceil)\right|^2} \right).$$

Note that interscale dependencies are also taken into account in [SCH 04] where a multivalued image wavelet thresholding is performed. Other works developed Bayesian estimation procedures imposing a prior on the noise-free data.

8.4.4. *Bayesian multichannel approaches*

As previously mentioned, Bayesian approaches require a prior data statistical modeling.

8.4.4.1. *Bernoulli-Gaussian priors*

A Bernoulli-Gaussian (BG) prior is an appropriate model to reflect the sparsity of the wavelet representation of natural images [ABR 98, LEP 99]. With this statistical model, some authors derived Bayesian estimates [BEN 03, ELM 05b].

Let us first present the method proposed in [BEN 03, ELM 05b]. In each subband of index $(j, \mathbf{m})$, the probability distribution $p_{j,\mathbf{m}}$ of the coefficients $(\mathbf{s}_{j,\mathbf{m}}(\mathbf{k}))_{\mathbf{k}\in\mathbb{K}}$ can be written as follows :

$$\forall \mathbf{u} \in \mathbb{R}^B, \quad p_{j,\mathbf{m}}(\mathbf{u}) = (1 - \epsilon_{j,\mathbf{m}})\delta(\mathbf{u}) + \epsilon_{j,\mathbf{m}}\, g_{\mathbf{0},\boldsymbol{\Gamma}^{(\mathbf{s})}_{j,\mathbf{m}}}(\mathbf{u}),$$

where $g_{\mathbf{0},\boldsymbol{\Gamma}^{(\mathbf{s})}_{j,\mathbf{m}}}$ denotes the $\mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}^{(\mathbf{s})}_{j,\mathbf{m}})$ multivariate normal probability density function. The mixture parameter $\epsilon_{j,\mathbf{m}}$ corresponds to the probability that a coefficient vector $\mathbf{s}_{j,\mathbf{m}}(\mathbf{k})$ contains useful information. In order to avoid degenerated MAP estimates, it is used to couple the multivariate prior model with hidden random variables $q_{j,\mathbf{m}}(\mathbf{k})$. The sequence $(q_{j,\mathbf{m}}(\mathbf{k}))_{\mathbf{k}\in\mathbb{K}_j}$ is an i.i.d. binary sequence of random variables defining the following conditional densities : for every $\mathbf{k} \in \mathbb{K}_j$,

$$p(\mathbf{s}_{j,\mathbf{m}}(\mathbf{k}) \mid q_{j,\mathbf{m}}(\mathbf{k}) = 0) = \delta(\mathbf{s}_{j,\mathbf{m}}(\mathbf{k})),$$
$$p(\mathbf{s}_{j,\mathbf{m}}(\mathbf{k}) \mid q_{j,\mathbf{m}}(\mathbf{k}) = 1) = g_{\mathbf{0},\boldsymbol{\Gamma}^{(\mathbf{s})}_{j,\mathbf{m}}}(\mathbf{s}_{j,\mathbf{m}}(\mathbf{k}))$$

with $\mathsf{P}(q_{j,\mathbf{m}}(\mathbf{k}) = 1) = \epsilon_{j,\mathbf{m}} \in [0, 1]$. In practice, the hyperparameters $\boldsymbol{\Gamma}^{(\mathbf{s})}_{j,\mathbf{m}}$ and $\epsilon_{j,\mathbf{m}}$ related to the BG priors can be estimated by a moment method or an Expectation Maximization (EM) technique. As the noise is Gaussian, the following conditional probability densities are easily derived : for every $\mathbf{k} \in \mathbb{K}_j$,

$$\begin{cases} p(\mathbf{r}_{j,\mathbf{m}}(\mathbf{k}) \mid q_{j,\mathbf{m}}(\mathbf{k}) = 0) & = & g_{\mathbf{0},\boldsymbol{\Gamma}^{(\mathbf{n})}}(\mathbf{r}_{j,\mathbf{m}}(\mathbf{k})) \\ p(\mathbf{r}_{j,\mathbf{m}}(\mathbf{k}) \mid q_{j,\mathbf{m}}(\mathbf{k}) = 1) & = & g_{\mathbf{0},\boldsymbol{\Gamma}^{(\mathbf{n})}+\boldsymbol{\Gamma}^{(\mathbf{s})}_{j,\mathbf{m}}}(\mathbf{r}_{j,\mathbf{m}}(\mathbf{k})). \end{cases}$$

Thus, a two-step estimation procedure can be used for noise removal :

1) For every $\mathbf{k} \in \mathbb{K}_j$, the estimate $\hat{q}_{j,\mathbf{m}}(\mathbf{k})$ of $q_{j,\mathbf{m}}(\mathbf{k})$ is set to 1 if :

$$\mathsf{P}(q_{j,\mathbf{m}}(\mathbf{k}) = 0 \mid \mathbf{r}_{j,\mathbf{m}}(\mathbf{k})) < \mathsf{P}(q_{j,\mathbf{m}}(\mathbf{k}) = 1 \mid \mathbf{r}_{j,\mathbf{m}}(\mathbf{k})),$$

otherwise $\hat{q}_{j,\mathbf{m}}(\mathbf{k})$ is set to 0. This implies that :

$$\hat{q}_{j,\mathbf{m}}(\mathbf{k}) = \begin{cases} 1 & \text{if } \mathbf{r}_{j,\mathbf{m}}(\mathbf{k})^{\top}\mathbf{M}_{j,\mathbf{m}}\mathbf{r}_{j,\mathbf{m}}(\mathbf{k}) > \chi_{j,\mathbf{m}}, \\ 0 & \text{otherwise} \end{cases}$$

where $\mathbf{M}_{j,\mathbf{m}}$ is the semi-definite positive matrix :

$$\mathbf{M}_{j,\mathbf{m}} = (\mathbf{\Gamma}^{(\mathbf{n})})^{-1} - (\mathbf{\Gamma}^{(\mathbf{s})}_{j,\mathbf{m}} + \mathbf{\Gamma}^{(\mathbf{n})})^{-1},$$

and the threshold $\chi_{j,\mathbf{m}}$ is defined by

$$\chi_{j,\mathbf{m}} = 2\ln\left(\frac{1-\epsilon_{j,\mathbf{m}}}{\epsilon_{j,\mathbf{m}}}\right) + \ln\left(\frac{\mid \mathbf{\Gamma}^{(\mathbf{s})}_{j,\mathbf{m}} + \mathbf{\Gamma}^{(\mathbf{n})}\mid}{\mid \mathbf{\Gamma}^{(\mathbf{n})}\mid}\right)$$

where $|\mathbf{A}|$ denotes the determinant of matrix $\mathbf{A}$.

2) On the one hand, if $\hat{q}_{j,\mathbf{m}}(\mathbf{k}) = 0$, it is expected that the related observation is dominated by the noise, according to the definition of the hidden variables. Hence, it is natural to set $\overset{\wedge}{\mathbf{s}}_{j,\mathbf{m}}(\mathbf{k}) = \mathbf{0}$. On the other hand, if $\hat{q}_{j,\mathbf{m}}(\mathbf{k}) = 1$, the Bayesian estimate of $\mathbf{s}_{j,\mathbf{m}}$ minimizing a quadratic cost is computed. It corresponds to the a posteriori conditional mean. The posterior distribution is Gaussian as the bivariate distribution of $(\mathbf{r}_{j,\mathbf{m}}(\mathbf{k}), \mathbf{s}_{j,\mathbf{m}}(\mathbf{k}))$ is Gaussian when $q_{j,\mathbf{m}}(\mathbf{k}) = 1$. It is easy to check that :

$$\forall \mathbf{k} \in \mathbb{K}_j, \quad \mathsf{E}[\mathbf{s}_{j,\mathbf{m}}(\mathbf{k}) \mid \mathbf{r}_{j,\mathbf{m}}(\mathbf{k}), q_{j,\mathbf{m}}(\mathbf{k}) = 1] = \mathbf{Q}_{j,\mathbf{m}}\mathbf{r}_{j,\mathbf{m}}(\mathbf{k})$$

where

$$\mathbf{Q}_{j,\mathbf{m}} \overset{\triangle}{=} \mathbf{\Gamma}^{(\mathbf{s})}_{j,\mathbf{m}}(\mathbf{\Gamma}^{(\mathbf{s})}_{j,\mathbf{m}} + \mathbf{\Gamma}^{(\mathbf{n})})^{-1}.$$

It appears that the estimator amounts to a shrinkage rule that performs a tradeoff between a linear estimation in the sense of a minimum mean square error and a hard thresholding.

An alternate approach to this two-step estimation procedure is the use of the a posteriori conditional mean which, for every $\mathbf{k} \in \mathbb{K}_j$, can be expressed as

$$\mathsf{E}[\mathbf{s}_{j,\mathbf{m}}(\mathbf{k}) \mid \mathbf{r}_{j,\mathbf{m}}(\mathbf{k})] = \mathsf{E}[\mathbf{s}_{j,\mathbf{m}}(\mathbf{k}) \mid \mathbf{r}_{j,\mathbf{m}}(\mathbf{k}), q_{j,\mathbf{m}}(\mathbf{k}) = 1]\mathsf{P}(q_{j,\mathbf{m}}(\mathbf{k}) = 1 \mid \mathbf{r}_{j,\mathbf{m}}(\mathbf{k}))$$

since $p(\mathbf{s}_{j,\mathbf{m}}(\mathbf{k}) \mid \mathbf{r}_{j,\mathbf{m}}(\mathbf{k}), q_{j,\mathbf{m}}(\mathbf{k}) = 0) = \delta(\mathbf{s}_{j,\mathbf{m}}(\mathbf{k}))$. Besides, we can write :

$$
\begin{aligned}
\forall \mathbf{k} \in \mathbb{K}_j, \; \mathsf{P}(q_{j,\mathbf{m}}(\mathbf{k}) = 1 \mid \mathbf{r}_{j,\mathbf{m}}(\mathbf{k})) &= \frac{p(\mathbf{r}_{j,\mathbf{m}}(\mathbf{k}) \mid q_{j,\mathbf{m}}(\mathbf{k}) = 1)\mathsf{P}(q_{j,\mathbf{m}}(\mathbf{k}) = 1)}{p(\mathbf{r}_{j,\mathbf{m}}(\mathbf{k}))} \\
&= \frac{\epsilon_{j,\mathbf{m}}\theta_1(\mathbf{r}_{j,\mathbf{m}}(\mathbf{k}))}{\epsilon_{j,\mathbf{m}}\theta_1(\mathbf{r}_{j,\mathbf{m}}(\mathbf{k})) + (1-\epsilon_{j,\mathbf{m}})\theta_0(\mathbf{r}_{j,\mathbf{m}}(\mathbf{k}))} \\
&\overset{\triangle}{=} \gamma_{\epsilon_{j,\mathbf{m}}}(\mathbf{r}_{j,\mathbf{m}}(\mathbf{k})).
\end{aligned}
$$

with the following definitions :

$$\theta_0 \overset{\triangle}{=} g_{\mathbf{0},\mathbf{\Gamma}^{(\mathbf{n})}}, \quad \theta_1 \overset{\triangle}{=} g_{\mathbf{0},\mathbf{\Gamma}^{(\mathbf{s})}_{j,\mathbf{m}}+\mathbf{\Gamma}^{(\mathbf{n})}}.$$

The optimal mean-square Bayesian estimate can be easily deduced :

$$\forall \mathbf{k} \in \mathbb{K}, \qquad \overset{\wedge}{\mathbf{s}}_{j,\mathbf{m}}(\mathbf{k}) = \gamma_{\epsilon_{j,\mathbf{m}}}(\mathbf{r}_{j,\mathbf{m}}(\mathbf{k}))\mathbf{Q}_{j,\mathbf{m}}\mathbf{r}_{j,\mathbf{m}}(\mathbf{k}). \tag{8.5}$$

Note that in [ELM 05b], interscale dependencies are taken into account in addition to cross-channel ones.

8.4.4.2. *Laplacian mixture model*

A Bayesian componentwise approach was proposed by Pižurica and Philips [PIŽ 06]. They have described a simple way for applying it to a multicomponent image, when the noise is componentwise decorrelated ($\mathbf{\Gamma}^{(\mathbf{n})} = \mathbf{\Gamma}_1^{(\mathbf{n})} = \sigma^2 \, \mathbf{I}_B$). For each component at each resolution level and in each oriented subband, the principle is to consider a mixture of two truncated Generalized Gaussian (GG) (also called generalized Laplacian) distributions where a Bernoulli random variable controls the switching between the central part of the distribution and its tails. More precisely, for each component b, the authors have considered as a prior the GG distribution :

$$\forall u \in \mathbb{R}, \quad p_{j,\mathbf{m}}^{(b)}(u) = \frac{\left(\lambda_{j,\mathbf{m}}^{(b)}\right)^{1/\beta_{j,\mathbf{m}}^{(b)}}}{2\Gamma(1/\beta_{j,\mathbf{m}}^{(b)})} \exp(-\lambda_{j,\mathbf{m}}^{(b)}|u|^{\beta_{j,\mathbf{m}}^{(b)}})$$

where $\Gamma(z) = \int_0^{+\infty} t^{z-1}e^{-t}dt$ is the Gamma function, $\lambda_{j,\mathbf{m}}^{(b)} > 0$ is the scale parameter and $\beta_{j,\mathbf{m}}^{(b)} > 0$ is the shape parameter. It is worth pointing out that the prior hyperparameters can be easily estimated from the fourth moments of the noisy coefficients $r_{j,\mathbf{m}}^{(b)}$ [SIM 96]. Then, they have defined a signal of interest as a noise-free coefficient which exceeds a given threshold $T_{j,\mathbf{m}}^{(b)}$. To estimate the signal of interest from the noisy observations, they have introduced a sequence of Bernoulli variables $q_{j,\mathbf{m}}^{(b)}(\mathbf{k})$ associated with the two hypotheses $\mathcal{H}_0$ "the noise-free signal is not of interest" and $\mathcal{H}_1$ "the noise-free signal is of interest" :

$$\mathcal{H}_0 : |s_{j,\mathbf{m}}^{(b)}(\mathbf{k})| \le T_{j,\mathbf{m}}^{(b)} \qquad \text{and} \qquad \mathcal{H}_1 : |s_{j,\mathbf{m}}^{(b)}(\mathbf{k})| > T_{j,\mathbf{m}}^{(b)}.$$

In other words, if $q_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = 1$, the coefficient $s_{j,\mathbf{m}}^{(b)}(\mathbf{k})$ is of interest. Therefore, it is possible to compute $\mathsf{P}\big(q_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = 1\big)$:

$$\mathsf{P}\big(q_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = 1\big) = 1 - \Gamma_{\text{inc}}\big(\lambda_{j,\mathbf{m}}^{(b)}(T_{j,\mathbf{m}}^{(b)})^{\beta_{j,\mathbf{m}}^{(b)}}, 1/\beta_{j,\mathbf{m}}^{(b)}\big)$$

where Γ_{inc} is the incomplete Gamma function. Hence, the following conditional probabilities can be easily derived :

$$p(s_{j,\mathbf{m}}^{(b)}(\mathbf{k})|q_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = 0) = \begin{cases} C_0 \exp(-\lambda_{j,\mathbf{m}}^{(b)}|s_{j,\mathbf{m}}^{(b)}(\mathbf{k})|^{\beta_{j,\mathbf{m}}^{(b)}}) & \text{if } |s_{j,\mathbf{m}}^{(b)}(\mathbf{k})| \le T_{j,\mathbf{m}}^{(b)} \\ 0 & \text{otherwise} \end{cases}$$

$$p\big(s_{j,\mathbf{m}}^{(b)}(\mathbf{k})|q_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = 1\big) = \begin{cases} 0 & \text{if } |s_{j,\mathbf{m}}^{(b)}(\mathbf{k})| \le T_{j,\mathbf{m}}^{(b)} \\ C_1 \exp\big(-\lambda_{j,\mathbf{m}}^{(b)}|s_{j,\mathbf{m}}^{(b)}(\mathbf{k})|^{\beta_{j,\mathbf{m}}^{(b)}}\big) & \text{otherwise} \end{cases}$$

where C_0 and C_1 are normalizing constants. For multivalued images, the authors proposed to exploit a local information from the different channels by defining the band

activity indicator $z_{j,\mathbf{m}}$ as the average over the B channels of the magnitudes of the B homologous noisy coefficients :

$$z_{j,\mathbf{m}}(\mathbf{k}) = \frac{1}{B} \sum_{b'=1}^{B} |r_{j,\mathbf{m}}^{(b')}(\mathbf{k})|.$$

Consequently, the local minimum mean square estimator is

$$\mathsf{E}[s_{j,\mathbf{m}}^{(b)}(\mathbf{k})|r_{j,\mathbf{m}}^{(b)}(\mathbf{k}), z_{j,\mathbf{m}}(\mathbf{k})]$$
$$= \mathsf{P}\big(q_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = 1|r_{j,\mathbf{m}}^{(b)}(\mathbf{k}), z_{j,\mathbf{m}}(\mathbf{k})\big)\mathsf{E}[s_{j,\mathbf{m}}^{(b)}(\mathbf{k})|r_{j,\mathbf{m}}^{(b)}(\mathbf{k}), q_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = 1]$$
$$+ \mathsf{P}\big(q_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = 0|r_{j,\mathbf{m}}^{(b)}(\mathbf{k}), z_{j,\mathbf{m}}(\mathbf{k})\big)\mathsf{E}[s_{j,\mathbf{m}}^{(b)}(\mathbf{k})|r_{j,\mathbf{m}}^{(b)}(\mathbf{k}), q_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = 0]$$

by assuming that the coefficients $(r_{j,\mathbf{m}}^{(b')}(\mathbf{k}))_{1\le b'\le B}$ are independent conditionally to $\mathcal{H}_0$ or $\mathcal{H}_1$. The sparseness of the wavelet representation allows us to consider that the second term takes very low values and to approximate $\mathsf{E}[s_{j,\mathbf{m}}^{(b)}(\mathbf{k})|r_{j,\mathbf{m}}^{(b)}(\mathbf{k}), q_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = 1]$ by $r_{j,\mathbf{m}}^{(b)}(\mathbf{k})$. Then, the following estimate (called ProbShrink) is derived for each subband $(j, \mathbf{m})$ and channel b :

$$\hat{s}_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = \mathsf{P}\big(q_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = 1|r_{j,\mathbf{m}}^{(b)}(\mathbf{k}), z_{j,\mathbf{m}}(\mathbf{k})\big)r_{j,\mathbf{m}}^{(b)}(\mathbf{k}).$$

After some manipulations, the explicit expression of the ProbShrink estimate is :

$$\hat{s}_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = \frac{\eta\big(r_{j,\mathbf{m}}^{(b)}(\mathbf{k})\big)\xi\big(r_{j,\mathbf{m}}^{(b)}(\mathbf{k})\big)\mu}{1 + \eta\big(r_{j,\mathbf{m}}^{(b)}(\mathbf{k})\big)\xi\big(r_{j,\mathbf{m}}^{(b)}(\mathbf{k})\big)\mu}r_{j,\mathbf{m}}^{(b)}(\mathbf{k})$$

where

$$\eta\big(r_{j,\mathbf{m}}^{(b)}(\mathbf{k})\big) = \frac{p\big(r_{j,\mathbf{m}}^{(b)}(\mathbf{k})|q_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = 1\big)}{p\big(r_{j,\mathbf{m}}^{(b)}(\mathbf{k})|q_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = 0\big)}$$

$$\xi\big(r_{j,\mathbf{m}}^{(b)}(\mathbf{k})\big) = \frac{p\big(z_{j,\mathbf{m}}(\mathbf{k})|q_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = 1\big)}{p\big(z_{j,\mathbf{m}}(\mathbf{k})|q_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = 0\big)}$$

$$\mu = \frac{\mathsf{P}\big(q_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = 1\big)}{1 - \mathsf{P}\big(q_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = 1\big)}.$$

In practice, it is used to set $T_{j,\mathbf{m}}^{(b)} = \sigma$ and the computation of the conditional densities $p\big(r_{j,\mathbf{m}}^{(b)}(\mathbf{k})|q_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = 1\big)$ and $p\big(r_{j,\mathbf{m}}^{(b)}(\mathbf{k})|q_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = 0\big)$ can be deduced from $p\big(s_{j,\mathbf{m}}^{(b)}(\mathbf{k})| q_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = 1\big)$ and $p\big(s_{j,\mathbf{m}}^{(b)}(\mathbf{k})|q_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = 0\big)$. Indeed, we have : $\forall \ell \in$

$\{0, 1\}$,

$$p\big(r_{j,\mathbf{m}}^{(b)}(\mathbf{k})|q_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = \ell\big)$$

$$= \int_{\mathbb{R}} g_{0,\sigma^2}(r_{j,\mathbf{m}}^{(b)}(\mathbf{k}) - s_{j,\mathbf{m}}^{(b)}(\mathbf{k}))p\big(s_{j,\mathbf{m}}^{(b)}(\mathbf{k})|q_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = \ell\big)ds_{j,\mathbf{m}}^{(b)}(\mathbf{k}),$$

where g_{0,σ^2} denotes the $\mathcal{N}(0, \sigma^2)$ probability density. For the statistical characterization of the band activity indicator $z_{j,\mathbf{m}}$, it is assumed that the coefficients $(r_{j,\mathbf{m}}^{(b')}(\mathbf{k}))_{1 \le b' \le B}$ are all distributed according to either $\mathcal{H}_0$ or $\mathcal{H}_1$. Therefore, the conditional density of $z_{j,\mathbf{m}}(\mathbf{k})$ given $q_{j,\mathbf{m}}^{(b)}(\mathbf{k})$ is the B times convolution of the conditional densities of $|r_{j,\mathbf{m}}^{(b')}(\mathbf{k})|$ given $q_{j,\mathbf{m}}^{(b)}(\mathbf{k})$.

8.4.4.3. *Gaussian scale mixture model*

Inspired by the work for monochannel images [POR 03], Scheunders and de Backer [SCH 07] have assumed that the unknown signal vector $\mathbf{s}_{j,\mathbf{m}}$ can be expressed as follows :

$$\mathbf{s}_{j,\mathbf{m}}(\mathbf{k}) = \sqrt{z_{j,\mathbf{m}}(\mathbf{k})}\mathbf{u}_{j,\mathbf{m}}(\mathbf{k})$$

where $\mathbf{u}_{j,\mathbf{m}}(\mathbf{k})$ is a zero-mean normal vector $\mathcal{N}(\mathbf{0}, \mathbf{\Gamma}^{(\mathbf{u}_{j,m})})$ and $z_{j,\mathbf{m}}(\mathbf{k})$ is a positive scalar random variable independent of $\mathbf{u}_{j,\mathbf{m}}(\mathbf{k})$. The prior probability density function of $\mathbf{s}_{j,\mathbf{m}}(\mathbf{k})$ can be considered as a Gaussian Scale Mixture (GSM) as we have :

$$p(\mathbf{s}_{j,\mathbf{m}}(\mathbf{k})) = \int_0^\infty p(\mathbf{s}_{j,\mathbf{m}}(\mathbf{k})|z_{j,\mathbf{m}}(\mathbf{k}))p(z_{j,\mathbf{m}}(\mathbf{k}))dz_{j,\mathbf{m}}(\mathbf{k}).$$

Indeed, $p(z_{j,\mathbf{m}}(\mathbf{k}))$ corresponds to a mixing density and $\mathbf{s}_{j,\mathbf{m}}(\mathbf{k})|z_{j,\mathbf{m}}(\mathbf{k}) \sim \mathcal{N}(\mathbf{0}, z_{j,\mathbf{m}}(\mathbf{k})\mathbf{\Gamma}^{(\mathbf{u}_{j,m})})$. The posterior mean estimate $\mathsf{E}[\mathbf{s}_{j,\mathbf{m}}(\mathbf{k})|\mathbf{r}_{j,\mathbf{m}}(\mathbf{k})]$ can then be expressed as :

$$\mathsf{E}[\mathbf{s}_{j,\mathbf{m}}(\mathbf{k})|\mathbf{r}_{j,\mathbf{m}}(\mathbf{k})] =$$
$$\int_0^\infty p(z_{j,\mathbf{m}}(\mathbf{k})|\mathbf{r}_{j,\mathbf{m}}(\mathbf{k}))\mathsf{E}[\mathbf{s}_{j,\mathbf{m}}(\mathbf{k})|\mathbf{r}_{j,\mathbf{m}}(\mathbf{k}), z_{j,\mathbf{m}}(\mathbf{k})]dz_{j,\mathbf{m}}(\mathbf{k}).$$

On the one hand, we note that the involved conditional posterior mean in the right hand side of the previous equation corresponds to the Wiener estimate :

$$\mathsf{E}[\mathbf{s}_{j,\mathbf{m}}(\mathbf{k})|\mathbf{r}_{j,\mathbf{m}}(\mathbf{k}), z_{j,\mathbf{m}}(\mathbf{k})] = z_{j,\mathbf{m}}(\mathbf{k})\mathbf{\Gamma}^{(\mathbf{u}_{j,m})}\big(z_{j,\mathbf{m}}(\mathbf{k})\mathbf{\Gamma}^{(\mathbf{u}_{j,m})} + \mathbf{\Gamma}^{(\mathbf{n})}\big)^{-1}\mathbf{r}_{j,\mathbf{m}}(\mathbf{k}).$$

On the other hand, Bayes rule allows us to express $p(z_{j,\mathbf{m}}(\mathbf{k})|\mathbf{r}_{j,\mathbf{m}}(\mathbf{k}))$ as :

$$p(z_{j,\mathbf{m}}(\mathbf{k})|\mathbf{r}_{j,\mathbf{m}}(\mathbf{k})) \propto p(\mathbf{r}_{j,\mathbf{m}}(\mathbf{k})|z_{j,\mathbf{m}}(\mathbf{k}))p(z_{j,\mathbf{m}}(\mathbf{k})).$$

In practice, in order to calculate this conditional probability, specifying the distribution of $z_{j,\mathbf{m}}(\mathbf{k})$ is required. In the case of monochannel images, Portilla *et al.* [POR 03] have proposed the (improper) Jeffrey's prior :

$$p(z_{j,\mathbf{m}}(\mathbf{k})) \propto 1/z_{j,\mathbf{m}}(\mathbf{k}).$$

In [SCH 07], another alternative has been considered when a single component noise-free image y is available, which is employed as an ancillary variable. It is assumed that the joint distribution of $(\mathbf{s}_{j,\mathbf{m}}(\mathbf{k}), y_{j,\mathbf{m}}(\mathbf{k}))$ is a GSM. Consequently, it is found that the probability of $\mathbf{s}_{j,\mathbf{m}}(\mathbf{k})$ conditioned on $y_{j,\mathbf{m}}(\mathbf{k})$ is a GSM. Then, by adopting the same strategy as previously (development of the conditional posteriori means and application of Bayes rule), it has been shown that the Bayesian estimate $\mathsf{E}[\mathbf{s}_{j,\mathbf{m}}(\mathbf{k})|\mathbf{r}_{j,\mathbf{m}}(\mathbf{k}), y_{j,\mathbf{m}}(\mathbf{k})]$ can be explicitly calculated.

8.4.5. *Variational approaches*

In wavelet-based variational approaches, the wavelet coefficients $(\overset{\wedge}{\mathbf{s}}_{j,\mathbf{m}})_{j,\mathbf{m}}$ of the denoised multichannel image $\overset{\wedge}{\mathbf{s}}$ are obtained by minimizing the objective function :

$$(\mathbf{u}_{j,\mathbf{m}})_{j,\mathbf{m}} \mapsto \frac{1}{2} \sum_{\mathbf{k}\in\mathbb{K}} \left((\mathbf{u}(\mathbf{k}) - \mathbf{r}(\mathbf{k}))^{\top} \mathbf{Q}^{-1}((\mathbf{u}(\mathbf{k}) - \mathbf{r}(\mathbf{k}))\right) + h\big((\mathbf{u}_{j,\mathbf{m}})_{j,\mathbf{m}}\big) \quad (8.6)$$

where $\mathbf{u}$ designates a generic multispectral image with B channels of size $L \times L$ and $(\mathbf{u}_{j,\mathbf{m}})_{j,\mathbf{m}}$ is its wavelet representation, $\mathbf{Q}$ is a definite positive matrix of size $B \times B$ and h is some appropriate function. The first quadratic term represents a data fidelity measure with respect to the observation model. The function h usually corresponds to a penalty term (also called regularization term) used to incorporate prior information about the original image $\mathbf{s}$. In the particular case when $\mathbf{Q} = \mathbf{\Gamma}^{(\mathbf{n})}$ and $\exp(-h(\cdot))$ is (up to a multiplicative constant) a probability density function, this approach amounts to a MAP approach where h is the potential associated with the prior distribution adopted for the wavelet coefficients of $\mathbf{s}$.

Some classical choices for h are :

$- h = i_C$ where i_C is the indicator function of a closed convex set C ($i_C\big((\mathbf{u}_{j,\mathbf{m}})_{j,\mathbf{m}}\big) = 0$ if $(\mathbf{u}_{j,\mathbf{m}})_{j,\mathbf{m}} \in C$ and $+\infty$ otherwise). This function imposes a hard constraint on the solution (e.g. the positivity of its pixel values).

$- h\big((\mathbf{u}_{j,\mathbf{m}})_{j,\mathbf{m}}\big) = \lambda \sum_{b=1}^{B} \sum_{\mathbf{k}\in\mathbb{K}} \|\nabla u^{(b)}(\mathbf{k})\|_2^2$ where $\nabla u^{(b)}$ is a discrete version of the gradient of $u^{(b)}$, $\|\cdot\|_2$ is the Euclidean norm of $\mathbb{R}^2$ and λ is a positive factor. This corresponds to a Tikhonov regularization [TIK 63], which serves to ensure the smoothness of the denoised image.

$- h\big((\mathbf{u}_{j,\mathbf{m}})_{j,\mathbf{m}}\big) = \lambda \sum_{b=1}^{B} \sum_{\mathbf{k}\in\mathbb{K}} \|\nabla u^{(b)}(\mathbf{k})\|_2$ which can be viewed as a discrete version of a Total Variation (TV) penalization [RUD 92, TEB 98]. A multivariate version of this function was also proposed in [BLO 98] (see also [AUJ 06] for extensions).

$- h\big((\mathbf{u}_{j,\mathbf{m}})_{j,\mathbf{m}}\big) = \sum_{b=1}^{B} \sum_{(j,\mathbf{m})} \lambda_{j,\mathbf{m}}^{(b)} \sum_{\mathbf{k}\in\mathbb{K}_j} |u_{j,\mathbf{m}}^{(b)}(\mathbf{k})|^{\beta_{j,\mathbf{m}}^{(b)}}$ where the exponents $\beta_{j,\mathbf{m}}^{(b)}$ and the scaling parameters $\lambda_{j,\mathbf{m}}^{(b)}$ are positive. This function is the potential of an independent Generalized Gaussian distribution for the wavelet coefficients [CHA 00, ANT 02]. Under some assumptions, it can also be interpreted as a regularity measure in terms of a Besov norm [CHA 98, LEP 01]. A particular case of interest is when the exponents are all equal to 1. The resulting ℓ^1 norm is used to promote the sparsity of the wavelet representation of the solution, by setting many of its coefficients to zero [TRO 06].

Note that a composite form consisting of a sum of several of the above functions can also be considered. Each of these penalizations may indeed present its own advantages and drawbacks. For example, the Tikhonov regularization tends to oversmooth image edges whereas the TV penalization may introduce visual staircase effects.
Another point to be emphasized is that the solution to the minimization of (8.6) can be obtained in different manners, according to the choice of h. Sometimes, an explicit expression of $\overset{\wedge}{\mathbf{s}}_{j,\mathbf{m}}$ can be derived. This arises, in particular, for Tikhonov regularization or, when $\mathbf{Q}$ is a diagonal matrix and a GG potential is used with $\beta_{j,\mathbf{m}}^{(b)} \in \{1, 4/3, 3/2, 2, 3, 4\}$ [CHA 07]. In other cases, iterative algorithms must be applied to compute $\overset{\wedge}{\mathbf{s}}_{j,\mathbf{m}}$. When a hard constraint is imposed, the constraint set C often is decomposed as the intersection of elementary closed convex sets and iterative techniques of projection onto each of these sets can be applied [COM 96, COM 03, COM 04]. One of such approaches is the well-known Projections Onto Convex Sets (POCS) algorithm. Provided that h is a convex function, other iterative convex optimization algorithms can be employed to bring efficient numerical solutions to the considered large-size variational problems (see [CHA 07, COM 07, COM 08] and references therein). These optimization methods are also useful when a redundant frame representation of the data is used instead of an orthonormal one.

8.4.6. *Stein-based approaches*

As already mentioned, the main problem in wavelet coefficient thresholding is the determination of the threshold value. Furthermore, in practice, it often appears preferable not to require some prior knowledge about the original data. A solution to alleviate such problems is to invoke Stein's principle [STE 81] which can be formulated as follows :

Proposition 1 *Using the notations of Section 8.4.1, let $f_{j,\mathbf{m}}^{(b)} : \mathbb{R}^d \to \mathbb{R}$ be a continuous, almost everywhere differentiable function such that :*

$$\forall \boldsymbol{\theta} \in \mathbb{R}^d,$$

$$\lim_{\|\mathbf{t}\| \to +\infty} f_{j,m}^{(b)}(\mathbf{t}) \exp\Big(-\frac{(\mathbf{t} - \boldsymbol{\theta})^\top (\boldsymbol{\Gamma}^{(\overline{\mathbf{n}}^{(b)})})^{-1} (\mathbf{t} - \boldsymbol{\theta})}{2} \Big) = 0;$$

$$\mathsf{E}[|f_{j,\mathbf{m}}^{(b)}(\overline{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k}))|^2] < +\infty \quad and \quad \mathsf{E}\Big[\Big\| \frac{\partial f_{j,\mathbf{m}}^{(b)}(\overline{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k}))}{\partial \overline{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k})}\Big\|\Big] < +\infty.$$

Then,

$$\mathsf{E}[f_{j,\mathbf{m}}^{(b)}(\overline{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k})) s_{j,\mathbf{m}}^{(b)}(\mathbf{k})] = \mathsf{E}[f_{j,\mathbf{m}}^{(b)}(\overline{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k})) r_{j,\mathbf{m}}^{(b)}(\mathbf{k})] -$$

$$\mathsf{E}\Big[\frac{\partial f_{j,\mathbf{m}}^{(b)}(\overline{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k}))}{\partial \overline{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k})}\Big]^\top \mathsf{E}[\overline{\mathbf{n}}_{j,\mathbf{m}}^{(b)}(\mathbf{k}) n_{j,\mathbf{m}}^{(b)}(\mathbf{k})]. \quad (8.7)$$

8.4.6.1. *Expression of the quadratic risk using Stein's formula*

A standard criterion in signal/image processing is the mean square error (also called quadratic risk) defined as

$$\mathsf{E}[\|\mathbf{s}_{j,\mathbf{m}}(\mathbf{k}) - \overset{\lambda}{\mathbf{s}}_{j,\mathbf{m}}(\mathbf{k})\|^2]$$

$$= \sum_{b=1}^{B} \mathsf{E}[|s_{j,\mathbf{m}}^{(b)}(\mathbf{k}) - f_{j,\mathbf{m}}^{(b)}(\overline{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k}))|^2]$$

$$= \sum_{b=1}^{B} \Big(\mathsf{E}[|s_{j,\mathbf{m}}^{(b)}(\mathbf{k})|^2] + \mathsf{E}[|f_{j,\mathbf{m}}^{(b)}(\overline{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k}))|^2] - 2\mathsf{E}[f_{j,\mathbf{m}}^{(b)}(\overline{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k})) s_{j,\mathbf{m}}^{(b)}(\mathbf{k})] \Big)$$

$$= \mathsf{E}[\|\mathbf{r}_{j,\mathbf{m}}(\mathbf{k}) - \overset{\lambda}{\mathbf{s}}_{j,\mathbf{m}}(\mathbf{k})\|^2] + 2 \sum_{b=1}^{B} \mathsf{E}\Big[\frac{\partial f_{j,\mathbf{m}}^{(b)}(\overline{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k}))}{\partial \overline{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k})}\Big]^\top \mathsf{E}[\overline{\mathbf{n}}_{j,\mathbf{m}}^{(b)}(\mathbf{k}) n_{j,\mathbf{m}}^{(b)}(\mathbf{k})].$$

The last equality has been obtained by using Stein's formula (see (8.7)). The advantage of the latter expression is that it only depends on the observed data and no longer on the unknown original data. This means that no prior information is necessary to build an estimate minimizing the quadratic risk. Another remark to be formulated is that the minimization of the risk will be performed by optimizing a limited number of variables parameterizing the estimator $f_{j,\mathbf{m}}^{(b)}$ (e.g. a threshold value). Due to the assumption of stationarity made at the end of Section 8.3, it can be noticed that the above risk is independent of $\mathbf{k}$, so that it will denoted $R_{j,\mathbf{m}}$ in the next paragraphs.

8.4.6.2. *Existing methods based on Stein's principle*

We will now present a variety of Stein based estimators ranging from component-wise thesholding rules to more sophisticated block-based methods.

– SUREshrink [DON 95] is a very popular componentwise method. It thus operates separately on each channel b. It consists of a soft-thresholding as given by (8.4), where subband-dependent threshold values $\chi_{j,\mathbf{m}}^{(b)}$ are computed in order to minimize the contribution $R_{j,\mathbf{m}}^{(b)}$ of the b-th channel to the quadratric risk $R_{j,\mathbf{m}}$. More exactly, the classical sample estimate $\tilde{R}_{j,\mathbf{m}}^{(b)}$ of $R_{j,\mathbf{m}}^{(b)}$ is employed. Then, the variables $|\mathbf{r}_{j,\mathbf{m}}^{(b)}(\mathbf{k})|$ are sorted in descending order, so that $|\mathbf{r}_{j,\mathbf{m}}^{(b)}(\mathbf{k}_1)| \geq |\mathbf{r}_{j,\mathbf{m}}^{(b)}(\mathbf{k}_2)| \geq \ldots \geq |\mathbf{r}_{j,\mathbf{m}}^{(b)}(\mathbf{k}_{L_j^2})|$. It can then be shown that if, for $i_0 \in \{2, \ldots, L_j^2\}$, $|\mathbf{r}_{j,\mathbf{m}}^{(b)}(\mathbf{k}_{i_0-1})| > \chi_{j,\mathbf{m}}^{(b)} \geq |\mathbf{r}_{j,\mathbf{m}}^{(b)}(\mathbf{k}_{i_0})|$, the risk estimate $\tilde{R}_{j,\mathbf{m}}^{(b)}$ is a second-order binomial increasing function of $\chi_{j,\mathbf{m}}^{(b)}$ over the considered interval. It can be deduced that the optimal threshold value over the interval $[|\mathbf{r}_{j,\mathbf{m}}^{(b)}(\mathbf{k}_{i_0})|, |\mathbf{r}_{j,\mathbf{m}}^{(b)}(\mathbf{k}_{i_0-1})|)$ is $|\mathbf{r}_{j,\mathbf{m}}^{(b)}(\mathbf{k}_{i_0})|$. So, the optimal threshold value over $\mathbb{R}_+$ is found by evaluating $\tilde{R}_{j,\mathbf{m}}^{(b)}$ on the finite set of candidate values $\{|\mathbf{r}_{j,\mathbf{m}}^{(b)}(\mathbf{k}_1)|, \ldots, |\mathbf{r}_{j,\mathbf{m}}^{(b)}(\mathbf{k}_{L_j^2})|, 0\}$.

– Starting from the Bayesian estimate given by (8.5), the authors in [BEN 05] derived a multivariate shrinkage estimator called SUREVECT. More precisely, the parameters $\epsilon_{j,\mathbf{m}}$ and $\mathbf{Q}_{j,\mathbf{m}}$ are directly adjusted so as to minimize $R_{j,\mathbf{m}}$. For simplicity, the matrix $\mathbf{\Gamma}_{j,\mathbf{m}}^{(s)}$ involved in the expression of the θ_1 function has not been included in the set of parameters estimated with a SURE procedure. An approximate value (like that obtained by a moment method) can be employed. It can be proved that

$$R_{j,\mathbf{m}} = \mathrm{tr}\big((\mathbf{Q}_{j,\mathbf{m}} - \mathbf{A}_{j,\mathbf{m}})\mathbf{C}_{j,\mathbf{m}}(\mathbf{Q}_{j,\mathbf{m}} - \mathbf{A}_{j,\mathbf{m}})^{\top} - \mathbf{A}_{j,\mathbf{m}}\mathbf{C}_{j,\mathbf{m}}\mathbf{A}_{j,\mathbf{m}}^{\top}\big)$$

with

$$
\begin{aligned}
\mathbf{A}_{j,\mathbf{m}} &\triangleq \mathbf{B}_{j,\mathbf{m}}\big(\mathbf{C}_{j,\mathbf{m}}\big)^{-1} \\
\mathbf{B}_{j,\mathbf{m}} &\triangleq \mathsf{E}[\gamma_{\epsilon_{j,\mathbf{m}}}(\mathbf{r}_{j,\mathbf{m}}(\mathbf{k}))\mathbf{r}_{j,\mathbf{m}}(\mathbf{k})(\mathbf{r}_{j,\mathbf{m}}(\mathbf{k}))^{\top}] - \mathbf{\Gamma}\big(\mathsf{E}[\gamma_{\epsilon_{j,\mathbf{m}}}(\mathbf{r}_{j,\mathbf{m}}(\mathbf{k}))]\mathbf{I}_B \\
&\quad + \mathsf{E}[\mathbf{\nabla}\gamma_{\epsilon_{j,\mathbf{m}}}(\mathbf{r}_{j,\mathbf{m}}(\mathbf{k}))(\mathbf{r}_{j,\mathbf{m}}(\mathbf{k}))^{\top}]\big) \\
\mathbf{C}_{j,\mathbf{m}} &\triangleq \mathsf{E}[\gamma_{\epsilon_{j,\mathbf{m}}}^2(\mathbf{r}_{j,\mathbf{m}}(\mathbf{k}))\mathbf{r}_{j,\mathbf{m}}(\mathbf{k})(\mathbf{r}_{j,\mathbf{m}}(\mathbf{k}))^{\top}]
\end{aligned}
$$

where it has been assumed that $\mathbf{C}_{j,\mathbf{m}}$ is invertible. It is easily shown that setting $\mathbf{Q}_{j,\mathbf{m}} = \mathbf{A}_{j,\mathbf{m}}$ allows us to minimize $R_{j,\mathbf{m}}$ and that the optimal value $\epsilon_{j,\mathbf{m}}$ within $[0, 1]$ should maximize $\mathrm{tr}(\mathbf{B}_{j,\mathbf{m}}\mathbf{C}_{j,\mathbf{m}}^{-1}\mathbf{B}_{j,\mathbf{m}}^{\top})$. It appears that this amounts to a mere optimization of a one-variable function which can be performed with conventional numerical optimization tools.

– More recently, we proposed an approach to take into account spatial and cross-channel dependences by using Sein's principle [CHA 08]. More precisely, the proposed estimator takes the form :

$$\hat{s}_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = f_{j,\mathbf{m}}^{(b)}(\bar{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k})) = \eta_{\chi_{j,\mathbf{m}}^{(b)}}(\|\bar{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k})\|^{\alpha_{j,\mathbf{m}}^{(b)}})\,(\boldsymbol{\rho}_{j,\mathbf{m}}^{(b)})^{\top}\bar{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k}),$$

where $\eta_{\chi_{j,\mathbf{m}}^{(b)}}(\cdot)$ is the thresholding function given by

$$\forall \tau \in \mathbb{R}_+, \qquad \eta_{\chi_{j,\mathbf{m}}^{(b)}}(\tau) = \left(\frac{\tau - \chi_{j,\mathbf{m}}^{(b)}}{\tau}\right)_+$$

and $\chi_{j,\mathbf{m}}^{(b)} \geq 0$, $\alpha_{j,\mathbf{m}}^{(b)} > 0$ and $\boldsymbol{\rho}_{j,\mathbf{m}}^{(b)} \in \mathbb{R}^d$. The vector $\boldsymbol{\rho}_{j,\mathbf{m}}^{(b)}$ corresponds to a linear parameter. This estimator generalizes the block-based estimator proposed in [CAI 01, ŞEN 02] as well as linear estimators. The parameters are computed similarly to the SUREShrink approach (more details can be found in [CHA 08]). Although the choice of the ROV is quite flexible, examples of some ROVs corresponding to spatial and across channel neigborhoods were considered in [CHA 08].

– Another efficient approach is SURE-LET (the SURE-Linear Expansion of Threshold [LUI 08]). Then, the considered expression of the estimator is

$$f_{j,\mathbf{m}}^{(b)}\left(r_{j,\mathbf{m}}^{(b)}(\mathbf{k})\right) = a_{j,\mathbf{m}}^{(b),1}\, r_{j,\mathbf{m}}^{(b)}(\mathbf{k}) + a_{j,\mathbf{m}}^{(b),2}\left(1 - \exp\left(-\frac{(r_{j,\mathbf{m}}^{(b)}(\mathbf{k}))^8}{(\omega_{j,\mathbf{m}}^{(b)})^8}\right)\right) r_{j,\mathbf{m}}^{(b)}(\mathbf{k})$$

where $\omega_{j,\mathbf{m}}^{(b)} > 0$ and $a_{j,\mathbf{m}}^{(b),1}$, $a_{j,\mathbf{m}}^{(b),2}$ are real-valued weighting factors. In the case when $(a_{j,\mathbf{m}}^{(b),1}, a_{j,\mathbf{m}}^{(b),2}) = (1,0)$ a linear estimator is obtained, whereas in the case when $(a_{j,\mathbf{m}}^{(b),1}, a_{j,\mathbf{m}}^{(b),2}) = (0,1)$ and $\omega_{j,\mathbf{m}}^{(b)} \gg 1$, a thresholding-like rule is obtained. Note that other possible choices of the nonlinear part of the estimator have been proposed in [PES 97, **?**]. This estimator possesses a number of attractive features :

- the determination of the optimal values of $a_{j,\mathbf{m}}^{(b),1}$ and $a_{j,\mathbf{m}}^{(b),2}$ minimizing $R_{j,\mathbf{m}}^{(b)}$ reduces to solving a set of linear equations.

- The use of a linear combination of more than two terms in the estimator expression can be addressed in a similar manner.

- The approach can be extended to the multichannel case [LUI 08] (interscale dependencies are also taken into account).

- When employing non-orthonormal representations (e.g. redundant frames), there is no equivalence between the minimization of $R_{j,\mathbf{m}}$ for each subband and the minimization of the mean square error in the space domain. However, for a SURE-LET estimator, it turns out that the minimization of the latter criterion can be easily carried out, so ensuring an optimal use of the flexibility offered by non-orthonormal representations.

8.5. Method comparisons

In this section, we will give some insights about the appropriateness of a noise removal method for a given application. Indeed, the previous survey has indicated that specifying a denoising method involves several choices such as :

– componentwise processing or joint processing,

– Bayesian strategy or non Bayesian approach,

– choice of the representation,

– computational complexity.

In what follows, we will briefly discuss these issues.

8.5.1. *Componentwise processing versus joint processing*

A great number of efficient denoising methods have been already developed for monochannel images. Hence, the temptation is strong to directly apply them to each component of a given noisy multichannel image. However, if correlations exist across the components, it is judicious to exploit them. A basic method is a two-step procedure which firstly removes such correlations through a principal component analysis or an independent component analysis and, then applies a monovariate denoising for each transformed component [GRE 88]. However, in general, it has been observed that jointly denoising the components outperforms the basic two-step method [BEN 05, LUI 08].

8.5.2. *Bayesian strategy versus non Bayesian approach*

Wavelet transforms are known to generate very compact representations of regular signals. This is an appealing property which simplifies the statistical prior modelling required in the Bayesian framework. In this respect, for both componentwise and joint processing, it is important to note that the wavelet coefficients have been modelled in different ways :

– marginal probabilistic models for the subband coefficients,

– joint probabilistic models accounting for interscale dependencies between the subband coefficients [ŞEN 02], [ROM 01, ELM 05b] or for spatial intrascale dependencies [MAL 97], [CHA 08].

The more accurate the prior distribution, the better the performance. Concerning multivariate estimators, it has been found that Bernoulli-Gaussian and Gaussian Scale Mixtures priors behave similarly in terms of signal-to-noise ratios for natural images, and they are clearly better than Gaussian priors [DEB 08].
It is worth pointing out that the Bayesian approach presents two main drawbacks. Firstly, there may exist a mismatch between the data and the model and, the structure of the estimator may thus be inappropriate. Secondly, the values of the hyperparameters may be suboptimal as they are derived from statistical inference exploiting the prior model. These shortcomings have made attractive non Bayesian methods. In the case of Stein-based approaches however, the choice of the estimator form is a key

issue. In [BEN 05], it has been shown that SUREVECT outperforms the tested Bayesian denoising methods. Combination of nonlinear functions (LET) also allows us to define a flexible class of estimators.

8.5.3. *Choice of the representation*

Very often, the performance of a denoising method (for a given noise level) is measured in terms of the average $\mathrm{MSE}_{\mathrm{im}}$ of the mean square errors $\mathrm{MSE}_{\mathrm{im}}^{(b)}$ in each channel b :

$$\mathrm{MSE}_{\mathrm{im}} = \frac{1}{B} \sum_{b=1}^{B} \mathrm{MSE}_{\mathrm{im}}^{(b)} \quad \text{with} \quad \mathrm{MSE}_{\mathrm{im}}^{(b)} = \frac{1}{L^2} \sum_{\mathbf{k} \in \mathbb{K}} |s^{(b)}(\mathbf{k}) - \overset{\curlywedge (b)}{s}(\mathbf{k})|^2.$$

When the noise is removed in the WT domain, the average mean square error $\mathrm{MSE}_{\mathrm{tr}}$ of the B mean square errors $\mathrm{MSE}_{\mathrm{tr}}^{(b)}$ in the transform domain is very often employed :

$$\mathrm{MSE}_{\mathrm{tr}} = \frac{1}{B} \sum_{b=1}^{B} \mathrm{MSE}_{\mathrm{tr}}^{(b)} \text{ with } \mathrm{MSE}_{\mathrm{tr}}^{(b)} = \frac{1}{L^2} \sum_{(j,\mathbf{m})} \sum_{\mathbf{k} \in \mathbb{K}_j} |s_{j,\mathbf{m}}^{(b)}(\mathbf{k}) - \overset{\curlywedge (b)}{s}_{j,\mathbf{m}}(\mathbf{k})|^2.$$

Notice however that only some of the reported methods explicitly introduce the mean square error criterion in the optimization process for the related estimator.

On the one hand, it must be pointed out that the minimization of the mean square error between the wavelet coefficients is equivalent to the minimization of the mean square error in the image domain if and only if the wavelet representation is orthonormal. Consequently, when the image is decomposed on a redundant frame, the minimization of $\mathrm{MSE}_{\mathrm{tr}}$ is suboptimal. On the other hand, it has been noticed by many authors that better results are often obtained with redundant decompositions. One of the main reasons for this fact is that these transforms have better translation invariance properties, which is beneficial to the reduction of Gibbs-like visual artifacts. Especially, curvelet [CAN 99] and dual-tree wavelet decompositions [SEL 05, CHA 06] have provided recent examples of multiscale transforms giving rise to redundant representations which are especially good for denoising purposes.

The suboptimality of the denoising on a redundant representation has been alleviated for LET estimators since it has been proved that it is possible to directly minimize an unbiased estimate of $\mathrm{MSE}_{\mathrm{im}}$ [LUI 08].

8.5.4. *Computational complexity*

As noticed in [DEB 08], it is clear that the computational load is increased by a joint processing rather than separate ones. Besides, in Bayesian approaches, the hyperparameter estimation may imply a high computational cost. Despite their versatility, variational methods requiring iterative solutions are also time-consuming.

Altough they yield improved results, redundant representations involve a significant increase of the computational complexity when the redundancy factor is high.

8.6. Conclusions and perspectives

We have given a broad overview of multivariate image denoising methods, with a special emphasis on wavelet-based approaches, since wavelets efficiently represent image features (both textures and edges). Under different noise correlation models, a variety of transform domain estimators have been described, ranging from Wiener to Stein-based via Bayesian or variational approaches. While most of these methods have been presented for orthonormal wavelet bases, some of them can be applied to more general frames. We would also like to mention that we do not have provided an exhaustive list of the available denoising methods. For example, the reader is refered to [BUA 05] for a description of other non-wavelet based denoising approaches.

Other degradations such as blurring may also affect multichannel images, essentially due to the optical sensor. Recent works now aim at developing advanced wavelet-based methods for restoring multichannel images degraded by a blur *and* an additive noise.

[ABR 98] ABRAMOVICH F., SAPATINAS T., SILVERMAN B. W., « Wavelet thresholding via a Bayesian approach », *Journal of the Royal Statistical Society*, vol. 60, p. 725-749, 1998, Series B.

[ABR 02] ABRAMS M. C., CAIN S. C., « Sampling, radiometry and image reconstruction for polar and geostationary meteorological remote sensing systems », *Image Reconstruction from Incomplete Data II Proc. SPIE*, vol. 4792, p. 207-215, 2002, P. J. Bone, M. A. Fiddy, R. P. Milane Ed.

[ANG 91] ANGELOPOULOS G., PITAS I., « Multichannel Wiener filters in color image restoration based on AR color image modelling », *Proc. Int. Conf. on Acoust., Speech and Sig. Proc.*, vol. 4, Toronto, Canada, p. 2517–2520, 14-17 Apr. 1991.

[ANT 02] ANTONIADIS A., LEPORINI D., PESQUET J.-C., « Wavelet thresholding for some classes of non-Gaussian noise », *Statistica Neerlandica*, vol. 56, n°4, p. 434–453, Dec. 2002.

[ATK 03] ATKINSON I., KAMALABADI F., JONES D. L., DO M. N., « Adaptive wavelet thresholding for multichannel signal estimation », UNSER M. A., ALDROUBI A., LAINE A. F., Eds., *Proc. of SPIE Conf. on Wavelet Applications in Signal and Image Processing X*, vol. 5207, San Diego, CA, USA, p. 28–39, Aug. 2003.

[AUJ 06] AUJOL J.-F., KANG S. H., « Color image decomposition and restoration », *Journal of Visual Communication and Image Representation*, vol. 17, n°4, p. 916–928, Aug. 2006.

[BEN 03] BENAZZA-BENYAHIA A., PESQUET J.-C., « Wavelet-based multispectral image denoising with Bernoulli-Gaussian models », *IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing*, Grado, Italy, Jun. 8-11 2003.

(a) SNR = −2.91 dB

(b) SNR = 4.07 dB

(c) SNR = 8.33 dB

(d) SNR = 8.59 dB

(e) SNR = 8.66 dB

(f) SNR = 8.87 dB

Figure 8.8. *Denoising results on the first component of a Landsat 7 satellite image (see Fig. 8.2). (a) Degraded image and restored images using (b) Wiener filter, (c) BLS-GSM [POR 03], (d) Probshrink [PIŽ 06], (e) SUREVECT [BEN 05] and (f) Block-based wavelet estimate [CHA 08]*

[BEN 05] BENAZZA-BENYAHIA A., PESQUET J.-C., « Building robust wavelet estimators for multicomponent images using Stein's principle », *IEEE Trans. on Image Proc.*, vol. 14, n°11, p. 1814–1830, Nov. 2005.

[BLO 98] BLOMGREN P. V., CHAN T. F., « Color TV : Total variation methods for restoration of vector valued images », *IEEE Trans. on Image Proc.*, vol. 7, n°3, p. 304–309, Mar. 1998.

[BUA 05] BUADES A., COLL B., MOREL J. M., « A review of image denoising algorithms, with a new one », *Multiscale Modeling and Simulation*, vol. 4, n°2, p. 490–530, 2005.

[CAI 01] CAI T. T., SILVERMAN B. W., « Incorporating information on neighboring coefficients into wavelet estimation », *Sankhya*, vol. 63, Series B., p. 127–148, 2001.

[CAN 99] CANDÈS E., DONOHO D. L., « Curvelets - a surprisingly effective nonadaptive representation for objects with edges », C. RABUT A. C., SCHUMAKER L. L., Eds., *Curves and Surfaces*, p. 105–120, Vanderbilt University Press, Nashville, TN, USA, 1999.

[CAN 06] CANDÈS E., DEMANET L., DONOHO D., YING L., « Fast discrete curvelet transforms », *Multiscale Modeling and Simulation*, vol. 5, n°3, p. 861–899, Mar. 2006.

[CHA 98] CHAMBOLLE A., DEVORE R. A., LEE N. Y., LUCIER B. J., « Nonlinear wavelet image processing : Variational problems, compression, and noise removal through wavelet shrinkage », *IEEE Trans. on Image Proc.*, vol. 7, 1998.

[CHA 00] CHANG S. G., YU B., VETTERLI M., « Adaptive wavelet thresholding for image denoising and compression », *IEEE Trans. on Image Proc.*, vol. 9, n°9, p. 1532–1546, Sep. 2000.

[CHA 06] CHAUX C., DUVAL L., PESQUET J.-C., « Image analysis using a dual-tree M-band wavelet transform », *IEEE Trans. on Image Proc.*, vol. 15, n°8, p. 2397–2412, Aug. 2006.

[CHA 07] CHAUX C., COMBETTES P. L., PESQUET J.-C., WAJS V. R., « A variational formulation for frame based inverse problems », *Inverse Problems*, vol. 23, p. 1495–1518, juin 2007.

[CHA 08] CHAUX C., DUVAL L., BENAZZA-BENYAHIA A., PESQUET J.-C., « A Nonlinear Stein Based Estimator for Multichannel Image Denoising », *IEEE Trans. on Signal Proc.*, vol. 56, n°8, p. 3855–3870, Aug. 2008.

[COH 92] COHEN A., DAUBECHIES I., FEAUVEAU J.-C., « Biorthogonal bases of compactly supported wavelets », *Comm. ACM*, vol. 45, p. 485–560, 1992.

[COI 92] COIFMAN R. R., MEYER Y., WICKERHAUSER M. V., « Wavelet analysis and signal processing », *Wavelets and their applications*, p. 153–178, Jones and Bartlett, Boston, MA, 1992.

[COI 95] COIFMAN R., DONOHO D., « Translation-invariant de-noising », ANTONIADIS A., OPPENHEIM G., Eds., *Wavelets and Statistics*, vol. 103 de *Lecture Notes in Statistics*, p. 125–150, Springer, New York, NY, USA, 1995.

[COM 96] COMBETTES P. L., « The Convex Feasibility Problem in Image Recovery », *Advances in Imaging and Electron Physics*, vol. 95, p. 155–270, New York : Academic Press, 1996.

[COM 03] COMBETTES P. L., « A block-iterative surrogate constraint splitting method for quadratic signal recovery », *IEEE Trans. on Signal Proc.*, vol. 51, 2003.

[COM 04] COMBETTES P. L., PESQUET J.-C., « Wavelet-constrained image restoration », *International Journal on Wavelets, Multiresolution and Information Processing*, vol. 2, n°4, Dec. 2004.

[COM 07] COMBETTES P. L., PESQUET J.-C., « A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery », *IEEE J. Selected Topics Signal Process.*, vol. 1, 2007.

[COM 08] COMBETTES P. L., PESQUET J.-C., « A proximal decomposition method for solving convex variational inverse problems », *Inverse Problems*, vol. 24, Dec. 2008, to appear.

[COR 03] CORNER B. R., NARAJANAN R. M., REICHENBACH S. E., « Noise estimation in remote sensing imagery using data masking », *Internat. J. Remote Sensing*, vol. 24, n°4, p. 689–702, 2003.

[CRO 76] CROISIER A., ESTEBAN D., C. G., « Perfect channel splitting by use of interpolation/decimation/tree decomposition techniques », *Int. Conf. on Inform. Science and Systems*, Patras, Greece, p. 443–446, Aug. 1976.

[DAU 92] DAUBECHIES I., *Ten Lectures on Wavelets*, CBMS-NSF, SIAM Lecture Series, Philadelphia, PA, USA, 1992.

[DEB 08] DE BACKER S., PIŽURICA A., HUYSMANS B., PHILIPS W., SCHEUNDERS P., « Denoising of multicomponent images using wavelet least-squares estimators », *Image and Vision Computing*, vol. 26, n°7, p. 1038–1051, Jul. 2008.

[DO 05] DO M. N., VETTERLI M., « The contourlet transform : an efficient directional multi-resolution image representation », *IEEE Trans. on Image Proc.*, vol. 14, n°12, p. 2091–2106, Dec. 2005.

[DON 93] DONOHO D. L., « Unconditional bases are optimal bases for data compression and for statistical estimation », *Appl. and Comp. Harm. Analysis*, vol. 1, n°1, p. 100–115, Dec. 1993.

[DON 94] DONOHO D. L., JOHNSTONE I. M., « Ideal spatial adaptation by wavelet shrinkage », *Biometrika*, vol. 81, p. 425–455, Sept. 1994.

[DON 95] DONOHO D. L., JOHNSTONE I. M., « Adapting to unknown smoothness via wavelet shrinkage », *J. American Statist. Ass.*, vol. 90, p. 1200–1224, Dec. 1995.

[ELM 05a] ELMZOUGHI A., BENAZZA-BENYAHIA A., PESQUET J.-C., « An interscale multivariate statistical model for MAP multicomponent image denoising in the wavelet transform domain », *Proc. Int. Conf. on Acoust., Speech and Sig. Proc.*, vol. 2, Philadelphia, PA, USA, p. 45–48, March 18-23 2005.

[ELM 05b] ELMZOUGHI A., BENAZZA-BENYAHIA A., PESQUET J.-C., « An interscale multivariate statistical model for MAP multicomponent image denoising in the wavelet transform domain », *Proc. Int. Conf. on Acoust., Speech and Sig. Proc.*, vol. 2, Philadelphia, USA, p. 45–48, Mar. 18-23 2005.

[FLA 98] FLANDRIN P., *Time-frequency and time-scale analysis*, Academic Press, San Diego, USA, 1998.

[GRE 88] GREEN A., BERMAN M., P.SWITZER, CRAIG M. D., « A transformation for ordering multispectral data in terms of image quality with implications for noise removal », *IEEE Trans. on Geoscience and Remote Sensing*, vol. 26, n°1, p. 65–74, Jan. 1988.

[GUY 90] GUYOT G., « Optical properties of vegetation canopies », STEVEN M. D., CLARK J. A., Eds., *Applications of Remote Sensing in Agriculture*, p. 19–43, Butterworth, 1990.

[HAA 10] HAAR A., « Zur Theory der orthogalen Funktionen Systeme », *Math. Annalen*, vol. 69, p. 331–371, 1910.

[HEI 06] HEIL C., WALNUT D. F., *Fundamental papers in wavelet theory*, Princeton University Press, 2006.

[JAI 89] JAIN A. K., *Fundamentals of digital image processing*, Englewood Cliffs, NJ : Prentice Hall, 1989.

[LAN 86] LANDGREBE D. A., MALARET E., « Noise in Remote-Sensing Systems : the effect on classification error », *IEEE Trans. on Geoscience and Remote Sensing*, vol. GE-24, n°2, p. 294–300, 1986.

[LAN 00] LANDGREBE D., « Information extraction principles and methods for multispectral and hyperspectral image data », CHEN C. H., Ed., *Information processing for remote sensing*, p. 3–38, World Scientific Publishing Co., Inc., NJ, USA, 2000.

[LEP 99] LEPORINI D., PESQUET J.-C., KRIM H., « Best basis representations based on prior statistical models », MÜLLER P., VIDAKOVIC B., Eds., *Bayesian Inference in wavelet based models*, vol. 141 de *Lecture notes in statistics*, p. 155–172, Springer, 1999.

[LEP 01] LEPORINI D., PESQUET J.-C., « Bayesian wavelet denoising : Besov priors and non-Gaussian noises », *Signal Processing*, vol. 81, p. 55–67, 2001.

[LUI 08] LUISIER F., BLU T., « SURE-LET multichannel image denoising : interscale orthonormal wavelet thresholding », *IEEE Trans. on Inform. Theory*, vol. 17, n°4, p. 482–492, April 2008.

[MAL 97] MALFAIT M., ROSSE D., « Wavelet-based image denoising using a Markov random field a priori model », *IEEE Trans. on Image Proc.*, vol. 6, n°4, p. 549–565, Apr. 1997.

[MAL 98] MALLAT S., *A wavelet tour of signal processing*, Academic Press, San Diego, CA, USA, 1998.

[MAL 08] MALLAT S., « Geometrical grouplets », *Appl. and Comp. Harm. Analysis*, 2008, To appear.

[MEY 90] MEYER Y., *Ondelettes et opérateurs. I*, Actualités mathématiques., Hermann, Paris, France, 1990.

[PES 96] PESQUET J.-C., KRIM H., CARFANTAN H., « Time-invariant orthogonal wavelet representations », *IEEE Trans. on Signal Proc.*, vol. 44, n°8, p. 1964–1970, Aug. 1996.

[PES 97] PESQUET J.-C., LEPORINI D., « A new wavelet estimator for image denoising », *IEE Sixth Int. Conf. Im. Proc. Appl.*, vol. 1, p. 249–253, Jul. 14-17 1997.

[PIT 90] PITAS I., VENETSANOPOULOS A., *Nonlinear digital filters*, Kluwer, Dodrecht, 1990.

[PIŽ 06] PIŽURICA A., PHILIPS W., « Estimating the probability of the presence of a signal of interest in multiresolution single- and multiband image denoising », *IEEE Trans. on Image Proc.*, vol. 15, n°3, p. 654–665, 2006.

[POR 03] PORTILLA J., STRELA V., WAINWRIGHT M. J., SIMONCELLI E. P., « Image denoising using scale mixtures of Gaussians in the wavelet domain », *IEEE Trans. on Image Proc.*, vol. 12, n°11, p. 1338–1351, Nov. 2003.

[RAP 08] RAPHAN M., SIMONCELLI E., « Optimal Denoising in Redundant Representations », *IEEE Trans. on Image Proc.*, vol. 17, n°8, p. 1342–1352, Aug. 2008.

[ROM 01] ROMBERG J. K., CHOI H., BARANIUK R. G., « Bayesian tree-structured image modeling using wavelet-domain hidden Markov models », *IEEE Trans. on Image Proc.*, vol. 10, n°7, p. 1056–1068, Jul. 2001.

[RUD 92] RUDIN L. I., OSHER S., FATEMI E., « Nonlinear total variation based noise removal algorithms », *Physica D*, vol. 60, n°1-4, p. 259–268., 1992.

[SCH 04] SCHEUNDERS P., « Wavelet thresholding of multivalued images », *IEEE Trans. on Image Processing*, vol. 13, n°4, p. 411-416, April 2004.

[SCH 07] SCHEUNDERS P., DE BACKER S., « Wavelet denoising of multicomponent images using Gaussian scale mixture models and a noise-free image as priors », *IEEE Trans. on Image Proc.*, vol. 16, n°7, p. 1865–1872, Jul. 2007.

[SEL 05] SELESNICK I. W., BARANIUK R. G., KINGSBURY N. G., « The dual-tree complex wavelet transform », *IEEE Signal Processing Magazine*, vol. 22, n°6, p. 123–151, Nov. 2005.

[ŞEN 02] ŞENDUR L., SELESNICK I. W., « Bivariate shrinkage with local variance estimation », *Signal Processing Letters*, vol. 9, n°12, p. 438–441, Dec. 2002.

[SIM 96] SIMONCELLI E. P., ADELSON E. H., « Noise removal via Bayesian wavelet coring », *Proc. Int. Conf. on Image Processing*, vol. 1, Lausanne, Switzerland, p. 379–382, Sep. 16-19 1996.

[SMI 84] SMITH M., BARNWELL T., « A procedure for designing exact reconstruction filter banks for tree structured subband coders », *Proc. Int. Conf. on Acoust., Speech and Sig. Proc.*, vol. 9, San Diego, CA, USA, p. 421–424, March 19-21 1984.

[STE 81] STEIN C., « Estimation of the mean of a multivariate normal distribution », *Annals of Statistics*, vol. 9, n°6, p. 1135–1151, 1981.

[STE 93] STEFFEN P., HELLER P. N., GOPINATH R. A., BURRUS C. S., « Theory of regular M-band wavelet bases », *IEEE Trans. on Signal Proc.*, vol. 41, n°12, p. 3497–3511, Dec. 1993.

[TEB 98] TEBOUL S., BLANC-FÉRAUD L., AUBERT G., BARLAUD M., « Variational approach for edge-preserving regularization using coupled PDE's », *IEEE Trans. on Image Proc.*, vol. 7, Mar. 1998.

[TIK 63] TIKHONOV A. N., « On the solution of ill-posed problems and the method of regularization », *Dokl. Akad. Nauk SSSR*, vol. 151, 1963.

[TRO 06] TROPP J. A., « Just relax : Convex programming methods for identifying sparse signals in noise », *IEEE Trans. on Inform. Theory*, vol. 52, p. 1030–1051, 2006.

[VAI 87] VAIDYANATHAN P. P., « Theory and design of M-channel maximally decimated quadrature mirror filters with arbitrary M, having the perfect-reconstruction property », *IEEE Trans. on Acous., Speech and Signal Proc.*, vol. 35, n°4, p. 476–492, Apr. 1987.

[WIE 49] WIENER N., *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*, Cambridge, Technology Press of Massachusetts Institute of Technology, and New York, Wiley, 1949.

Chapitre 9

Unsupervised classification for multivariate images

9.1. Introduction

Unsupervised classification or clustering is a common technique for exploratory data analysis. It has been applied with success in many data analysis problems such as computer vision, marketing, biology, medicine and chemistry. Since in many applications there is no a priori information about the data, unsupervised methods are the most appropriate tools to reveal the relationships between data. A good clustering algorithm should reveal the number of classes automatically ; however, there are few techniques which could furnish this information without a priori knowledge. Generally, the appropriate number of classes is estimated by comparing some validity indices for partitions resulted after different parameterizations of the same algorithm.

A classical application of clustering algorithms is to perform image segmentation by pixel classification. Image segmentation consists in partitioning a source image into homogeneous regions according to some homogeneity criterion defined on image's attributes (intensity, color, texture). A multivariate image is formed by a set of images, each of them describing the same scene from a different perspective. In this way, one image represents a different variable or attribute of the multivariate image. Almost all physical measures can be used to obtain multivariate images : temperature, polarization, mass, wavelength etc. In multivariate images, pixels represent the answer of a small region from the analyzed scene to a particular excitation. Depending on the image resolution and the scene to be analyzed, one pixel may include more than

one object (or chemical compounds, or minerals etc.). It is the case of remote sensing images or microscopy images where a single pixel covers a large surface which may include different objects or chemical compounds. So it is not unrealistic to use methods based on a linear mixing model which could reveal hidden variables with a realistic meaning. These methods are often used as pre-processing step for dimension reduction. More than that, they render the clustering results meaningful.

Following, the similarity measures for multivariate data clustering are discussed in Section 9.2. Next, some dimension reduction methods are presented and we continue by presenting some probability based clustering methods. The application of Parzen-Watershed clustering method to multivariate image segmentation is presented in Section 9.5 and finally we conclude.

9.2. Similarity measures for pixel clustering

The core of almost all clustering methods is represented by the notion of similarity (or dissimilarity) between data. There is not a unique definition in the literature for this concept resulting in a multitude of similarity measures which, applied on the same data set could lead to completely different clustering results. However, few of them are often used by clustering algorithms and we present them in the following.

9.2.1. *Minkowski metrics*

The most common similarity measure for multivariate data is the Euclidean distance which is a particular case of the Minkowski metrics. Minkowski metrics are also defined as being particular instances of the r metric family. Let $X_i = (x_{i1}, ..., x_{id})$ and $X_k = (x_{k1}, ..., x_{kd})$ be two multivariate data of dimension d, then the r metric family is defined by :

$$L_r(X_i - X_k) = \left(\sum_{j=1}^{d} |x_{ij} - x_{kj}|^r \right)^{1/r} \tag{9.1}$$

where $r \in \mathbb{R}_+$, $r \geq 1$. Among all Minkowski metrics, Manhattan distance ($r = 1$) and Euclidian distance ($r = 2$) are often used to solve multivariate data clustering problems.

All Minkowski metrics satisfy the following properties :
1) $L_r(X_i, X_i) = 0, \forall (i)$;
2) $L_r(X_i, X_k) = L_r(X_k, X_i), \forall (i, k)$;
3) $L_r(X_i, X_k) \geq 0, \forall (i, k)$;

4) $L_r(X_i, X_k) = 0$, if and only if $(X_i = X_k)$;

5) $L_r(X_i, X_k) \leq L_r(X_i, X_m) + L_r(X_m, X_k), \forall(i, m, k)$;

In [GOW 86] it is shown that any metric defined by a particular instance of r should only satisfy properties 1) and 4) in order to be used as a similarity measure. Lately, metrics for which the triangle inequality (condition 5) is not satisfied were proposed and used to solve problems of multivariate data clustering. These metrics are called *fractionary metrics* and they are defined for $r \in \mathbb{R}_+$, $0 < r < 1$. In [C.C 01, ALE 99, K.B 00] it has been shown that fractionary metrics are less concentrated than Minkowski metrics for uniform distributions of data. So the main reason for which they are used as a similarity measure for multivariate data is to overcome the *concentration phenomenon* which affects all r metrics [D.F 07]. Hence their utility for improving clustering results is not yet well established since it has been shown that Minkowski metrics may be less concentrated than fractionary metrics for non uniform data distributions [D.F 07].

If we consider multivariate objects as one dimensional vectors, it is important to note that Euclidian distance in particular and all r metrics in general, offer information concerning the similarity from their amplitude point of view. In multivariate image analysis this means to compare the absolute intensity of pixels. These metrics do not offer a complete description of the similarity between pixels since the correlation considerations are ignored.

9.2.2. *Mahalanobis distance*

One way to take into account the correlation between multivariate data is to use the Mahalanobis distance [MAH 36] as a similarity measure. For two multivariate data X_i and X_k, it is defined as follows :

$$D(X_i, X_k) = (X_i - X_k)\Sigma^{-1}(X_i - X_k)^T \tag{9.2}$$

where Σ is the covariance matrix. If Σ is the identity matrix, then the Euclidean distance is obtained.

9.2.3. *Spectral angle*

The spectral angle is the d-dimensional extension of the two-dimensional angle. If we consider pixels as being points into a d-dimensional space and if for each point we draw a vector to the origin, then the spectral angle measures the angle between pairs

of such vectors ; small values of the spectral angle correspond to good similarities between pixels. For two d-dimensional vectors X_i and X_k, the spectral angle is defined by :

$$\alpha = \arccos \frac{X_i^T X_k}{\|X_i\|\|X_k\|} \tag{9.3}$$

The spectral angle is used as similarity measure by the Spectral Angle Mapper algorithm [F.A 93]. This method computes the spectral angle between each pixel and a target spectrum. It is to note that the spectral angle is not sensitive to the difference of pixel illumination ; this is an important drawback since pixels of different intensities but good correlation will be regrouped into the same class. From another point of view, the spectral angle is an important tool to measure the linear correlation between pixels' spectra. Hence it is not able to distinguish between positive and negative correlations.

9.2.4. *Pearson's correlation coefficient*

Pearson's correlation coefficient overcomes this drawback since it considers the mean extracted data. It is defined as follows :

$$R = \frac{(X_i - \bar{X}_i)(X_k - \bar{X}_k)}{\sqrt{\sum_{j=1}^{d}(x_{ij} - \bar{X}_i)^2 \sum_{j=1}^{d}(x_{kj} - \bar{X}_k)^2}} \quad \text{where} \quad \bar{X}_i = \sum_{j=1}^{d} x_{ij} \tag{9.4}$$

The Spectral Correlation Mapper algorithm [O.C 00] is an improvement of the Spectral Angle Mapper which uses the Pearson's correlation coefficient as a similarity measure. As well as the spectral angle, Pearson's correlation coefficient cannot distinguish between pixels with similar spectra but different intensities.

9.3. Dimension reduction techniques

Dimension reduction techniques are an important pre-processing step for multivariate data clustering. There are two main reasons to perform dimension reduction before gathering data into classes. The first one is that multivariate data include many redundant information as well as noise, so relevant information is often hidden by attributes with no interest for user. The second reason is to avoid the so called "*curse of dimensionality*" also known as "*the empty space phenomenon*" [D.W 83]. We may summarize the benefits of dimension reduction methods for multivariate data clustering by the following :

– reduction of redundant information,

– avoiding the empty space phenomenon,

– computational complexity reduction,

– identifying hidden factors in the data set (*e.g.* the presence of mineral species, chemical compounds or other regions of interest which may have a physical meaning),

– data visualization.

In [FOD 02], the dimension reduction problem is defined as follows : given a d-dimensional random vector $x = (x_1, \ldots, x_d)$, find a lower dimensional representation of it, $s = (s_1, \ldots, s_k)$ with $k < d$ that captures the content of the original data, according to some criterion. A set of multivariate data X contains n realizations of the d dimensional random vector x. For multivariate image applications, one pixel is modeled by a d dimensional random vector x and a multivariate image by a $n \times d$ matrix X. The components of x are called *features* or *attributes* while the components of s are called *hidden variables*.

There are two major categories of methods to perform dimension reduction : feature selection and feature extraction. Generally, feature selection methods are used when the class label attribute is known, while for multivariate data clustering, these methods become computationally expensive. For these applications, feature extraction methods are preferred for dimension reduction.

Two major types of feature extraction techniques exist in the literature : linear and non-linear. Linear methods consider that each observation vector x is a linear combination of some hidden variables s :

$$x_i = w_{i,1}s_{1i} + \ldots + w_{i,K}s_{K,i} \tag{9.5}$$

or

$$x_i = \sum_{k=1}^{K} w_{i,k}s_{ki} \tag{9.6}$$

with $w_{i,k}$ being a weight matrix. For a $n \times d$ observation matrix X we obtain :

$$X = WS \tag{9.7}$$

The components of S are obtained by projecting the n observations of X onto k directions estimated according to some criterion defined on the observation matrix X. There are methods based on second-order statistics (Principal Component Analysis - PCA or Factor Analysis - FA) while others are using higher order statistics (Independent Component Analysis - ICA). The main drawback of these methods for multivariate image analysis is that they do not take into account the non-negativity constraint which makes that the results provided by them do not allow a physical interpretation. Blind source separation methods (BSS) which take into account the non negativity constraint such as non-negative matrix factorization algorithms, overcome this drawback providing meaningful factors ; such methods do not make use of the data statistics and the only constraint is the non negativity of the results. Following we will briefly present some linear dimension reduction approaches.

9.3.1. *Principal component analysis*

As stated in [FOD 02], PCA is "the best, in the mean-square error sense, linear dimension reduction technique". It consists in finding the eigenvectors as well as the eigenvalues of the covariance matrix of X [JOL 02, SMI]. Then the principal components are obtained by a linear combination of the eigenvectors and the observation matrix X.

$$S = BX \qquad (9.8)$$

where B is the eigenvectors matrix of the covariance matrix of X and S are the principal components.

The dimension reduction is performed by choosing those eigenvectors whose corresponding eigenvalues are greater than a fixed threshold. A general and comprehensive description of the method can be found in [JOL 02, SMI].

9.3.2. *Independent component analysis*

Independent component analysis is a signal processing tool which can also be used to perform dimension reduction of multivariate data. While PCA provides results where principal components are uncorrelated, ICA methods are based on the higher order statistics providing that the linear projections are not necessarily orthogonal. The uncorrelatedness condition is replaced with that of independence which is more general ; independence implies uncorrelatedness but vice versa is not always true for non gaussian variables. The ICA model can be expressed by equation (9.7), where X is the observation matrix, W is the mixing matrix and S is the source matrix.

Generally, ICA is performed in two steps : defining the objective function and the algorithm which optimizes it. Different statistical tools can be used to measure the independence of random variables which are leading to various objective functions to be optimized.

Kullback-Leibler divergence

Kullback-Leibler (KL) divergence [KUL 59] allows to estimate the mutual independence of random variables by measuring the distance between their probability density functions (*pdfs*). For two probability densities P and Q of a random variable X, the KL divergence is defined by :

$$D_{KL}(P\|Q) = \sum_i P_i \log \frac{P_i}{Q_i} \qquad (9.9)$$

Statistical independence is obtained by minimizing this criterion.

Mutual information

A common contrast function used to test the statistical independence of random variables is based on the mutual information. It can be derived from the Kullback-Leibler divergence and for two random variables X and Y it is defined as :

$$I(X, Y) = \sum_{y \in Y} \sum_{x \in X} p(x, y) \log \left(\frac{p(x, y)}{p_1(x) . p_2(y)} \right) \qquad (9.10)$$

where $p(x, y)$, $p_1(x)$ and $p_2(y)$ are respectively the joint probability of (x, y), the marginal probability of x and the marginal probability of y.

As well as for the Kullback-Leibler divergence, the statistical independence can be obtained by minimizing this contrast function.

Nongaussianity

Statistical independence can also be achieved under the assumption that nongaussian variables are independent [HYV 00]. We will mention two measures for estimating the nongaussianity of a random variable :

Kurtosis

Kurtosis or the fourth order cumulant is a common measure nongaussianity which derives from the development of the second characteristic function of a random variable. For a random variable X it is defined as the fourth order cumulant divided by the square of the second order cumulant :

$$kurt(X) = \frac{\kappa_4(X)}{\kappa_2^2(X)} \qquad (9.11)$$

In another formulation, the kurtosis is also defined as the forth order moment divided by the square of the variance of the random variable X minus 3. In this way, the kurtosis of a gaussian random variable with zero mean and unit variance equals zero.

$$kurt(X) = \frac{\mu_4}{\sigma^4} - 3 = \frac{E\left\{X^4\right\}}{\left(E\left\{X^2\right\}\right)^2} - 3 \qquad (9.12)$$

The kurtosis of a nongaussian random variable is nonzero. Random variables with a negative kurtosis are called subgaussian while those with a positive kurtosis are called super gaussians.

Negentropy

A second measure of the nongaussianity of a random variable X is the negentropy. It is based on the entropy of a random variable and on the fact that *a gaussian variable has the largest entropy among all random variables of equal variances*, [HYV 00]. In this way, the negentropy of a random variable X is defined by the difference between the entropy of a gaussian variable V_{gauss} (with the same mean and standard deviation) $H(V_{gauss})$ and its own entropy $H(X)$:

$$J(X) = H(V_{gauss}) - H(X) \qquad (9.13)$$

The negentropy of a random variable is always positive and equals zero only if the random variable is gaussian. The statistical independence is obtained by searching nongaussian variables which is performed in fact by maximizing the negentropy. For approximating negentropy a classical approach is to use high order statistics.

9.3.3. *Non-negative matrix factorization approaches*

The non negativity of the results is often required by applications where physical phenomenon are analyzed as well as for applications where records are non-negative data. In multivariate imaging, pixels quantify the amount of light reflected by a surface, which is a non-negative value. Another examples of application demanding non negative constraints is electron energy loss spectroscopy (EELS) [BON 05]. One objective of the dimension reduction methods is to reveal meaningful factors which might have a physical sense. PCA and ICA methods provide decomposition on orthogonal/independent basis for which factor loads can be either positive or negative ; this makes impossible the identification of meaningful factors which might reveal the presence of some particular object or compounds in the image. Non-negative matrix factorization techniques can be an appropriate substitute for the classical methods presented above. Their principle is based on the fact that one factor is either present (and so it has a positive contribution) or absent (and in this case its contribution is zero).

Positive Matrix Factorization

In the context of blind source separation (BSS), the Positive Matrix Factorization (PMF), [PAA 94, PAA 97] is the first method addressing to this problem. Its first aim was to perform a factorial analysis of environmental observations by finding the minimal number of causes (factors) which might explain the observation data set. Let X be the observation matrix, W the factor matrix and H their contributions, then factorial decomposition proposed by Paatero consists in minimizing the following objective function :

$$f(W, H) = \|X - WH\|^2 \qquad (9.14)$$

imposing that $W \geq 0$ and $H \geq 0$. For optimization, the *Alternative Least Square* (ALS) algorithm was proposed.

Non-negative Matrix Factorization

The Non-negative Matrix Factorization algorithm (NMF), [D.D 99] was designed for *a part based decomposition* of the observation matrix X. It consists in finding the matrices $W \geq 0$ and $H \geq 0$ which give the best approximation of X :

$$X \approx WH \tag{9.15}$$

In [LEE 01] there are proposed two formulations of this algorithm :

1) Minimize the Euclidean distance $\|X - WH\|^2$ subject to $W, H \geq 0$ under the following update rules :

$$H_{a\mu}^+ \leftarrow H_{a\mu} \frac{(W^T X)_{a\mu}}{(W^T W H)_{a\mu}} \tag{9.16}$$

$$W_{ia}^+ \leftarrow W_{ia} \frac{(XH^T)_{ia}}{(WHH^T)_{ia}} \tag{9.17}$$

2) Minimize the KL divergence $D(X\|WH)$ subject to $W, H \geq 0$ under the following update rules :

$$H_{a\mu}^+ \leftarrow H_{a\mu} \frac{\sum_i W_{ia} X_{i\mu}/(WH)_{i\mu}}{\sum_k W_{ka}} \tag{9.18}$$

$$W_{ia}^+ \leftarrow W_{ia} \frac{\sum_\mu H_{a\mu} X_{i\mu}/(WH)_{i\mu}}{\sum_\nu H_{a\nu}} \tag{9.19}$$

The non-negative matrix factorization problem can be extended to include additional constraints such as the sparseness of the factor matrix W or on their contributions H, [HOY 02] and [HOY 04].

9.3.4. *A magnitude-correlation representation of multivariate data*

In the following we will consider pixels as spectra and spectra as vectors. The mathematical definition of vectors identity states that two identical vectors have the same magnitude and the same direction ; the assumption that the vector magnitude is independent of direction is also true, meaning that the magnitude and direction offer complementary information about vectors. Moreover, vectors can be uniquely and

completely defined by their magnitude and direction (or shape). When comparing two spectra or two pixels we consider their magnitude and their directions or their shape.

One way to compare vectors magnitude is the Euclidian distance ; it is also the most common used similarity measure for multivariate data clustering. Euclidian distance is often cited in the literature regarding image classification and unsupervised learning and is characterized as *"...the simplest measure of vector distance, but not an accurate one"* [FUK 90]. It is the basic similarity measure used by *Isodata* and *K-means* clustering algorithms.

Shape or correlation information is not commonly used for multivariate data clustering because in most applications of cluster analysis the main dissimilarities between objects depend on object magnitude, not on their shape. But ignoring shape information about the data may induce severe errors in the overall result. The use of shape information can be justified and it is of high importance "*for the situations where all variables have been measured on the same scale and the precise values taken are important only to the extent that they provide information about the object's relative profile*" [B.E 01]. Our strategy is to combine magnitude and shape information into a unique representation of the data.

There are several similarity measures which test the similarity of shape between two vectors. Spectral angle equation (9.3) or the correlation coefficient equation (9.4) are common statistic measures of the strength of the linear relationship between two random variables. Although these measures do not describe completely the shape similarities between two random variables they are commonly used in this purpose by the data analysis community.

The magnitude-shape representation is achieved by expressing the vectors' magnitude and correlation relatively to some reference vectors. Magnitude is expressed by taking as reference the origin and so computing the norm of the data using one of the Minkowski metrics. The linear correlation is expressed by considering as reference a d-dimensional random vector ; a reasonable choice in this purpose could be the first eigenvector eig_1 obtained by PCA. In this way, each data will be described by two values, one expressing its norm computed by one of the Minkowski metrics and one expressing (even though not in a complete manner) its shape, or more precisely its linear correlation to a reference vector.

$$\text{FinalData} = \begin{pmatrix} \mid X_1 \mid_r & \mid X_2 \mid_r & \dots & \mid X_n \mid_r \\ R(X_1, eig_1) & R(X_2, eig_1) & \dots & R(X_n, eig_1) \end{pmatrix} \quad (9.20)$$

Now, data can be represented into a two dimensional feature space, visualized and analyzed by specific methods.

9.4. Clustering methods for multivariate images

Clustering methods can be divided into three main groups : hierarchical, centroid based or partitional and density based methods. Each group can be further subdivided ; the best-known clustering hierarchical methods are hierarchical agglomerative algorithms (*single-link*, *complete-link* or *Ward algorithm*), [SNE 73, KIN 67, MUR 84] while partitional clustering methods are represented by the *K-means* [FOR 65, J.A 79] and its fuzzy extension *fuzzy C-means* [BEZ 81]. The main drawback of these methods is that they are well adapted for convex shape clusters and the fact that the number of clusters do not derive automatically after the clustering process. Density based clustering methods are designed to overcome these drawbacks. We will mention Denclust [Den98], CLUPOT [COO 81], DBSCAN [DBS96] and Mean-Shift [CHE 95] as the best-known representants of this group of methods. A general description of clustering methods for multivariate images can be found in [T.N 05]. Following we will present two density based methods which were compared and used for microscopy image segmentation : *Parzen-Watershed* and *Support Vector Clustering*.

9.4.1. *Parzen-Watershed algorithm*

Parzen-watershed algorithm is a clustering technique which does not make any assumption concerning the shape nor the number of classes [J.C 05], [BON 00], [M.H 01]. For multivariate data, a dimension reduction pre-processing step is mandatory. Following we will summarize this method for the case of two-dimensional data but the method can also be applied for 3 or 4 dimensional data.

The first step of this approach consists in representing the data as points into a two-dimensional feature space. The feature space can be represented as a rectangular grid containing $n \times m$ elements y_i. In this way, any function f can be defined by its value $f(y_i)$ which can be visualized as an image containing $n \times m$ pixels. Then each data point is represented in the image as a pixel having a different grey level from the other pixels, see figure 9.1(b). The second step (see figure 9.1(c)), consists in estimating the probability density function $p(x_i)$ of the whole data set in the new feature space. This can be done according to the Parzen method [PAR 62] for which the point distribution (one object corresponds to one point x_i in feature space) is transformed into a quasi-continuous distribution $p(x_i)$ through a convolution by a smoothing kernel :

$$p(x_i) = \lambda \sum_{k=0}^{N} K\left(\frac{x_i - y_k}{h}\right) \tag{9.21}$$

where K is a smoothing kernel of size h and λ is a normalizing factor. The estimated *pdf* $p(x_i)$ is generally characterized by several modes, or local maxima, separated by valleys or local minima. Each mode is considered to match with a class of objects.

Thus, next step figure 9.1(d), consists in estimating the position of the boundaries between the different classes in the feature space. Although this task is trivial for a one dimensional feature space, it is much complicated for an n dimensional feature space, $(n > 1)$. To split the feature space, the watershed function issued from mathematical morphology [S.B 82] or the SKIZ (SKeleton by Influence Zones) method [SER 82] can be used.

(a)	(b)	(c)	(d)
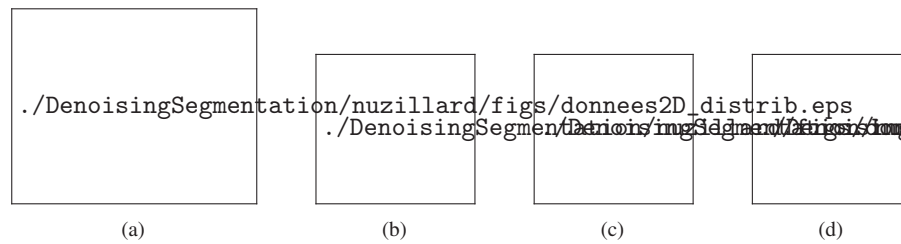
Figure 9.1. *Illustration of the Parzen-Watershed algorithm for two-dimensional data : a) scatterplot of two-dimensional data, b) image representation of two-dimensional data, c) estimated pdf, d) influence zones.*

Then, last step consists in returning to the real image space. This can be done easily because one knows where any pixel x_i was mapped in the feature space, when the scatterplot was built. Thus, for any pixel of the image space, it only needs to carry back the label c found in the feature space, to the real space. As for any clustering method, the question of selecting a number of classes is relevant. For this method, the number of classes is considered to be equal to the number of modes of the estimated *pdf*. The number of modes itself is related to the width of the smoothing kernel h. Estimating the number of classes consists in looking at the number of modes M as a function of the parameter h. The decreasing curve $M = f(h)$ displays some plateaus, which gives stable solutions for the number of classes.

9.4.2. *Support Vector Clustering*

For this method, data points are mapped from the feature space to a higher dimensional space using a Gaussian kernel. In this new feature space, we look for the hyperplane which separates the data set from the origin with the largest margin. This hyperplane when mapped back to the data space forms a set of contours which enclose the data points. These contours are interpreted as cluster boundaries. Points enclosed by each separated contour are gathered in the same cluster. As the width parameter of the Gaussian kernel decreases, the number of connected contours in the data space increases, leading to an increasing number of clusters [A.B 01]. SVC can deal with outliers by employing a soft constant margin that allows the hyperplane in feature

space not to separate all points. For large values of this parameter, one can deal with overlapping clusters [A.B 01]. To separate the data from the origin, one solves the following quadratic equation [B.S 01] :

$$min_{w \epsilon F, \xi \epsilon R^l \rho \epsilon R} \quad \left(\frac{1}{2} \parallel w \parallel^2 + \frac{1}{\nu l} \sum_i \xi_i - \rho \right) \tag{9.22}$$

subject to :

$$(w.\Phi(x_i)) \geq \rho - \xi_i, \quad \xi_i \geq 0 \tag{9.23}$$

Since nonzero slack variables ξ_i, are penalized in the objective function, we can expect that if w and ρ solve this problem, then the decision function :

$$f(x) = sgn\left((w.\Phi(x_i)) - \rho\right) \tag{9.24}$$

will be positive for most samples x_i contained in the data set, while $\parallel w \parallel$ will be still small. The actual trade-off between these two goals is controlled by ν . Using multipliers $\alpha_i, \beta_i \geq 0$ we introduce a Lagrangian :

$$L(w, \xi, \rho, \alpha, \beta) = \frac{1}{2} \parallel w \parallel^2 + \frac{1}{\nu l} \sum_i \xi_i - \rho \\ - \sum_i \alpha_i((w.\Phi(x_i)) - \rho + \xi_i) - \sum_i \beta_i \xi_i \tag{9.25}$$

and set the derivatives with respect to the primal variables w, ξ, ρ, equal to zero, yielding :

$$w = \sum_i \alpha_i \Phi(x_i) \tag{9.26}$$

$$\alpha_i = \frac{1}{\nu l} - \beta_i \leq \frac{1}{\nu l}, \ \sum_i \alpha_i = 1 \tag{9.27}$$

All samples x_i for which $\alpha_i > 0$ are called support vectors. So the decision function f is transformed into a kernel expansion.

$$f(x) = sgn\left(\sum_i \alpha_i k(x_i, x) - \rho \right) \tag{9.28}$$

Substituting equations (9.26) and (9.27) into (9.25) we obtain the dual problem :

$$min \quad \left(\frac{1}{2} \sum_{ij} \alpha_i \alpha_j k(x_i, x_j) \right) \tag{9.29}$$

subject to :

$$0 \leq \alpha_i \leq \frac{1}{\nu_l}, \ \sum_i \alpha_i = 1 \tag{9.30}$$

We can recover ρ by exploiting that for any $\alpha_i > 0$ the corresponding pattern x_i satisfies :

$$\rho = (w.\Phi(x_i)) = \Sigma_i \alpha_i k(x_j, x_i) \tag{9.31}$$

Although this procedure can recover exactly the boundaries of the classes, in our application we used the same SKIZ procedure [SER 82] in order to separate the data space. The binary function corresponds to the connected components in the data space which will be used as seeds in the SKIZ procedure, [NUZ 06]. This method is illustrated in figure 9.2

(a)   (b)   (c)   (d)
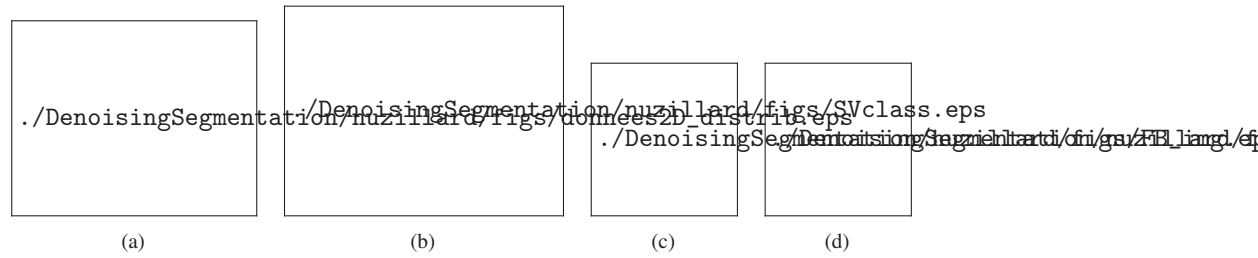
Figure 9.2. *Illustration of the SVC algorithm for two-dimensional data : a) scatterplot of two-dimensional data, b) classes boundaries, c) connected components represented as a binary function in an image space, d) influence zones.*

For this method, the number of classes is considered to be equal to the number of the connected components which is also related to the kernel width as well as to the hyperplane separation constant ν. In [NUZ 06] a comparison between Parzen-Watershed method and SVC is performed on artificial bi-dimensional data containing showing that Parzen-Watershed performs better than SVC when we deal with overlapped classes.

Density based clustering - drawbacks

Even though density based clustering techniques perform better than other clustering methods when dealing with arbitrary shape clusters, it is reasonable to mention their main inconvenient. First of all, these methods may consider low-density clusters as noise or outliers. Second, the computing time for estimating the probability density function is very demanding. Estimating the right parameters for *pdf* estimation is very problematic too. They also have problems with overlapped clusters. Although some scientists [T.N 05] are skeptical regarding the use of these clustering techniques for multivariate images, they proved to be a strong tool for these applications especially when dimension reduction is performed before clustering.

9.5. Application on multispectral image segmentation

9.5.1. *Data description*

A multispectral image of a cross section of a barley grain, acquired in microspectrofluorometry is analyzed in order to identify external tissues of the barley grain. The central part contains starch and the external layers serve as protection. Those contain pure natural fluorescent components : cutin, ferulic acid, lignin whose signatures recover partially. Confocal laser microspectrofluorometry enables to visualize autofluorescence of tissues. Many laser excitations associated with a set of various filters in reception allow getting pseudo multi-spectral images. The grains were furnished by INRA Clermont-Ferrand and the images were recorded by INRA Nantes thanks to M.-F. Devaux. The data set contains 19 images each one of 512x512 pixels. For the multispectral image, presented in figure 9.3 the tissues are revealed by image segmentation. A probability based clustering method (the Parzen-Watershed algorithm) is used to perform image segmentation. PCA, NMF algorithm as well as *the magnitude-shape representation* are used for dimension reduction.
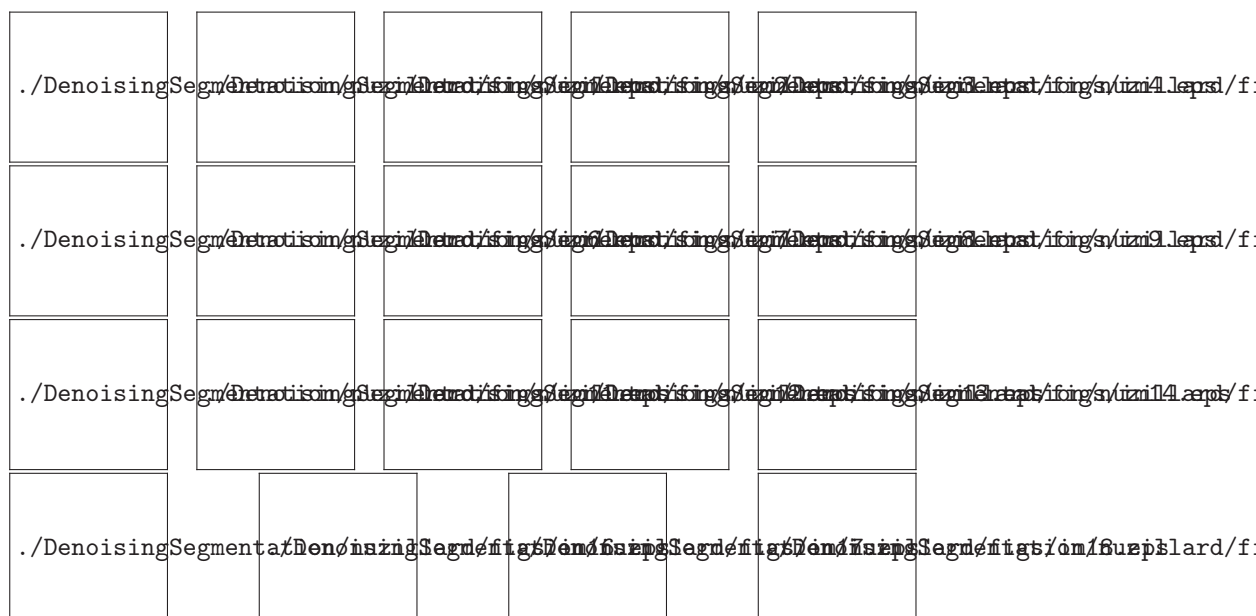
Figure 9.3. *Multicomponent image representing a cross section into a barley grain*

9.5.2. *Clustering results*

Three dimension reduction methods were tested in our experiments as pre-treatment step for pixel clustering : linear projection methods (*PCA* and *NNMF*) and *the magnitude-shape representation*. Linear projection methods realize some linear transformations of the data onto orthogonal or independent directions. Generally, the most part of the energy is gathered in two or three components. This allows us to perform the dimension reduction, by ignoring the low energy components. The energy distribution for the PCA and NMF applied on the multispectral image under study is presented in table 9.1. The four first resulted components after linear transformation are presented in figure 9.5.

PC	77.4%	16.8%	2.5%	1.3%
NNF	42%	41%	12%	5%

Tableau 9.1. *Energy distribution*

Figure 9.4. *The four first principal components.*

One advantage of linear projection methods such NMF is that besides the implicit dimension reduction they perform a more realistic transformation of data. By taking into account positivity constraints required by the physical laws, the results become physically interpretable. Lignin and ferulic acid spectra can be easily identified with the first two column vectors in the recovered matrix, see figure 9.6, meaning that the second line in figure 9.5 corresponds to a concentration map of these two fluorescent components.
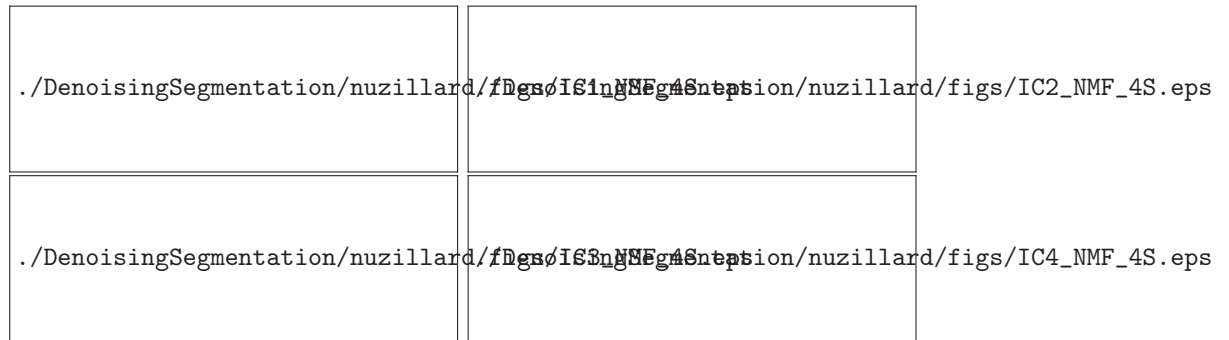
Figure 9.5. *Results of the dimension reduction by NMF algorithm.*

(a)

(b)
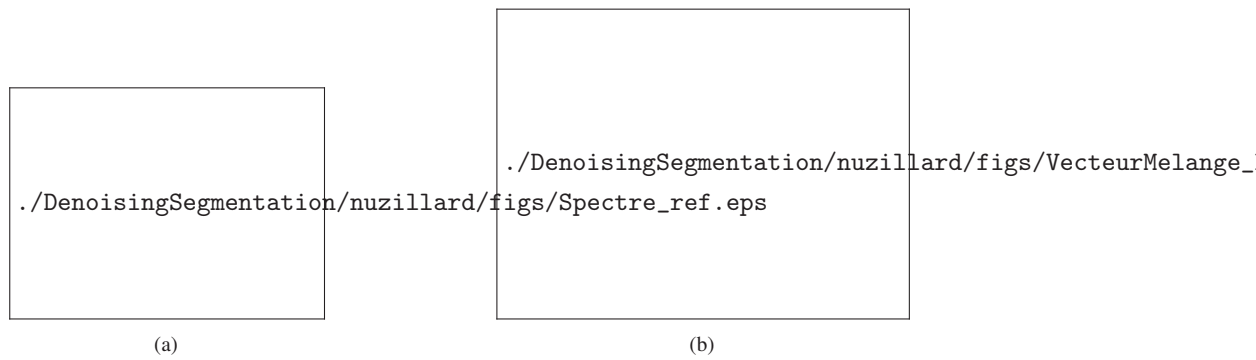
Figure 9.6. *(a) Reference spectra : cutin, ferulic acid and lignin. (b) The four first column vectors resulted after the NMF*

Parzen-watershed clustering results are presented in figures 9.8, 9.9, 9.10. The stability index for both magnitude-shape representation and NMF show 5 and 4 classes with significant reliability, figure 9.7 b) and c) while for PCA, figure 9.7 a) indicates a partitioning in four or three classes. The results of the pixel classification are almost identical, as we can see in figures 9.8, 9.9 and 9.10.

(a)   (b)   (c)
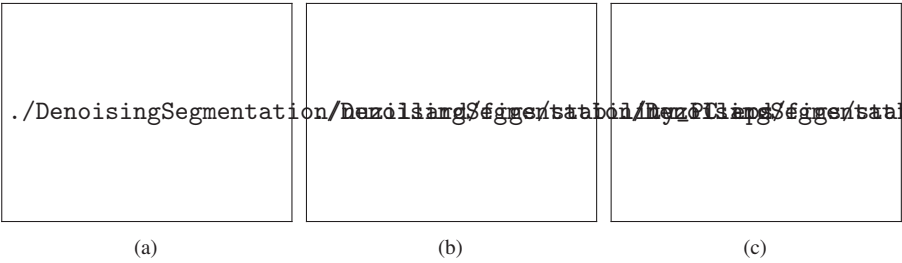
Figure 9.7. *Stability of the number of classes : a) PCA representation, b) NMF representation, c) magnitude-shape representation*

(a)   (b)   (c)
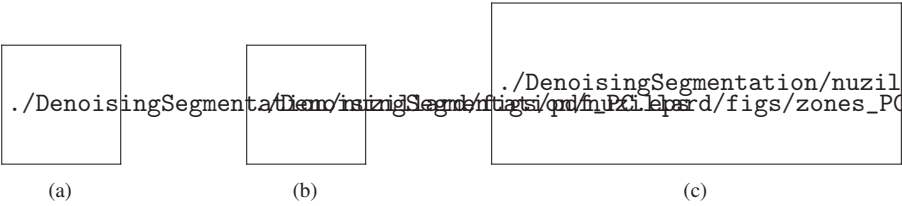
Figure 9.8. *Parzen-watershed clustering results : PCA was used to perform dimension reduction. a) estimated pdf, b) influence zones, c) segmented image*

(a)   (b)   (c)
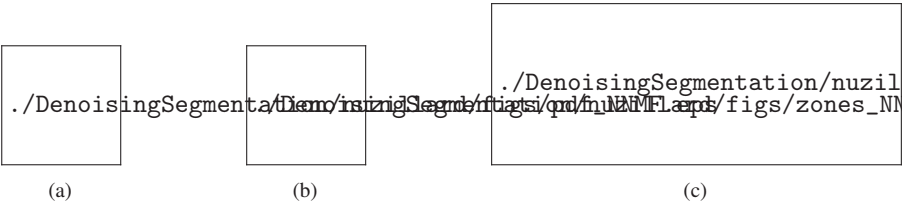
Figure 9.9. *Parzen-watershed clustering results : NMF was used to perform dimension reduction. a) estimated pdf, b) influence zones, c) segmented image*

./DenoisingSegmentation/nuzillard/figs/img_seg_MS.eps

./DenoisingSegmentation/nuzillard/figs/ond_fn_MS.eps ./DenoisingSegmentation/nuzillard/figs/zones_MS.eps

(a)   (b)   (c)
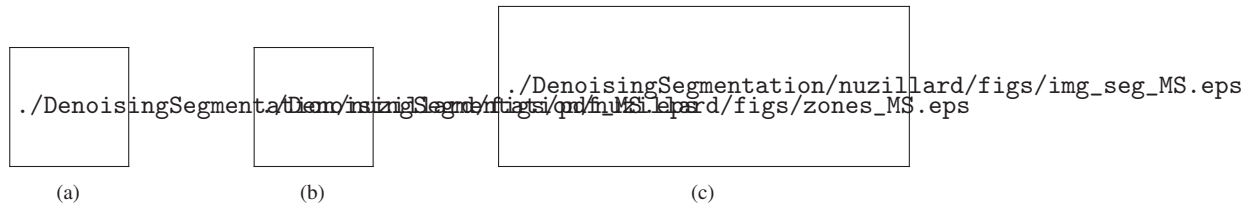
Figure 9.10. *Parzen-watershed clustering results : magnitude-shape representation was used to perform dimension reduction. a) estimated pdf, b) influence zones, c) segmented image*

Even though clustering validation can be realized by internal validity indices which take into account the intrinsic relationships between data, it is more realistic to use external indices (such as *truth maps* representing a priori information about the data) for class validation when these are available. Figure 9.11 represents a theoretical cross-section of a barley grain which can be used to validate clustering results. Better segmentation results are obtained when the dimension reduction was performed by NMF and *magnitude-shape representation* since the husk (where the most part of the cutin is presented) is better revealed.

./DenoisingSegmentation/nuzillard/figs/barley_grain.eps

Figure 9.11. *Theoretical section of a barley grain*

9.6. Summary

Clustering is an appropriate tool for multivariate image segmentation. However the clustering algorithm should be carefully chosen. Centroid based algorithms are well adapted for convex shape clusters ; for these methods as well as for the hierarchical algorithms, the choice of the similarity measure is often application dependent. Spectral angle and Pearson's correlation coefficient offer better results than the Euclidean distance when spectral considerations are important or when we dispose of some pixels reference spectra. Non Euclidean metrics can also be used instead to improve clustering performances, but their utility remains application dependent. In the absence of a priori information, the choice of the appropriate similarity measure can only be done by comparing the results obtained for different similarity criteria but this will increase the computing time. The main advantage of centroid based algorithms is the computing time, which is a very important aspect for applications dealing with large amount of data such as multivariate images. However, these methods present some important drawbacks (the number of classes imposed a priori and convex shape clusters) which may result in erronnated clustering results. Probability based clustering methods overcome these drawbacks and they are an efficient tool for multivariate image segmentation.

We outline the importance of the dimension reduction as pre-processing step for pixel clustering. Linear projection methods are often used in this purpose. Among these, methods which take into account the non negativity constraint are especially appropriate for multivariate images since the hidden variables revealed have a physical meaning. Similarity measures can also be used to extract relevant features about the data. In this case, a reference has to be chosen and all data should be expressed by their similarity relative to this reference. In *magnitude-shape representation* data are expressed by their norm and by their spectral similarities relative to the first eigenvector obtained by PCA. This gives a more general description of the data.

A sensitive problem regarding clustering methods is the validation of the results. Most of the validity indices are appropriate for particular shaped clusters (gaussian) and they are sensitive also to the size of the classes. This may have as result a poor pixel partitioning. External indices should be used instead when they are available.

[A.B 01] A. BEN-HUR D. HORN H. S.-V. V., « Support vector Clustering », *Journals of Machine Learning Research*, vol. 2, p. 125–137, 2001.

[ALE 99] ALEXANDER HINNEBURG CHARU C. AGGARWAL D. A. K., « What Is the Nearest Neighbor in High Dimensional Spaces ? », *Lecture Notes in Computer Science*, vol. 1540, p. 217-235, 1999.

[B.E 01] B. EVERITT S. LANDAU M. L., *Cluster analysis*, Oxford University Press, London, 4th édition, 2001.

[BEZ 81] BEZDEK J. C., *Pattern recognition with fuzzy objective function algorithms*, Plenum Press, New York, 1981.

[BON 00] BONNET N., « Artificial intelligence and pattern recognization techniques in microscopic image processing and analysis », *Adv. Imag. Electron. Phys.*, vol. 14, p. 1-77, 2000.

[BON 05] BONNET N., NUZILLARD D., « Independent component analysis : A new possibility for analyzing series of electron energy loss spectra », *Elsevier, Ultramicroscopy*, , n°102, p. 327-337, 2005.

[B.S 01] B. SCHÖLKOPF J-C. PLAT J. S.-T. A. S. R. W., « Estimating the support of a high dimensional distribution », *Neural Computing*, vol. 13, p. 1443-1471, 2001.

[C.C 01] C.C.AGGARWAL A. HINNEBURG D. K., « On the surprising behavior of distance metrics in high dimensional space », *Lecture Notes in Computer Science*, vol. 1973, p. 420-434, 2001.

[CHE 95] CHENG Y., « Mean Shift, Mode Seeking, and Clustering », *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, p. 790-799, 1995.

[COO 81] COOMANS D., MASSART D., « Potential methods in pattern recognition. Part 2 CLUPOT-an unsupervised pattern recognition technique », *Anal. Chim. Acta*, vol. 133, p. 225-239, 1981.

[DBS96] *A density-based algorithm for discovering clusters in large spatial databases with noise*, 1996.

[D.D 99] D.D.LEE, H.S.SEUNG, « Learning the parts of objects by non-negative matrix factorization », *Nature*, , n°401, p. 788-791, 1999.

[Den98] *An efficient approach to clustering in large multimedia databases with noise*, 1998.

[D.F 07] D. FRANCOIS V. W., VERLEYSEN M., « The Concentration of Fractional Distances », *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, n°7, p. 873-886, 2007.

[D.W 83] D. W. SCOTT J. R. T., « Probability density estimation in higher dimensions », *J.E. Gentle (ed.), Computer Science and Statistics : Proceedings of the Fifteenth Symposium on the Interface*, Amsterdam, New York, Oxford, North Holland-Elsevier Science Publishers, p. 173-179, 1983.

[F.A 93] F. A. KRUSE A. B. LEFKOFF J. B. B. K. B. H. A. T. S. P. J. B. A. F. H. G., « The Spectral Image Processing System (SIPS) - Interactive Visualization and Analysis of Imaging Spectrometer Data : Remote Sensing of Environment », *Special issue on AVIRIS*, vol. 44, p. 145 - 163, mai - juin 1993.

[FOD 02] FODOR I. K., A Survey of Dimension Reduction Techniques,   technical report, Lawrence Livermore National Laboratory, 2002.

[FOR 65] FORGY E. W., « Cluster analysis of multivariate data : efficiency versus interpretability of classification », *Biometrics Abstract, Biometric Society meeting*, vol. 21, Riverside, page768, 1965.

[FUK 90] FUKUNAGA K., *Introduction to statistical pattern recognition*, Academic Press, Inc., San Diego, 2nd édition, 1990.

[GOW 86] GOWER J. C., LEGENDRE P., « Metric and Euclidean properties of dissimilarity coefficients », *Journal of Classification*, vol. 3, p. 5-48, 1986.

[HOY 02] HOYER P. O., « Non-negative sparse coding », *In Proc. IEEE Workshop on Neural Networks for Signal Processing*, Martigny, Switzerland, p. 557-565, 2002.

[HOY 04] HOYER P. O., « Non-negative Matrix Factorization with sparseness constraints », *Journal of Machine Learning Research*, vol. 5, p. 1457-1469, 2004.

[HYV 00] HYVÄRINEN A., OJA E., « Independent Component Analysis : Algorithms and Applications », *Neural Networks*, vol. 13, n°4-5, p. 411-430, 2000.

[J.A 79] J. A. HARTIGAN M. A. W., « A K-means clustering algorithm », *Applied Statistics*, vol. 28, p. 100-108, 1979.

[J.C 05] J. CUTRONA N. BONNET M. H. F. H., « Advances in the segmentation of the multi-component microanalytical images », *Ultramicroscopy*, vol. 103, p. 141-152, 2005.

[JOL 02] JOLLIFFE I., *Principal Component Analysis*, Springer Series in Statistics, 2002.

[K.B 00] K. BEYER J. GOLDSTEIN R. R., SHAFT U., « When Is Nearest Neighbor Meaningful ? », *VLDB 2000, Proceedings of 26th International Conference on Very Large Data Bases*, Cairo Egypt, p. 10-14, 10-14 septembre 2000.

[KIN 67] KING B., « Step-wise clustering procedures », *J. Am. Stat. Assoc.*, vol. 69, p. 86-101, 1967.

[KUL 59] KULLBACK S., *Information theory and statistics*, John Wiley, New York, 1959.

[LEE 01] LEE D. D., SEUNG H. S., « Algorithms for non-negative matrix factorization », *In Advances in Neural Information Processing, (NIPS*2000)*, MIT Press,, page556562, 2001.

[MAH 36] MAHALANOBIS P., « Proceedings of the National Institute of Science of India », vol. 12, India, p. 49-55, décembre 1936.

[M.H 01] M. HERBIN N. BONNET P. V., « Number of clusterings and influence zones », *Pattern Recognition Letters*, vol. 22, p. 1557–1568, 2001.

[MUR 84] MURTAGH F., « A survey of recent advances in hierarchical clustering algorithms which use cluster centers. », *Computer Journal*, vol. 26, p. 354-359, 1984.

[NUZ 06] NUZILLARD D., LAZAR C., « Comparison of Two Unsupervised Methods of Classification for Segmenting Multi-spectral Images », *International Conference on Acoustics, Speech, and Signal Processing, ICASSP*, vol. 3, Toulouse, France, 2006.

[O.C 00] O. CARVALHO P. M., « Spectral Correlation Mapper (SCM) : An Improving Spectral Angle Mapper », *Ninth JPL Airborne Earth Science Workshop*, Pasadena CA, JPL Publication 00-18, p. 65-74, 2000.

[PAA 94] PAATERO P., TAPPER U., « Positive matrix factorization : A non-negative factor model with optimal utilization of error estimates of data values », *Environmetrics*, vol. 5, p. 111-126, 1994.

[PAA 97] PAATERO P., « Least squares formulation of robust non-negative factor analysis », *Chemometrics and Intelligent Laboratory Systems*, vol. 37, p. 23-35, 1997.

[PAR 62] PARZEN E., « On the estimation of a probabitity density function and mode », *Annals Math. Stats.*, vol. 33, p. 1065–1076, 1962.

[S.B 82] S. BEUCHER F. M., *Mathematical Morphology in Image Processing*, Dekker, New York, 1982.

[SER 82] SERRA J., *Image Analysis and Mathematical Morphology*, Academic Press New York, New York, 1982.

[SMI] SMITH L. I., A tutorial on Principal Components Analysis, Tutorial.

[SNE 73] SNEATH P. H., SOKAL R. R., *Numerical Taxonomy*, Freeman, London, UK, 1973.

[T.N 05] T. N. TRAN R. WEHRENS L. M. B., « Clustering multispectral images : a tutorial », *Chemometrics and Intelligent Laboratory Systems*, vol. 77, n°1-2, p. 3-17, 2005.

Chapitre 10

Blind determination of noise type for spaceborne and airborne remote sensing

10.1. Introduction

Spaceborne and airborne remote sensing systems are efficient tools for providing users and customers operative valuable information for different applications as hydrology, meteorology, environment monitoring, etc. The majority of modern remote sensing systems are multichannel (multispectral), i.e. they either include separate subsystems with different characteristics (wavelengths, polarizations, etc.) in their stuff or simply form images in several or even many subbands [KAA 07].

However, this also gives rise to many serious problems, as it requires to correct the sensors response to avoid the effects of systematic disturbances, such as patterns and random variability or temporal noise, that all together seriously reduce the quality of available data. It is well known that patterns can be often removed in the frequency domain without appreciable loss of spatial or spectral resolution. Temporal noise cannot. As a result, noise is one of the basic factors that degrade Multichannel Remotely Sensed data and prevent accurate and reliable retrieval of useful information [AIA 06]. Availability of a priori information on noise type and properties allows applying more effective methods for image processing, namely, filtering, edge detection, feature extraction [BAR 03, KUL 04].

Indeed, noise determines the accuracy with which absorption features can be distinguished in the spectra and later how objects can be detected on the ground. Note that the amplitude of noise may be quite large, up to $20 - 30\%$ of the unperturbed signal

Chapitre rédigé par B. Vozel, S.K. Abramov, K. Chehdi, V.V. Lukin, M.L. Uss, N.N. Ponomarenko, J. Astola.

[**?**, **?**].

Concerning the hyperspectral feature of data, the signal is usually land cover dependent and varies with both wavelength and atmospheric absorption. Consequently, information on noise is required with each spectrum to determine if detection of a particular absorption feature is possible, the more so as noise level decreases with both an increase in wavelength and atmospheric absorption. The reflectance components are generally spatially autocorrelated and also correlated with one another. Their homogeneity may vary with the wavelength. Since a region may be homogeneous in one band, but not in another, the noise identification, and subsequently the estimation of its statistical parameters are clearly to be carried out on separate bands, provided the noise can be reasonably assumed spatially stationary within each spectral band and despite it may also be spatially and spectrally correlated.

In practice, it is usual that some a priori information is available but not all desirable properties of noise and other distortions are at disposal. Typically, some preliminary assumptions can be implied depending a type of imaging system that has obtained a given image. For example, for optical images it is usually assumed that dominant noise is additive and Gaussian [**?**]. It is often assumed that this noise is i.i.d. (spatially uncorrelated). If this noise is not intensive, in some conditions blur influence can be more prevailing than the influence of the additive noise. Then one has to deal with image restoration although filtering the additive noise is paid attention too. Impulse noise can be present as well [BAR 03] and its probability can be unknown. Blur parameters can be unknown as well [AIS 08]. Thus, there are, at least, three types of distortions that can degrade an optical (sub-band, component) image and often their parameters are unknown. Moreover, recently it has been shown that additive noise in some optical systems can be not stationary (strictly saying, it can be signal-dependent [FOI 07]) and assumption that noise is spatially uncorrelated can be not valid as well [LUK 08a]. Analogous effects are observed in multi- and hyperspectral remote sensing where noise properties are not yet understood well enough [BAR 05] but clearly they change in wide limits in different sub-band images.

Thus it becomes necessary or, at least, desirable to retrieve such information either in imaging system itself before image transmission (storage, pre-processing) or before offering it to customers or at user's side to perform effective processing by applying automatic, blind and robust processing of images.

Here automatic means that images are to be processed without participation of a human, in non-interactive mode. Blind means that data processing is to be performed under conditions of limited a priori information on noise/impairment type and characteristics or a range of variation of these characteristics can be rather wide. Robust in wide sense means that image processing methods should perform well enough if a priori information, preliminary estimates and decisions are not very reliable and/or accurate.

Because of this, in parallel with attempts to improve robustness [BAR 03, GON 02, AYS 07], a lot of attention is nowadays paid to other direction, namely, design of multistage automatic or semiautomatic procedures for which noise type and statistical

characteristics are estimated and evaluated at the first stage and some methods that incorporate the obtained knowledge are applied at the second stage [FOI 07].

Similar situation holds for multichannel radar remote sensing. Although a simplified assumption that dominant noise is multiplicative, in general, valid [TSY 05], impulse and additive noise components can be present as well [MEL 00, LUK 08b]. Whilst in many papers multiplicative noise is supposed spatially uncorrelated [TSY 05, MEL 01], it is more adequate [LUK 08b, LOP 07] to consider speckle spatially correlated. Besides, multiplicative noise might have either almost Gaussian or obviously non-Gaussian PDF (Probability Distribution Function) depending upon a radar type (side-look aperture or synthetic aperture) and a number of looks in an SAR (Synthetic Aperture Radar) [MEL 00, OKT 07]. Thus, again we deal with several factors that can degrade a radar image and several properties (statistical characteristics) of noise can be unknown in advance.

Ideally one should know in advance noise type(s) and all characteristics of them : PDF (or, at least, variance) of additive noise and its spatial correlation ; PDF (or, in the worst case, relative variance) of multiplicative noise and its spatial correlation ; PDF or, at least, probability of occurrence for impulse noise ; point spread (blur) function, etc. A simpler and more realistic task that can be solved in a blind and automatic manner is to determine dominant factor(s) that degrade an image at hand, to estimate basic statistical characteristics of the determined dominant factors and to adjust the corresponding parameters of image processing methods and algorithms to be applied. Just this task is mainly addressed in this Chapter.

Our contribution deals with revisiting the obtained results, giving a reader a general understanding of what can be done [LUK 05a, LUK 06b, VOZ 06a, LUK 06a, PON 06a], how blind processing can be carried out, what accuracy can be provided [LUK 07c, LUK 08a, LUK 08b, USS 07, LUK 09], and how the obtained decisions and estimates can be used at final stages of image processing [LUK 02c, LUK 02b, KLA 05, VOZ 06a]. Our main intention is processing of multichannel images (multi- and hyperspectral, multi-frequency and multi polarization radar). For such kinds of images, noise type(s) and statistical characteristics can be considerably different in component (sub-band) images [KUL 04, MEL 00, PON 07, BAR 05, O. 05]. Because of this, we insist that noise /distortion type determination is to be performed component-wise. Similarly, blind evaluation of noise/distortion parameters should be done component-wise as well. Data processing at both aforementioned stages can be carried out sequentially (component by component) or in parallel. At the final stage of multichannel image processing both component-wise or vector-like (3-D) methods [KAA 07, BAR 03, KUL 04, PON 07, O. 05, RAD 06] can be realized, the choice depends upon many factors.

The Chapter structure corresponds to two stages of blind automatic procedure considered above. Section 10.2 deals with blind determination of noise/distortion type. Blind estimation of noise statistical characteristics is considered in section 10.3. Information

about possible approaches, their advantages, drawbacks, accuracy and application restrictions is given there. Finally, the conclusions are given and directions of future work are presented.

10.2. Blind determination of noise type

Knowledge of the ratio of signal to noise in hyperspectral data is one of the main parameters for any user to first optimize both data quality investigation and final theoretical interpretation of the data content, and then to devise practical and useful tools for the end-user [BAR 05].

As noise (mainly its temporal part) represents really the limitation to image signal-to-noise ratio (SNR), it is still a valid challenge to dispose of reliable algorithms with which to identify the nature of noise, to estimate its amplitude, and finally remove it. In the following, some solutions will be recalled and some others will be devised for identifying the noise typically introduced by multispectral scanners and by hyperspectral imaging spectrometers. The major part of noise in such sensors are generally assumed to be additive to the signal. Its level decreases with both an increase in wavelength and atmospheric absorption. However, before stating the final observation model, it is worth analyzing the image at hand to be processed and thus accurately estimating noise statistical and spatial correlation characteristics from it.

10.2.1. *General approach*

We focus our attention on the three main and usual classes of degradations, corresponding each to a possible hypothesis that will be naturally translated by the observation model retained for subsequent image processing tasks : the first one corresponding to an additive noise hypothesis, the second one to a multiplicative noise hypothesis, and the latter to an impulse noise hypothesis.

Now, the question is to decide which model among the three previously considered is most likely to have occured during the image forming stage or is simply predominant. A common general idea of blind noise nature identification is the following. An observed image is divided into many overlapping or non-overlapping blocks of small size (typically 5×5 or 7×7 pixels). Then, for each k-th block the local mean $\bar{I}_k$ and the local variance σ_k^2 are calculated. All couple of estimates $(\sigma_k^2, \bar{I}_k)$ is represented by a single dot on a scatter plot. As a result, the scatterplot relative to the whole band may be regarded as the joint probability density function (PDF) of estimated local variance and estimated local mean. Textural components give high local variance values, whereas homogeneous areas produce clusters of scatterpoints. In [**?**], we have shown it is possible to effectively identify the nature of the noise by recording variations of such local statistics but computed only in the homogeneous regions of the observed image. Because the variance of the observed signal measured on homogeneous areas will be approximately equal to the variance of the noise. Then, a regression line can be easily

fitted to the corresponding scatterpoints.

If the regression line has zero slope, then the noise is decided to be an additive one. The intercept of the regression line yields a first estimate of its variance. If the recording can be approached by a line passing through the origin (the intercept is zero), then the noise is decided to be a multiplicative one. Then, its estimated variance can be obtained by the slope of the line. Note that the previous analysis can be directly extended to decide which kind of noise is prevailing in case of more complex models of noise such as mixed multiplicative and additive observation noise. And finally, if the recording can not be viewed as a line passing through zero, then the noise is declared as an impulsive one.

One interesting feature of the scatterplot based methods is that they are known to be also able to correctly measure correlated noise, with few and appropriate adjustments. However, in all cases, one serious difficulty with them is that they ideally require a priori knowledge of the location of homogeneous regions.

As a result, it is possible that among the selected regions, assumed homogeneous, one can have fraction of or totality of areas that correspond to non-intensive texture or to low contrast edges and details, giving biased local estimations, and then erroneous identification results. Different approaches essentially based on local statistics calculation can be proposed in order to overcome these troubles [SEN 02].

Another direction is to make use of refined classification and estimation procedures [VOZ 06a], both to increase the identification rate and also to improve the estimation of statistical noise parameters. We first consider the detection of homogeneous regions by proposing an improvement of classical image multi-thresholding methods. The goal is to achieve the precise determination of homogeneous zones in numerical images by pixels classification. The thresholds and the modes are obtained by minimization of a new energy of gravitational clustering initialized with the significant peaks of a cumulated histogram. Then, the best modes and the best thresholds are calculated by alternate optimization of an energy of multi-thresholding, leading to a piece-wise quadratic potential. This energy is built from a total uniformity criterion which measures the homogeneity of a given map of regions.

Finally, an unsupervised classification is performed by use of a supervised variational classification approach which minimizes an adapted energy of transitions of phases. The potential which controls the classification process is built from the previously determined best thresholds and modes. The whole procedure is described below.

10.2.2. *Classification*

Now, our purpose is to obtain the most precise detection of homogeneous regions in each spectral band. We choose to use a partitioning method, based on histogram transformation, because the histogram operator basically reveals the unknown mixture between the informational content of homogeneous regions (quasi-uniform grey level distribution) and noise.

This refined advanced classification method was recently proposed by some of the

co-authors, leading to noticeable improvement of classical image multi-thresholding methods [KLA 05].

The histogram transformation that we propose to use, considers an alternative modelling of the histogram by a dynamic representation. The values of the histogram are comparable with the masses h_i (occurence frequency) and the samples with the positions i (intensity level) of a dynamic system of the previously defined masses. The initial vector of the positions $\vec{x}^0$ is defined by $x_i^0 = i$, whenever the level i is initially present in the processed image. We call *centre of gravity* relative to a position, the barycentre of the positions located within a radius lower than a maximum preset distance and named *gravitational radius* $R(x_i)$. The energy of gravitational clustering measures the dispersion of the positions with regard to the relative centres of gravity. It is defined by :

$$Q_{\mathrm{gc}}(\vec{h}, \vec{x}) \triangleq \frac{1}{2} \sum_{i \in I} \sum_{j \in I} h_i h_j \delta_{i,j}(\vec{x}) |x_i - x_j|^2 \qquad (10.1)$$

where I is the set of real intensities, $\delta_{i,j}(\vec{x}) \triangleq H(R(x_i, x_j) - |x_i - x_j|)$ and $R(x_i, x_j) = \sup(R(x_i), R(x_j))$. $H()$ is the Heaviside step function.

In practice, a gravitational equation is built to cancel the derivative of the energy of gravitational clustering : $\vec{x}$ minimizes $Q_{\mathrm{gc}}(\vec{h}, \vec{x})$.

The condition $R(x_i, x_j) = \sup(R(x_i), R(x_j)) > |x_i - x_j|$ simply translates the determination of the relative centres of gravity. The trajectory is simply controlled by the proximity and the importance of specific masses. The positions are attracted by their relative centres of gravity. $R(x_i)$ is essential to precisely quantify the multi-modality with regard to the positions of the modes.

Thus, the gravitational clustering returns a vector of the positions gradually constant on each class. Visually, the vector of the positions takes a staircase shape. The positions are converted into integer values to preserve a meaning in the set of intensities. The histogram corresponding to $\vec{x}^\infty$ is denoted histogram of the centres of gravity $h^{\mathrm{gc}}(\cdot)$. It is multimodal and the modes $h^{\mathrm{gc}}(x_i^\infty) = \mu_{x_i^\infty}$ are given by the various levels of the vector $\vec{x}^\infty$.

The thresholds $\vec{s}$, and the modes $\vec{m}$ must be ordered so that each mode lies between two consecutive thresholds. Following previous stage, a first set of thresholds and modes are obtained. Now, the aim is to determine the best thresholds $\vec{s}$ and the best modes $\vec{m}$ with regard to a criterion, denoted *total uniformity criterion*, translating better search for homogeneous regions and measuring the quality of a map of homogeneous regions. Its principle is built on the idea that an area is *uniform* if the dispersion of the grey levels is weak. We can measure the uniformity of an area by estimating the *intra-area variance*. The *total uniformity* criterion of a map is estimated by the weighted average, by the size of the different areas, of the intra-area variances.

Then, the *energy of multi-thresholding* is built in order to coincide exactly with the *total uniformity* criterion if the modes are equal to the intra-area means. It measures the quality of the modes m_k and the thresholds s_k, knowing their number N. This quality is estimated by regarding the vector of the thresholds as an approximation in a finite

number of samples of the histogram. Then, the best modes $\vec{m}^\infty$ and the best thresholds $\vec{s}^\infty$ are calculated by alternate optimization of this energy of multi-thresholding. The optimal solution consists of a continuation which alternatively minimizes the multi-thresholding energy with regard to the modes and the thresholds.

$$Q_{\mathrm{mt}}(\vec{h}, \vec{s}, \vec{m}) \triangleq \frac{1}{2} \sum_{k=0}^{N} \sum_{\substack{x_j > s_k}}^{x_j < s_{k+1}} h_j |x_j - m_k|^2$$
$$+ \frac{1}{2} \sum_{k=1}^{N-1} \sum_{x_j = s_k} h_j \min\left(|x_j - m_k|^2, |x_j - m_{k+1}|^2\right) \tag{10.2}$$

Assuming the modes $\vec{m}^n$ known, the energy of multi-thresholding $Q_{\mathrm{mt}}(\vec{h}, \vec{s}, \vec{m}^n)$ is convex with respect to the thresholds $\vec{s}$. The single minimum $\vec{s}^n$ is simply given by the middle of the interval ranging between two modes : $s_k^n = \left(m_k^n + m_{k+1}^n\right)/2$. Assuming the thresholds $\vec{s}^n$ known, the vector of the modes which minimizes the multi-thresholding energy is given by the barycentres M_k of the thresholds whose weights depend on the thresholds and the histogram : $m_k^{n+1} = M_k(\vec{h}, \vec{s}^n)$.

When the image is seriously degraded, it is necessary to consider an *a priori* homogeneity constraint to improve the classification result.

Such a method exists in the variational framework [SAM 00] and then [AUB 02a]. The *a priori* homogeneity constraint can be introduced by adding to the usual classification term, defined by a potential $W(I)$, an additional regularization term $\varphi(|\nabla I|)$ translating the homogeneity constraint :

$$Q_{\mathrm{pt}}(W, \varphi, \varepsilon; I) \triangleq \int_\Omega \left[\varepsilon \varphi(|\nabla I|) + \frac{1}{\varepsilon} W(I) \right] \tag{10.3}$$

The potential $W(\vec{s}, \vec{m}, I)$ controlling the classification process is built from the previously determined best thresholds and modes according to alternate optimization of the multi-thresholding energy : the stable phases are given by the modes $\vec{m}^\infty$, the unstable phases are given by the thresholds $\vec{s}^\infty$. As a result, the entire scheme is becoming attractively unsupervised.

Two different schemes can be considered to get the solution. The first scheme realizes a direct minimization of the dynamic equation associated with $Q_{\mathrm{pt}}(W, \varphi, \varepsilon; I)$. A second scheme performs an alternate minimization using the semi-quadratic algorithm [SAM 00], [AUB 02a], [AUB 02b].

The method of classification using the alternate minimization is better but much slower and more constraining than the method using direct minimization. The results obtained by the latter are quite good and easier to control. They are used hereafter.

10.2.3. *Identification*

Now, in order to identify the nature of the noise, we use either one or the other homogeneous regions map I_h, obtained at the output of the previous classification stage.

We calculate on each local estimation kernel $\Phi_{SW}(I)$, where SW are scanning windows with different increasing size, unbiased local statistics of interest $\mathrm{mean}_{\Phi(I)_{i,j}}$, $\mathrm{var}_{\Phi(I)_{i,j}}$:

$$\Phi_{SW}(I)_{i,j,i',j'} \triangleq \left\{ \begin{array}{l} 1 \text{ if } I_h(i',j') \text{ is the most frequent label,} \\ 0 \text{ otherwise,} \\ (i',j') \in SW \end{array} \right. \tag{10.4}$$

The initial method was based on the discrimination of the statistical series according to the window size. The analysis of the corresponding local statistical series shows that it is not relevant to discriminate them only according to the window size. Indeed, it is better to consider the number of pixels taken into account for their calculation. We thus ensure that two local statistical estimators, belonging to the same set and then calculated with the same number of pixels, follow the same statistical law. It should be noted that this law is a priori unknown, except for the Gaussian additive noise variance. In this last case, it is a χ^2 law.

For each number of pixels taken into account, we obtain a coherent set of statistics.

We now consider the most representative set is that which contains the most elements at the output of the preceding classification procedure of local statistical estimators. We fit a polynomial regression to the set of variance estimates, according to the average estimates $\mathrm{mean}_{\Phi(I)_{i,j}}$. From a practical point of view, the implementation is not more complex than for the linear regression. We decide the noise is additive if the zero order polynomial regression gives the best approximation within the meaning of least squares criterion. The noise is multiplicative if the best approximation is obtained by the second order regression. The experiments carried out have demonstrated that blind noise identification system performs with appropriate accuracy [**?**] and this allows passing to more detailed consideration of the methods for blind evaluation of statistical characteristics for particular noise types under assumption that it is identified correctly.

10.3. Methods for blind evaluation of noise statistical characteristics

As said and shown above, an image can be corrupted by a given type of noise (additive, multiplicative, impulse) or by a certain combination of two or even three types of noise (additive+impulse, multiplicative+impulse, additive+multiplicative, all three types). Certainly, less number of noise types is present, simpler the task to evaluate the corresponding characteristics is. Thus, let us consider first simpler practical situations and then pass to more complex. Note that throughout this section, we focus on evaluating statistical characteristics of noise with spatially invariant statistical and spatial correlation characteristics. First, recall the requirements to blind methods for evaluation of noise statistical characteristics. Ideally, a blind method of variance evaluation should meet the following requirements :

1) to produce an unbiased estimate (in particular, to give an estimate equal to zero for noise free images) ;

2) to ensure variance of noise variance estimate as close to potential minimum as possible;

3) to provide a method applicability for entire range of possible values of noise variance;

4) to make a method applicable, i.e. providing appropriate accuracy, to any image, i.e., to images characterized by different spectral-correlation properties, and to both cases of i.i.d. and spatially correlated noise;

5) to provide appropriate computational efficiency of algorithmic realization of a method.

The requirements 1) and 2) relate to a method accuracy. Thus, let us briefly discuss how the method accuracy can influence the performance of algorithms of further image processing that exploit the obtained estimates of additive noise variance $\hat{\sigma}_n^2$ or multiplicative noise (relative) variance $\hat{\sigma}_\mu^2$. The paper [ABR 04] dealt with pure multiplicative noise case although as shown in [ABR], the results and conclusions are very similar for pure additive noise case. In [ABR 04] it has been supposed that instead of a true value σ_μ^2, its estimate $\hat{\sigma}_\mu^2 = \kappa \sigma_\mu^2$ is used in filtering of images corrupted by pure multiplicative Gaussian noise where κ is a factor describing noise variance estimation errors and it has values from 0.6 to 1.6.

Therefore, the obtained results of analysis have demonstrated that it is necessary to provide κ within the limits from $0.8 \ldots 0.9$ up to $1.2 \ldots 1.4$. This means that in practice the estimate $\hat{\sigma}_\mu^2$ can be slightly biased and, if the estimate is biased, then, desirably, $\langle \hat{\sigma}_\mu^2 \rangle > \sigma_\mu^2$ where $\langle . \rangle$ denotes ensemble averaging. For example, it can be assumed that a method for blind evaluation of noise variance of multiplicative noise is good if it provides $\hat{\sigma}_\mu^2 \in [1.0\sigma_\mu^2, 1.1\sigma_\mu^2]$.

Estimates $\hat{\sigma}_\mu^2$ should be characterized by not only bias $\Delta_\mu = \langle \hat{\sigma}_\mu^2 \rangle - \sigma_\mu^2$, but also by variance of such estimates $\sigma_{\sigma_\mu}^2 = \langle (\hat{\sigma}_\mu^2 - \sigma_\mu^2)^2 \rangle = \theta_\mu^2$. Then, taking into account the analysis results presented above, $\langle \hat{\sigma}_\mu^2 \rangle \pm 3\sigma_{\sigma_\mu}$ should be within the interval $[0.8\sigma_\mu^2, 1.4\sigma_\mu^2]$. The relative (normalized) root mean square error $\sigma_{\mu_{rel}} = \sigma_{\sigma_\mu}/\sigma_\mu^2$ of estimates $\hat{\sigma}_\mu^2$ should be less than $0.05 \ldots 0.08$ in order to fit the aforementioned interval of κ values assuming Gaussian distribution of estimates $\hat{\sigma}_\mu^2$.

Similar conclusions and practical recommendations ($\langle \hat{\sigma}_n^2 \rangle \pm 3\sigma_{\sigma_n}$ should be within the interval $[0.8\sigma_n^2, 1.4\sigma_n^2]$) hold for additive noise case [ABR] for which $\Delta_n = \langle \hat{\sigma}_n^2 \rangle - \sigma_n^2$ and $\sigma_{\sigma_n}^2 = \langle (\hat{\sigma}_n^2 - \sigma_n^2)^2 \rangle = \theta_n^2$, $\sigma_{n_{rel}} = \sigma_{\sigma_n}/\sigma_n^2$. Analogous practical requirements to accuracy of noise variance estimation follow from the paper [MEL 01] where the designed hard switching locally adaptive filters are based on comparison of local activity indicators to some thresholds. These thresholds are set with taking into account a priori known noise variance or its estimate. One more question is what is potential minimum of variance ($\sigma_{\sigma_n}^2$ or $\sigma_{\sigma_\mu}^2$) of noise variance estimates. Consider a more simple case of additive noise. Suppose an ideal situation that one has a noisy image $\{ I_{ij}^n = I_{ij}^{tr} + n_{ij} \}$, where and $\{ I_{ij}^{tr} \}$ is the true image, and n_{ij} denotes i.i.d. zero mean additive noise with variance σ_n^2 and I, J define image size. Then, $\hat{\sigma}_n^2$ is

$\hat{\sigma}_n^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} (I_{ij}^n - I_{ij}^{tr})^2 / (IJ - 1)$. For i.i.d. noise $\sigma_{\sigma_n}^2 = (\mu_4 - \mu_2^2)/N$ where μ_2 and μ_4 are the second and the fourth central moments of noise PDF, $N = IJ$ is the number of samples of an observed noise. The second moment μ_2 is equal to the true value $\sigma_{\sigma_n}^2$ while the fourth central moment μ_4 depends upon noise PDF. For Gaussian noise $\mu_4 = 3\sigma_n^4$, for additive noise with other PDFs μ_4 is also proportional to σ_n^4 but proportionality factor is smaller than 3 for light-tail PDFs and it can be considerably larger than 3 for heavy-tail PDFs.

Then, for Gaussian noise, one gets potentially minimal $\sigma_{\sigma_{min}}^2 = 2\sigma_n^4/N$ and, thus, minimal $\sigma_{n_{rel}} = \sqrt{2/N}$. Thus, for any method it can not be expected that good accuracy of variance estimation can be provided for small size images. If noise is not i.i.d., then the accuracy of noise variance estimation characterized by $\sigma_{n_{rel}}$ is poorer.

The requirements 3) and 4) at the beginning of this subsection relate to a method universality. Obviously, a "good" method of blind evaluation of noise variance should provide appropriate accuracy for entire range of possible values of noise variance. Otherwise, a method applicability is considerably restricted. Similarly, a method can be considered "good" and practically applicable if it provides appropriate accuracy for any image, i.e., for images characterized by different spatial spectral-correlation properties.

Usually the most complicated cases are images containing a lot of edges, details and/or texture regions, i.e. a large percentage of pixels belong to heterogeneous regions [LUK 02b, LUK 05a, PON 03]. While many methods provide appropriate accuracy for real life images containing many homogeneous regions, majority of them fail being applied to images containing a large percentage of heterogeneous fragments [PON 03].

Although the common i.i.d. assumption about noise is considered, it often does not take place in practice. There are quite many factors that can result in spatial correlation of noise in images. These factors are image oversampling, interpolation used inside an imaging system, the dependence of linear resolution on distance [PON 08, PAL 07], etc. Fig.10.1a shows an L-band SAR image with manually selected homogeneous regions of rectangular shape marked by digits. A 2-D autocorrelation function (ACF) (see Fig.10.1b) has been calculated for a homogeneous 32x32 fragment of this image. As seen, the ACF has small side lobes; and its main lobe is not a Dirac -function. Essential correlation is observed for central horizontal cross-section of the 2-D ACF. Similar estimates of 2-D ACFs have been obtained for other homogeneous regions of the image in Fig.10.1a. This means that in this case the noise cannot be considered i.i.d.

Note that one might have no a priori information that allows deciding in advance if it is possible to consider the noise i.i.d. or not. Obtaining of such information requires additional efforts and a priori known or pre-estimated variance of noise [PON 08]. Then one comes to a closed loop. A way out is to design and apply such techniques for blind evaluation of noise variance that are able to perform well enough for reasonably wide limits of variation of noise spatial correlation properties (see subsection

10.3.4 for details).

(a) L-band SAR image   (b) 2-D ACF for one fragment 32×32
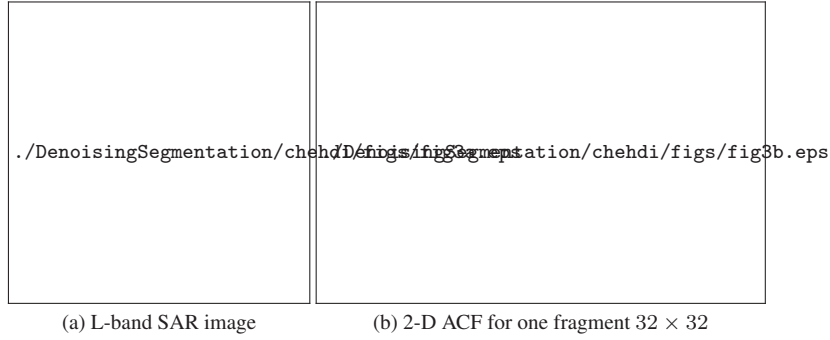
Figure 10.1. *(a) A real life L-band SAR image with marked homogeneous regions and (b)* 32×32 *pixel estimate of 2-D ACF for one fragment*

A method of blind evaluation of noise statistics should provide appropriate computational efficiency of its algorithmic realization (implementation). This is especially true if one deals with processing multispectral or hyperspectral images of large size. Then a more accurate but considerably "slower" method can be of no use since analysis should be performed in a limited time.

Therefore, below we consider possible approaches to blind evaluation of noise characteristics from the viewpoint of aforementioned requirements.

10.3.1. *Estimation of additive noise variance*

Blind methods can be classified into five main groups [LUK 06a] :

1) the methods based on curve (line) fitting (making up linear regression) into obtained sets of local estimates $\{\hat{\sigma}_{loc_k}(\overline{I}_{loc_k})\}$ where $\hat{\sigma}_{loc_k} = (\sum_{i,j \in G_k}(I_{ij}^n - \overline{I}_{loc_k})^2/(|G_k| - 1))^{1/2}$ and here k denotes the index of the k-th block, $|G_k|$ denotes the number of pixels in a k-th block, $|G_k|$ are equal for all k and are of the order of tens (for example, 25, or 49, or 81) [LUK 08a, LUK 08b] ; such approach is, in particular, considered for speckle noise in images formed by SARs ;

2) the methods that also imply obtaining the sets of local estimates $\{\hat{\sigma}_{loc_k}(\overline{I}_{loc_k})\}$ at the first stage with processing these sets by finding the estimate histogram maximum or by applying some robust (based on order statistics, bootstrap, quantiles, sample myriad, etc.) techniques [LUK 05a, LUK 06b] at the final stage ;

3) the methods that evaluate noise variance in spectral domain using wavelet, discrete cosine (DCT) and other orthogonal transforms [PON 03, MAL 98, SEN 02] and analysis of spectral coefficient absolute values ;

4) different combined techniques that, for example, use image segmentation or classification before applying linear regression, apply robust methods for processing spectral coefficient absolute values [PON 03], use iterative procedures for improving the accuracy of obtained estimates [LUK 02b] ;

5) methods of noise parameters estimation within a maximum likelihood framework recently proposed in [USS 07, LUK 09].

The basic idea behind the first group of methods is the following. It is assumed that for small size blocks both normal and abnormal local estimates are formed. Normal ones are the estimates close to a true value of additive or multiplicative noise variance (depending upon a decision what is noise type). Normal estimates are obtained for blocks that belong to image homogeneous or quasi-homogeneous regions. Abnormal estimates are those ones that considerably differ from a true value ; commonly abnormal estimates are sufficiently larger than a true value.

For additive noise case, there can be many local estimates $\{\hat{\sigma}_{loc_k}(\overline{I}_{loc_k})\}$ that are quite close to the true value σ_n. Then, by fitting horizontal line one can get an estimate of $\hat{\sigma}_n$. Similarly, in case of pure multiplicative noise, such line is fitted that comes via coordinate system origin, and its slope characterizes $\hat{\sigma}_\mu$. Note that such line fitting should be robust [?] in order to minimize the influence of abnormal estimates (see also example in [LUK 06a]. Since for additive noise the statistics of local estimates practically does not depend upon local mean, horizontal line fitting in scatter-plots can be replaced by simpler operations for the methods of the group 2) or refined by combining different techniques for the methods of the group 4). These methods will be described and analyzed below more in detail.

A distinctive feature of the methods of the group 3) is that they operate in "spectral domain" where spectrum can be obtained in different orthogonal bases : wavelet, DCT, etc. In general, all these methods rely on the following common assumption. The basic energy of an image or an image fragment (block) is concentrated in a relatively rather small percentage of spectral coefficients of an orthogonal transform used. These spectral coefficients possess rather large amplitudes (absolute values). Statistics of other spectral coefficients is basically determined by noise. And this makes possible to somehow process absolute values of such spectral coefficients in order to obtain an estimate of additive noise variance. This processing is used for separating image and noise information in spectral domain. Note that these properties are also put into basis of transformed based denoising that has become very popular nowadays [TSY 05, POR 03, LUK 07c, MAL 98, SEN 02].

The methods based on the abovementioned properties differ in details. For the methods based on using wavelet transforms commonly only spectral coefficients for high-high (HH) component are considered and processed [MAL 98, SEN 02]

in a robust manner. The need in using robust estimates stems from distribution of wavelet coefficients that obeys mixed Gaussian and Laplacian distribution [SEN 02]. The Gaussian PDF is the contaminated one and Laplacian is the contaminating distribution with smaller probability. Similarly, for methods based on DCT in blocks, it is possible to obtain a set of local estimates of additive noise variance as $\hat{\sigma}_n^2(p,q) = (1.483\mathrm{med}\{y_{kl}(p,q)\})^2$, where $y_{kl}(p,q)$ denotes the pq-th spectral coefficient determined for the kl-th block, $0 \leq k \leq I - M_b + 1, 0 \leq l \leq J - N_b + 1$ M_b and N_b define the block size. In [PON 03], a three-stage procedure is put forward. At the first stage one has to perform local DCT for all possible positions of overlapping blocks and, thus, to get the set of arrays $\{y_{kl}(p,q)\}$. At stage 2, the estimations $\hat{\sigma}_n^2(p,q)$ are obtained for all p and q except $p = 0$, $q = 0$. For the third stage, several ways to process the set of estimations $\hat{\sigma}_n^2(p,q)$ can be followed : 1) choose the N_s smallest values and perform their averaging ; 2) choose the N_s smallest values and find the median value ; 3) calculate the myriad of the N_s smallest or for all estimates $\hat{\sigma}_n^2(p,q)$. Recall that the sample myriad is a maximum likelihood estimator for Cauchy distribution. Thus, several special robust operations can be used to minimize the influence of large spectral coefficients on the final estimate. In [PON 03], we have recommended $N_s = 16$.

The methods of the group 3) possess several valuable advantages. First, they provide rather good accuracy even if images contain a lot of texture, edges and fine details. For many considered noise-free images [PON 03] the noise variance estimates were equal to zero or to appropriately small values (recall the requirements to blind evaluation methods stated at the beginning of subsection 10.3). For noisy images the obtained estimates of noise variance have been characterized by good accuracy [PON 03]. Since for many orthogonal transforms there are fast algorithms [MAL 98], the corresponding methods for blind evaluation of additive noise variance are rather fast. However, these methods have several drawbacks. The main drawback is that these methods provide appropriate accuracy only if noise is i.i.d. [LUK 08a]. This deals with the fact that assumption about mixed Gaussian and Laplacian distribution of spectral coefficients is not anymore valid for spatially correlated noise. More details will be given in subsection 10.3.4.

Finally, the methods of the group 5) are based on estimation of noise (additive, multiplicative, mixed additive and impulse, mixed multiplicative and impulse) parameters within a maximum likelihood framework. One advantage is that a well-developed theory is put into their basis. Their accuracy, as it will be shown in subsection 10.3.3, is high for images of different structure including textural ones. However, till the moment these methods are designed for the case of i.i.d. additive or multiplicative noise. Besides, processing takes a lot of time. These disadvantages restrict practical application of these methods.

Therefore, let us concentrate below on considering the methods of the groups 2) and 4). Instead of operating with a scatter-plot of local estimates

$\{\hat{\sigma}_{loc_k}(\overline{I}_{loc_k}), k = 1, \ldots, N_{bl}\}$ (N_{bl} is the number of blocks) as the methods of the group 1) do, the methods of the groups 2) and 4) "work" with a distribution of local estimates $\{\hat{\sigma}^2_{loc_k}\}$.

The basic assumptions put into basis of this group of methods are the following :

1) there are quite many "normal" local estimates, i.e., the estimates quite close to the true value of noise variance, and such estimates form one mode of the distribution of $\{\hat{\sigma}^2_{loc_k}\}$;

2) abnormal estimates form a one-side heavy tail of the distribution and the influence of this tail on the final estimate is avoided (minimized) by applying robust or other appropriate methods for processing of $\{\hat{\sigma}^2_{loc_k}\}$.

The simplest method of distribution mode finding is based on searching a histogram maximum and a coordinate that corresponds to it. However, as it is shown in [LUK 02b], this method suffers from several drawbacks. In case of unknown noise variance it is difficult to select appropriate subinterval size Δh or, equivalently, the number of subintervals N_h (or bins).

And the method accuracy is even worse if an image under analysis is textural [LUK 02b]. Appearance of the histogram and preliminary analysis of statistics of normal and abnormal local estimates show that it is possible to use the following model PDF for describing the statistical properties of the set of block variance estimates :

$$\rho(x) = p \cdot \rho_{normal}(m, \sigma^2) + (1 - p) \cdot \rho_{abnormal}(0, M) \qquad (10.5)$$

where $\rho_{normal}(m, \sigma^2)$ is the PDF with the mean $m = \sigma^2_n$ and variance σ^2 (do not mix σ^2 with additive noise variance σ^2_n and multiplicative noise relative variance σ^2_μ) that relates to normal local estimates, $\rho_{abnormal}(0, M)$ denotes the PDF which is non-zero within the limits from 0 to M , $M >> m$ (that relates to abnormal local estimates), p is the parameter that, in fact, shows the percentage of blocks belonging to image homogeneous regions. In [LUK 02b] instead of (10.5) the following simplified model has been used :

$$\rho(x) = p \cdot \rho_{gaussian}(m, \sigma^2) + (1 - p) \cdot \rho_{uniform}(0, M) \qquad (10.6)$$

In model (10.6), the variance σ^2 is determined by m (the true value to be estimated), the block size, the PDF of noise and spatial correlation properties of noise. Variance σ^2 is proportional to m^2 and approximately inversely proportional to the number of pixels in blocks. In practice, the ratio σ/m for 5×5 blocks is of the order $0.3 \ldots 0.4$ for i.i.d. noise.

On one hand, the model (10.5) can be replaced by the simplified model (10.6) under certain conditions. First, the PDF $\rho_{normal}(m, \sigma^2)$ can be replaced by $\rho_{gaussian}(m, \sigma^2)$ if according to the central limit theorem the normal local estimates $\hat{\sigma}^2_{n_{lock}}$ defined by $\hat{\sigma}^2_{n_{lock}} = \sum_{i,j \in G_k}(I^n_{ij} - \overline{I}_{loc_k})^2/(|G_k| - 1)$ obtained in image homogeneous image regions obey Gaussian distribution where $|G_k|$ defines the number of pixels belonging to the k-th block. The central limit theorem holds for Gaussian i.i.d. noise starting

from $|G_k| = 25$, i.e. for 5×5 blocks [LUK 02b]. However, for spatially correlated and/or heavier PDF noise the block size should be larger [LUK 08a]. Second, other model PDFs can be used instead of $\rho_{uniform}(0, M)$ to simulate statistical properties of abnormal estimates, but the choice of this model usually does not influence much the performance of the designed blind methods of noise variance evaluation.

On the other hand, by applying the simplified model (10.6) for which PDF data can be easily simulated and PDF parameters can be varied, it is possible and convenient to compare the accuracy of different methods, to analyze method accuracy for different model parameters and, respectively, for images with different properties. In particular, the model (10.6) has been used for design of new methods in the papers [LUK 02b, LUK 05a, LUK 06b, LUK 07b].

As shown above, even for the histogram method, it is worth having some preliminary estimate $\hat{\sigma}^2_{n_{prel}}$ in order to set the algorithm parameter Δh in a proper manner. Similar task arises for the myriad based method [LUK 02b] where a tunable parameter of a sample myriad is to be set with taking into account $\hat{\sigma}^2_{n_{prel}}$. Recall that the sample myriad is defined as [J.G 01] :

$$\hat{\beta}_k = \text{myriad}\{K_m, x_1, \ldots, x_n\} = \arg\min_\beta \sum_{i=1}^{n} ln[K_m^2 + (x_i - \beta)^2] \qquad (10.7)$$

where K_m is a tunable parameter, n is a sample size (it is equal to N_{bl} for the considered application), and $x_1, \ldots, x_n$ are the elements of a data sample (the set of local variance estimates obtained in blocks). Note that for the considered application we exploit not the property of the myriad estimator as a maximum likelihood one but its ability to find a distribution mode [LUK 05b].

The accuracy of the myriad based method can be considerably improved with properly selected K_m. However, again there is a problem in adjusting the parameter K_m that should be determined as $K_m = 0.3\hat{\sigma}_{n_{prel}}$ since one does not know m. Thus, one needs to have some preliminary estimate $\hat{\sigma}^2_{n_{prel}} \approx m$.

Another drawback is that the myriad method [LUK 02b] quickly loses its efficiency for p smaller than 0.2 [LUK 05a], i.e., for highly textured images. Summarizing the above said, the presented results demonstrate the following shortcomings of the histogram and myriad based methods. First, they both require having at disposal a preliminary estimate $\hat{\sigma}^2_{n_{prel}}$ for setting appropriate parameters Δh or K_m, respectively. Second, although quite many images met in practice are characterized by a rather large p in the model (10.6) of local estimates, it is desirable to design methods applicable for images with p of about 0.1, i.e. for highly textural images with a small percentage of blocks in image homogeneous regions. Thus, an important task is to design a method for obtaining preliminary estimate $\hat{\sigma}^2_{n_{prel}}$.

One early work [ZEL 96] intended on designing a method for obtaining an approximate (rough) estimate $\hat{\sigma}^2_{n_{prel}}$ proposes to find the 30% percentile (the $0.3N_{bl}$-th order statistic) of the set of local estimates and to use it as $\hat{\sigma}^2_{n_{prel}}$ or even as a final estimate $\hat{\sigma}^2_{n_{fin}}$. Later the drawback of this method has been understood. The bias of $\hat{\sigma}^2_{n_{prel}}$

depends upon p and is small enough if p is about 0.6 that holds for many natural images but, certainly, not for highly textural images. However, a positive feature of this method called the quantile method [LUK 02b] was that it ran us into idea to analyze more in detail statistical properties of order statistics for the distribution (10.6).

A thorough analysis of statistical properties of order statistics for the model distribution (10.6) carried out in the paper [LUK 05a] has shown the following. A q-th order statistic for which the following condition is valid $q \approx pN_{bl}/2$ produces practically unbiased estimate of the mode of the distribution (10.6) and, moreover, allows estimating p. Another observation is that this order statistic possesses the smallest variance among all order statistics $q = 1, \ldots, N_{bl}$. This property was put into bootstrap based algorithm for determination of $\hat{q}$. Recall that among a set of possible estimators bootstrap based method allows finding and selecting an estimator that produces a smallest variance of estimates.

However, better results can be obtained if the bootstrap based method is used in a certain combination with other methods. For example, the bootstrap based method can be used for producing the final estimate $\hat{\sigma}^2_{n_{fin}}$ if $\hat{p} = 2\hat{q}/N_{bl}$ occurs to be small. Otherwise, it is applied for obtaining $\hat{\sigma}^2_{n_{prel}}$ that is further used for calculation of, for instance, K_m for the myriad based method that, in turn, produces $\hat{\sigma}^2_{n_{fin}}$ [LUK 05a]. Such modifications have considerably improved robustness (in wide sense) of the corresponding combined bootstrap-myriad estimator and made it applicable for processing highly textural images characterized by non-zero but small p.

Meanwhile, recall that the bootstrap based method exploits finding and using only one, the $\hat{q}$-th order statistic. Other order statistics with indices close to $\hat{q}$ are not exploited although they are characterized by variance close to the variance of the $\hat{q}$-th order statistic [LUK 05a].

This observation served as a basis for several recently designed modifications that can be referred as inter-quantile based methods [LUK 08a, LUK 06b, LUK 07b]. The core of the inter-quantile approach is the use of the set $\{\hat{\sigma}^2_{n_{loc_k}}, k = 1, \ldots, N_{bl}\}$ and its ordering in ascending order with obtaining $X^{(k)} \geq X^{(k-1)}, k = 2, \ldots, N_{bl}$. Then the set $\{X^{(k)}\}, k = 1, \ldots, N_{bl}$ is processed with finding such $X^{(k_{left})}$ that $X^{(k_{left}+s)} - X^{(k_{left})}$ is the smallest among all $X^{(k_{left}+s)} - X^{(k_{left})}, k = 1, \ldots, N_{bl} - s$ where s is the method parameter. To provide the appropriate performance of the inter-quantile method for images with $p \geq 0.1$, it is recommended to select $s = [0.1N_{bl}]_2$ where $[.]_2$ denotes rounding-off to the nearest even integer. Then, the estimate $\hat{\sigma}^2_{n_{prel}} = X^{(k_{left}+s/2)}$.

In general, this estimate can be used as $\hat{\sigma}^2_{n_{fin}}$ since it is accurate enough [LUK 06b]. However, there are additional ways to improve the performance of the inter-quantile based method [LUK 06b, LUK 07b]. These ways deal with the fact that $p \approx (k_{left} + s/2)/N_{bl}$ and this allows determining what order statistics $X(k)$ correspond to normal local estimates : $k \leq pN_{bl}$. In turn, this allows performing

different operations with these order statistics including averaging, robust processing, fitting polynomials, selecting s^{opt} for the second iteration of inter-quantile method, etc.

In subsection 10.3.4 we will give the detailed description of the algorithm proposed and tested in [LUK 07b] with taking into account possible spatial correlation of noise. An alternative way deals with the use of image pre-segmentation [LUK 06a, LUK 07b] used also for determination of noise type (see Section 10.2). Recall that performance of the myriad based method depends upon a weight of a heavy tail of the distribution (10.6) [LUK 06a]. The presence of this one side tail "pulls" the estimate towards heavy tail side [LUK 02b] and this leads to increasing the positive bias Δ_n. This is an inherent property of robust estimators of contaminated non-symmetric distribution location and some mode finders [LUK 05b].

If abnormal local estimates of noise variance that form "right-hand" tail of the distribution (10.6) are somehow removed, it is possible to expect that the bias can be considerably decreased. Recall here that image classification methods described in section 10.2 allow obtaining preliminary information about image homogeneous regions.

For getting such discrimination map of pixels that belong to image homogeneous regions, we have applied the following post-processing of a classified image. Within 5×5 scanning window it has been tested to how many segmentation classes the pixels belong [LUK 06a]. We propose to calculate local estimates $\{\hat{\sigma}_{loc_k}(\overline{I}_{loc_k})\}$ only for those scanning window (block) positions for which all scanning window pixels belong to the same segmentation class. Then, the corresponding pixel of discrimination map is set equal to $DM_{ij} = 128$ (homogeneous region). Whenever they belong to two different segmentation classes, $DM_{ij} = 0$ (edge/detail neighborhood), and for three to more segmentation classes, $DM_{ij} = 255$ (texture).

In this manner one obtains a trimmed set of local estimates $\{\hat{\sigma}_{loc_k}(\overline{I}_{loc_k})\}_{trim}$ where abnormal estimates are mainly removed. After trimming performed on basis of class discrimination the amount of normal estimates has partly reduced. At the same time, the amount of abnormal estimates has decreased much more radically. As the result, the distribution of $\{\hat{\sigma}_{loc_k}(\overline{I}_{loc_k})\}_{trim}$ has considerably less heavy tail than the distribution of the original estimates $\{\hat{\sigma}_{loc_k}(\overline{I}_{loc_k})\}$. In such situations, less robustness of distribution mode estimate is required to provide appropriate bias and variance [LUK 05b]. Pre-segmentation can be also combined with other methods of mode finding, e.g., with the improved inter-quantile method [LUK 07b] and with Spectral domain techniques, e.g., with the DCT based method [PON 03].

In general, the analysis of simulation results presented in the papers [LUK 02b, LUK 05a, LUK 06b, LUK 06a, PON 03, LUK 07b, LUK 05b] allows concluding the following :

1) estimation bias Δ_n commonly contributes more than estimation error variance θ_n^2 to the aggregate error ε_n especially if an image size is rather large and noise variance is small ; because of this, in method design, it is necessary to pay more attention to diminishing bias absolute value than to decreasing θ_n^2 ;

2) if one deals with an image of larger size and/or characterized by larger p, it is possible to expect better accuracy of blind estimation of noise variance ;

3) the most crucial practical cases in the sense of the largest relative error (worst accuracy) are small p, small variance of noise, and small size of an image to be processed ; note that all these factors do not depend upon a user.

10.3.2. *Estimation of multiplicative noise variance*

Generally speaking, almost all methods applicable to blind evaluation of additive noise variance can be also used for estimation of multiplicative noise relative variance σ_μ^2. Note that it was understood more than twenty years ago that pure multiplicative noise can be converted to pure additive noise by applying forward homomorphic transforms of logarithmic type. Such logarithmic transform can be defined as $I_{ij}^h = a \log_b(I_{ij})$ where a and b are the transform parameters [OKT 07]. After estimating $\hat{\sigma}_n^2$ in the obtained image I_{ij}^h, it becomes possible to calculate $\hat{\sigma}_\mu^2 = \hat{\sigma}_n^2 (\ln b^2 / a^2)$ in the original image corrupted by multiplicative noise. This way is potentially possible for applying the spectral domain methods [PON 03, SEN 02] for estimating $\hat{\sigma}_n^2$ in I_{ij}^h. But we would like to stress that while using homomorphic transforms one should be careful since appropriately good results are obtained only if multiplicative noise is Gaussian and $\sigma_\mu^2 \leq 0.03$ [OKT 07]. Because of this, below we concentrate on considering the methods of blind evaluation of σ_μ^2 without using homomorphic transforms and keeping in mind that in practice the values σ_μ^2 can be up to 1.0, this happens for one-look SAR images [MEL 00].

Consider line fitting results. In the case of pure multiplicative noise, such line is fitted that it comes via coordinate system origin, and its slope characterizes $\hat{\sigma}_\mu$. But providing of appropriate robustness of line (curve) fitting algorithms to outliers becomes a serious problem if a percentage of outliers (abnormal local estimates for the considered application) is large, i.e. if p in the model (10.6) is small. This is a critical drawback of the methods based on line fitting that restricts their practical use.

Meanwhile, other methods of blind evaluation earlier described in subsection 10.3.1 for pure additive noise case can be very similarly applied for obtaining $\hat{\sigma}_\mu^2$. The only difference with respect to the additive noise case is in obtaining initial local estimates. Instead of $\sigma_{n_{loc_k}}^2 = \sum_{i,j \in G_k} (I_{ij}^n - \bar{I}_{loc_k})^2 / (|G_k| - 1), k = 1, \ldots, N_{bl}$ one has to obtain $\hat{\sigma}_{\mu_{loc_k}}^2 = \hat{\sigma}_{n_{loc_k}}^2 / \bar{I}_{loc_k}^2, k = 1, \ldots, N_{bl}$. Then all other operations that have to be performed with $\{\hat{\sigma}_{n_{loc_k}}, k = 1, \ldots, N_{bl}\}$ according to the methods given in subsection 10.3.1 are to be carried out for the set $\{\hat{\sigma}_{\mu_{loc_k}}, k = 1, \ldots, N_{bl}\}$. This can be the myriad based method [LUK 02b], the bootstrap based technique [LUK 05a] or the combined bootstrap-myriad based method [LUK 05a], the inter-quantile based methods [LUK 06b, LUK 07b], the combined methods presuming image pre-segmentation [LUK 08a, LUK 06a], etc.

The advantages and drawbacks of these methods for the case of pure multiplicative

noise are the same as discussed in the previous subsection. However, there are two differences to be mentioned. First, as seen from the definition $\hat{\sigma}^2_{\mu_{loc_k}} = \hat{\sigma}^2_{n_{loc_k}}/\overline{I}^2_{loc_k}, k = 1, \ldots, N_{bl}$, calculation of any local estimate uses not only estimation of $\hat{\sigma}^2_{n_{loc_k}}$ in a k-th block but also estimation of a local mean $\overline{I}_{loc_k}$ where both are random variables. This worsens relative accuracy of the local estimates $\hat{\sigma}^2_{\mu_{loc_k}}$ in comparison to the initial local estimates $\hat{\sigma}^2_{n_{loc_k}}$. Second, the magnitude order of estimates $\hat{\sigma}^2_{n_{loc_k}}, k = 1, \ldots, N_{bl}$ is larger than that of the normal estimates in the set $\hat{\sigma}^2_{\mu_{loc_k}}, k = 1, \ldots, N_{bl}$ which are smaller than or, in the least case, comparable to unity and only the abnormal estimates of multiplicative noise variance can be larger than 1 [PON 06b]. If images are represented as 16 bit data arrays, estimates $\hat{\sigma}^2_{n_{fin}}$ can be even larger like for some sub-bands of hyperspectral AVIRIS images [PON 07]. Irrespectively to the bit number used for image representation, final estimates $\hat{\sigma}^2_{\mu_{fin}}$ are usually smaller than unity and very rarely they can be slightly larger than unity. The latter might happen for one-look SAR images with fully developed speckle.

One more practical aspect is worth considering. The local estimates $\hat{\sigma}^2_{\mu_{loc_k}}$ and $\hat{\sigma}^2_{n_{loc_k}}$, $k = 1, \ldots, N_{bl}$ are obtained in blocks tessellating either entire image or image fragments determined by discrimination map (then one obtains the corresponding trimmed sets $\{\hat{\sigma}^2_{\mu_{loc_k}}(\overline{I}_{loc_k})\}_{trim}$, or $\{\hat{\sigma}^2_{n_{loc_k}}(\overline{I}_{loc_k})\}_{trim}$). These blocks can be non overlapping, partly overlapping or fully overlapping as in DCT based filtering [LUK 07c]. The use of non overlapping blocks leads to the fastest calculation of the local estimate set and its processing with obtaining final estimates. However, the accuracy of estimation in this case is slightly worse than in the case when partly or fully overlapping blocks are used. Overlapping or non overlapping of blocks practically does not influence bias of final estimates but affect estimation error variance.

In the case of overlapping blocks, computation load increases radically. But it can be reduced by means of applying running algorithms for determining the local estimates of variance and mean.

Similarly to additive noise case, the local estimates $\{\hat{\sigma}^2_{\mu_{loc_k}}(\overline{I}_{loc_k})\}$ can be distorted by clipping effects if (for Gaussian multiplicative noise) $\overline{I}_{loc_k}(1 + 3\sigma_\mu) > I_{max}$, where I_{max} denotes the maximal value for a given image representation. This means that if σ_μ is rather large, e.g. 0.3, one should be careful with local estimates $\hat{\sigma}^2_{\mu_{loc_k}}$ obtained for blocks with local mean $\overline{I}_{loc_k} > I_{max}/2$. Such estimates can occur smaller than they should be due to clipping effect.

It is difficult to compare the performance of the methods for blind evaluation of additive noise variance and multiplicative noise variance in terms of bias Δ, variance θ^2 and or aggregate error ε since estimated σ^2_n and σ^2_μ are values of different order (see above). However, it is possible to compare relative errors $\sigma_{nrel} = \sqrt{\varepsilon_n}/\sigma^2_n$ and $\sigma_{\mu rel} = \sqrt{\varepsilon_\mu}/\sigma^2_\mu$ where $\varepsilon_n = \Delta^2_n + \theta^2_n$ and $\varepsilon_\mu = \Delta^2_\mu + \theta^2_\mu$, subscripts n and μ relate to additive and multiplicative noise, respectively. Analysis and comparisons have shown that for the same test image the values σ_{nrel} and $\sigma_{\mu rel}$ are quite close if PSNR of the images corrupted by additive and multiplicative noise are approximately the

same and the same method for blind evaluation of noise variance is applied. For typical images corrupted by multiplicative noise with σ_μ^2 in the limits $0.01 \ldots 0.2$ the provided $|\Delta/\sigma_n^2|$ is usually smaller than 0.12 and θ_n^2/σ_n^2 is less than 0.03 for image size 512×512 pixels. Thus, the requirements formulated in subsection 10.3 can be satisfied.

Other conclusions [LUK 02b, PON 03] coincide with those one given for additive noise case at the end of previous subsection, i.e. estimate bias Δ_μ has larger effect on $\varepsilon_\mu = \Delta_\mu^2 + \theta_\mu^2$ than θ_μ^2 especially for images of rather large size. Accuracy is better if an image has large size, is characterized by larger p and σ_μ^2 is not very small. On the contrary, it is not worth expecting unbiased estimation and small $\sigma_{\mu rel}$ if p is small, multiplicative noise is not intensive and an image to be processed has small size, e.g., 100×100 pixels.

For the real life SAR image represented in Fig.10.1a, the local estimates $\{\hat\sigma_{\mu_l}, l = 1, \ldots, 7\}$ obtained on the seven marked selected homogeneous regions are equal to $0.166, 0.158, 0.169, 0.163, 0.169, 0.117$, and 0.162 [LUK 08a]. Thus, one can expect a blind estimate to be about 0.16. The improved inter-quantile based method described in [LUK 08a] produces $\hat\sigma_{\mu_{fin}}^2 = 0.156$, i.e. quite close to the predicted value. Similar estimates (about $0.14 \ldots 0.16$) have been obtained for other images formed by the same SAR [MEL 00, OKT 07, EGI 01]. Here we would like to thank A.I. Kalmykov Center for Earth Remote Sensing, National Academy of Science and National Space Agency of Ukraine, for offering most real scene images used in this Chapter.

10.3.3. *Evaluation of mixed noise parameters*

Recall that by mixed noise we mean the presence of, at least, two different components of noise in a given image. In our consideration a mixed noise can be additive+impulse [BAR 03], multiplicative+ impulse [MEL 00, LUK 05c], multiplicative+additive [LUK 08b]. Let us start from considering first two combinations.

For the very beginning, let us clarify what is a model of mixed noise and what is considered by impulse noise. A model of a one-channel image corrupted by mixed additive and impulse noise can be

$$I_{ij}^i = \begin{cases} I_{ij}^{tr} + n_{ij} & \text{with probability } 1 - P; \\ A_{ij} & \text{with probability } P; \end{cases} \tag{10.8}$$

and for mixed multiplicative and impulse noise case it can be

$$I_{ij}^i = \begin{cases} I_{ij}^{tr} \mu_{ij} & \text{with probability } 1 - P; \\ A_{ij} & \text{with probability } P; \end{cases} \tag{10.9}$$

where P denotes the probability of impulse noise, I_{ij}^{tr} is the true value of the ij-th image pixel, A_{ij} is a value of impulse observed in the ij-th pixel, n_{ij} and μ_{ij} denote additive and multiplicative noise, respectively [MEL 00]. It is a question what A_{ij} is

to be considered as an impulse. For simplicity, for Gaussian additive or multiplicative noise let us assume that an ij-th pixel is corrupted by an impulse if $A_{ij} \notin [I_{ij}^{tr} - 3\sigma_n, I_{ij}^{tr} + 3\sigma_n]$ or $A_{ij} \notin [I_{ij}^{tr}(1 - 3\sigma_\mu), I_{ij}^{tr}(1 + 3\sigma_\mu)]$, respectively.

There are several possible approaches to determine P and σ_n^2 or σ_μ^2 for the images obeying the models (10.8) or (10.9). They are the following :

1) estimate in a robust manner σ_n^2 or σ_μ^2, then apply some impulse detector that takes into account the obtained estimate $\hat{\sigma}_n^2$ or $\hat{\sigma}_\mu^2$ (depending upon a priori known or pre-determined type of fluctuative noise) and estimate P as $\hat{P} = N_{imp}/IJ$ where N_{imp} defines the number of detected impulses ;

2) apply some impulse detector that does not require a priori knowledge of noise type and statistical characteristics, determine $\hat{P} = N_{imp}/IJ$, remove impulses and then apply some method for blind evaluation of $\hat{\sigma}_n^2$ or $\hat{\sigma}_\mu^2$ that have been described in previous subsections ;

3) use the methods based on maximum likelihood approach to joint estimation of $\hat{P}$ and $\hat{\sigma}_n^2$ or $\hat{\sigma}_\mu^2$ depending upon fluctuative noise type known in advance [USS 07, LUK 09].

A pre-requisite of approach 1) is that there exist methods [A.V 03, PON 05, DOL 01] able to perform impulse detection under conditions of a priori known type of fluctuative (additive or multiplicative) noise and a priori known or pre-estimated $\hat{\sigma}_n^2$ or $\hat{\sigma}_\mu^2$. The methods [A.V 03, PON 05, DOL 01] are based on "joint analysis" of several local activity indicators or a local histogram calculated for a given scanning window. Based on such analysis, a decision is undertaken concerning a scanning window central pixel whether or not it is corrupted by an impulse or it relates to another class (edge, small sized detail, etc.).

However, there are two problems with these techniques. First, all these methods have been successfully tested for Gaussian additive or multiplicative noise and for rather small values of P, about $0.01 \ldots 0.02$, slightly larger P have been analyzed only in [DOL 01]. Second, in case of impulse noise with $P > 0.02 \ldots 0.03$ it becomes problematic to apply blind methods for evaluation of $\hat{\sigma}_n^2$ or $\hat{\sigma}_\mu^2$ described in subsections 10.3.1 and 10.3.2 [LUK 05c].

The reason is the following. The presence of even one pixel corrupted by an impulse in a block G_k for which local estimate is formed according to conventional algorithm $\hat{\sigma}_{n_{loc_k}}^2 = \sum_{i,j \in G_k}(I_{ij}^i - \overline{I}_{loc_k})^2/(|G_k| - 1)$ can lead to considerable increase of this estimate in comparison to the case when impulse is absent. The probability that at least one impulse is present in a block depends upon P and block size $|G_k|$. If at least one impulse is present, a local estimate obtained in image homogeneous region block occurs to be abnormal. Thus, for $|G_k| = 25$ or 49 even P about $0.02 \ldots 0.03$ becomes crucial in the sense that p in the model (10.6) becomes small. Note that a local mean estimate in a block is also influenced by impulse noise. There is one way out – to apply robust estimates of local variance in all blocks [LUK 05c]. For Gaussian noise

such robust estimate can be obtained as :

$$\hat{\sigma}^2_{rob_k} = (1.483 \text{med}_{G_k} \{|I_{ij \in G_k} - \text{med}_{G_k}|\})^2, \tag{10.10}$$

where med_{G_k} denotes a median for this block. Such algorithm provides good, almost unbiased local estimates even if $P = 0.1$ [LUK 05c]. But the estimate (10.10) can be used only for Gaussian fluctuative noise. If fluctuative noise is non-Gaussian, other scale estimators should be applied [SUO 95] and this is possible only if fluctuative noise PDF is a priori known.

The approach 2) implies using impulse detectors that do not require a priori knowledge of fluctuative noise type and statistical characteristics. Currently there are quite many such impulse detectors. Probably, the oldest and most well-known was proposed by Abreu at al. [ABR 96]. There are also many other ones [LEE 02, ZHA 00, ENG 01], see also references in [CHA 04, LUK 05c, ENG 01]. Although many methods of impulse noise detection exist, we are not ready to give a strict recommendation what is the best detector for the considered approach 2).

It is worth noting that there are impulse detectors that are not suited for solving the task of impulse noise parameter determination. This follows from considering a dependence of N_{imp} on detector parameters. Really, one has $\langle N_{imp} \rangle = \sum_{i,j} (P P^{imp}_{i,j} + (1 - P) P^{fa}_{i,j}) = IJ[\frac{P}{IJ} \sum_{i,j} P^{imp}_{i,j} + \frac{(1-P)}{(IJ)} \sum_{i,j} P^{fa}_{i,j}] = IJ[P\overline{P}^{imp} + (1 - P)\overline{P}^{fa}]$, where $P^{imp}_{i,j}$ and $P^{fa}_{i,j}$ denote probabilities of correct detection of impulse in a pixel (i, j) and false alarm probability, respectively, $\overline{P}^{imp}$ and $\overline{P}^{fa}$ are the corresponding probabilities $P^{imp}_{i,j}$ and $P^{fa}_{i,j}$ averaged over entire image. Then an estimate $\hat{P}$ should be obtained as :

$$\hat{P} = \frac{N_{imp}/(IJ) - \overline{P}^{fa}}{\overline{P}^{imp} - \overline{P}^{fa}} \tag{10.11}$$

If one does not know $\overline{P}^{imp}$ and $\overline{P}^{fa}$, the estimate $\hat{P} = N_{imp}/(IJ)$ might occur considerably biased. Probabilities $P^{imp}_{i,j}$ and $P^{fa}_{i,j}$ depend upon detector parameters (most often thresholds) and for a preset threshold depend upon an image at hand. Most existing impulse detectors are empirical. Although many of them are quite efficient, it is difficult to derive dependence of $P^{imp}_{i,j}$ and $P^{fa}_{i,j}$ on a threshold or several thresholds for them.

To remove detected impulses, there are quite many algorithms , see, e.g., [LUK 05c, ABR 96, LEE 02, ZHA 00, ENG 01, CHE 00b]. Most of them are based on applying some robust estimators to data sample formed from pixel values in a neighborhood of a detected impulse.

An advantage of both approaches 1) and 2) is that the corresponding algorithms can perform rather quickly. Although the algorithm (10.10) differs from conventional one, it is fast as well. Computational effectiveness of algorithms for processing local estimates has been considered above and many of them can be implemented and realized easily.

Consider now the approach 3) more in detail. For simplicity, consider the case of mixed additive and impulse noise. The method generalization to multiplicative+impulse noise case is straightforward.

A peculiarity of the approach [USS 07, LUK 09] is that it has been developed for processing images corrupted by mixed noise with a priori unknown parameters. This approach is based on statistical theory of parameter estimation where image filtering problem as well as image and noise parameters estimation problems are solved simultaneously. One more peculiarity of this approach is that 2-D fractal Brownian motion (fBM) models are used for modeling real-life images. An image is divided into fragments and a likelihood function (LF) is obtained for each fragment. Then a joint LF is considered for all fragments and it is optimized simultaneously for fBM field parameters and noise variance.

If impulse noise is present, the core of the method is an iterative procedure of impulse noise detection and estimation of additive noise variance using pixels that are considered "uncorrupted" by impulses.

For detecting the pixels corrupted by impulses, a novel detector has been designed that allows setting a threshold that provides a preset false alarm probability ($P_{i,j}^{fa} = \overline{P}^{fa} = const$). It also allows estimating $P_{i,j}^{imp}$. Then, knowing $P_{i,j}^{fa}$ and $P_{i,j}^{imp}$, we can obtain $\overline{P}^{fa}$ and $\overline{P}^{imp}$, and use (10.11) for estimating P.

Consider the algorithm more in detail. Suppose now that the true image is corrupted by additive and impulse noise

$$I_{ij}^n = I_{ij}^{tr} + n_{ij} \tag{10.12}$$

$$I_{ij}^i = I_{ij}^n(1 - b_{i,j}^{imp}) + A_{ij}b_{i,j}^{imp} \tag{10.13}$$

$$\langle n_{ij} \rangle = 0, \langle n_{ij}n_{i_1j_1} \rangle = \sigma_n^2 R_n(i,j,i_1,j_1) \tag{10.14}$$

where n_{ij} is a normally distributed ergodic random field with zero mean, a variance σ_n^2 and correlation matrix R_n, I_{ij}^n denotes here the image corrupted by additive noise, I_{ij}^i is the image corrupted by mixed additive and impulse noise, b^{imp} is a binary matrix that characterizes corrupted pixels position ($b_{i,j}^{imp} = 1$ if an ij-th pixel is corrupted, $b_{i,j}^{imp} = 0$ otherwise), A_{ij} is an impulse noise value at position with coordinates (i,j) (if impulse is present).

By definition [FAL 03], fBm-field is a Gaussian process $W_{t,s}^H$ (original coordinates at the point $(0,0) - W_{0,0}^H = 0$, $H \in [0,1]$) with correlation function :

$$\langle W_{t,s}^H \dot{W}_{t_1,s_1}^H \rangle = \frac{\sigma_x^2}{2}\left[\sqrt{t^2 + s^2}^{2H} + \sqrt{t_1^2 + s_1^2}^{2H} - \sqrt{(t-t_1)^2 + (s-s_1)^2}^{2H} \right] \tag{10.15}$$

where $\langle . \rangle$ denotes ensemble expectation, σ_x^2 is a variance of increment of fBm process on unit distance, H is a Hurst coefficient.

To take into account image heterogeneous structure, let parameters of fBm-field depend on spatial coordinates ($H = H(t, s)$,$\sigma_x = \sigma_x(t, s)$). But within a small image fragment it is possible to treat them as constant values.

To describe impulse noise, we need to specify some model for b^{imp} and A_{ij}. One particular and most typical case is a spatially uncorrelated impulse noise that has been often considered in literature [MEL 00]. For such a noise we have

$$P(b_{i,j}^{imp} = 1) = P, P(b_{i,j}^{imp} = 0) = 1 - P \qquad (10.16)$$

Impulse noise values A_{ij} are also spatially uncorrelated random values with PDF $f(A_{ij})$. Uniform distribution in the range $[a, b]$ can be a convenient choice for $f(A_{ij})$. Below we consider the problem of simultaneous (joint) estimation of the true image pixels value I_{ij}^{tr}, fBm-field parameters and noise parameters – σ_n^2 and P, based on the observed field $\{I_{ij}^i\}$ and taking into account aforementioned assumptions on statistical properties of noise-free image and noise.

The introduced fBm-model describes an image only locally. Thus, an image is processed in a $N \times N$ scanning window (SW) fashion.

It is proposed to divide the estimation algorithm into two stages : detection and estimation stages, respectively [LUK 09]. At detection stage all pixels are marked either as corrupted by impulse noise or as "uncorrupted" (but corrupted by additive noise). Uncorrupted pixels are then used to estimate fBm-field and noise variance (estimation stage).

Since detection should be based on estimated parameters and estimation should be based on "uncorrupted" pixels (for which $\hat{b}_{ij}^{imp} = 0$), then detection and estimation stages are to be repeated in iterative manner. At the initialization stage, preliminary impulse noise detection is performed. To do this, we have chosen a simple yet rather effective SD-ROM algorithm [ABR 96]. As a result, a primary estimate $\hat{b}^{imp}$ of the matrix b^{imp} is obtained.

Initial value of the variance estimate has to be set at initialization stage. If some a priori information is available it should be taken into account. Then, the algorithm proposed in [LUK 09] solve the four tasks, described now. At the first stage, fBm-fields parameters are estimated based on "uncorrupted" pixels. Then, at the second stage, noisy image filtering is carried out. Here the corrupted pixels values are predicted based on "uncorrupted" pixels in local neighborhood (SW) of a current pixel. The estimated parameters of the true image are used at the third stage to improve impulse noise detection result. At the last (fourth) stage, mixed noise parameters ($\hat{\sigma}_n^2$, P) are estimated. These operations are repeated until convergence condition is satisfied. This condition means that impulse noise detection results at previous and current iterations are identical. Convergence of this algorithm depends on impulse noise properties and can be verified in experiment. For uncorrelated impulse noise, experiments have confirmed algorithm convergence [LUK 09].

The proposed algorithm advantages are very accurate estimation of fluctuative noise variance and probability of impulse noise even if $P = 0.15$ [USS 07, LUK 09]. In particular, the accuracy of estimation of fluctuative noise variance is better than for other

blind methods described in subsections 10.3.1 and 10.3.2. Estimation of P is carried out with high accuracy - bias absolute value does not exceed 0.20%. The method allows not only estimating noise variance $\hat{\sigma}_n^2$ and probability P, but also to evaluate variances of these estimates for a given noisy image. Besides, as the result of image processing it is denoised and its fractal parameters are estimated. However, the method realization requires extensive computations because of the algorithm complexity and necessity to perform several iterations. This is the basic drawback of the maximum likelihood based approach that restricts its practical application especially for multi-channel images. One way out is to carry component-wise processing in parallel.

One more type of mixed noise mentioned at the beginning of this subsection is mixed multiplicative and additive noise. Presence of both components, in particular, in radar images has been recently indicated by tests with noise type identification system [LUK 08b] and it also stems from theoretical and practical research carried out for multi polarization radar images [LOP 07]. Although multiplicative noise is dominant, there are practical situations when it is not worth neglecting the presence of additive noise [LUK 08b]. Then, both estimates $\hat{\sigma}_n^2$ and $\hat{\sigma}_\mu^2$ are to be obtained.

SLAR Real-life Example The preliminary estimates of noise variances for Ka-band SLAR image presented in Fig.10.2a were the following. The multiplicative noise relative variance $\hat{\sigma}_\mu^2$ was of the order $0.004\ldots0.005$ and additive noise variance $\hat{\sigma}_n^2$ was about $10\ldots16$. Suppose these estimates are accurate enough. Then it means that if an image homogeneous region mean $\overline{I}$ is about 20, additive noise variance is larger than fluctuation variance induced by multiplicative component $\hat{\sigma}_\mu^2(\overline{I})^2 \approx 2$. Therefore, for such image homogeneous regions, additive noise influence occurs to be dominant. If an image homogeneous region has a mean intensity larger than $50\ldots60$, multiplicative noise becomes prevailing. Recall that in practice radar images in 8-bit representation can contain quite many homogeneous regions with $\overline{I}$ less than 50; in particular, they can correspond to water surface and bare soils with large moisture content.
In order to study the problem more in detail, the Ka-band SLAR image in Fig.10.2a has been classified using the procedure [KLA 05]. The obtained classification map with eight clusters is presented in Fig.10.2b. This classification map has been processed (see details in subsection 10.3.1) in order to get the map of image homogeneous regions. This map is represented in Fig.10.2c. Gray color pixels correspond to homogeneous regions, white color pixels relate to edges and details whilst black ones mainly correspond to texture. Using this map, a scatter-plot of local estimates of variance has been obtained.
To provide better accuracy of the local estimates we recommend to use overlapping blocks with side size $m = 7$. For convenience, this scatter-plot (Fig.10.3) is presented in such a manner that abscises axis corresponds to $\overline{I}_{loc}^2$ since the supposed dependence of local variance on the local mean is $\sigma_{loc}^2 = \sigma_n^2 + \sigma_\mu^2\overline{I}_{loc}^2$. Thus, if a curve

$\sigma_{loc}^2 = a + bx$ is fitted to data, its parameters a and b directly correspond to the estimates $\hat{\sigma}_n^2$ and $\hat{\sigma}_\mu^2$, and $x = \overline{I}_{loc}^2$.

Fig.10.3 shows that again there are normal and abnormal local estimates. The normal

(a)   (b)   (c)
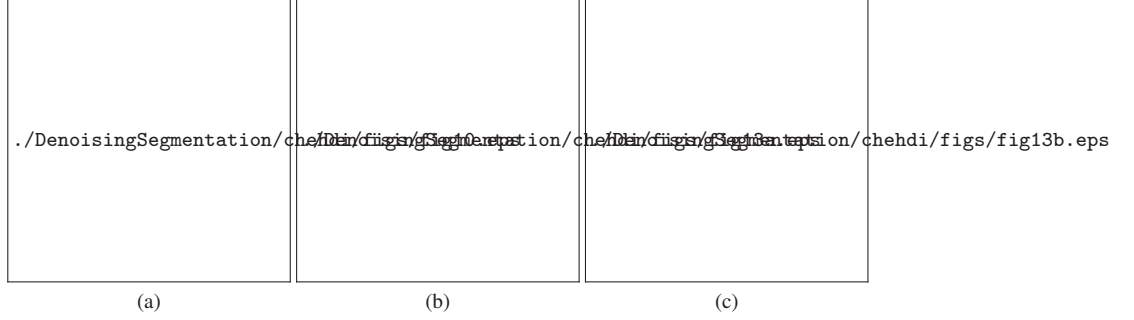
Figure 10.2. *a) Real Ka-band SLAR image b) Segmentation result and c) the corresponding discrimination map*

local estimates are grouped in clusters. Abnormal estimates are considerably larger than normal ones for the same local mean. They are mainly observed for $\overline{I}_{loc}^2$ of about $1000 \ldots 2500$.

For estimating the parameters of mixed noise the technique [LUK 08a] is to be ad-
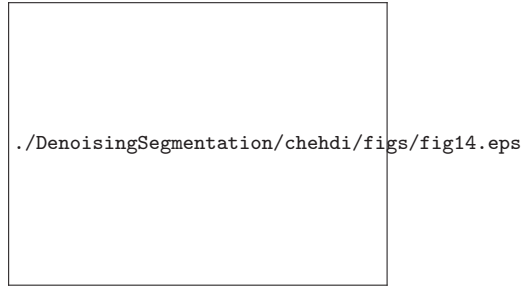
Figure 10.3. *The obtained scatter-plot with approximation line*

vanced in order to obtain partial estimates $\hat{\sigma}_{cl}^2$ for classified clusters. Recall that after image classification [KLA 05], the cluster number and centers are detected and denoted N_{cl} and $C_{cl_m}, m = 1, \ldots, N_{cl}$. As can be seen in Fig.10.3, some neighbor clusters can be easily separated (with large $\overline{C_{cl_m} C_{cl_{m+1}}}$ for the considered case) or partly overlap (with rather small $\overline{C_{cl_m} C_{cl_{m+1}}}$). Let us assume that a k-th block is referred to an m-th cluster if the corresponding central pixel of the block is referred to an m-th cluster (level) in the classified image. Thus, we obtain subsets of blocks for

each cluster. Then, the method [LUK 08a] is applied to each subset of local variances with obtaining $\hat{\sigma}^2_{cl_m}, m = 1, \ldots, N_{cl}$. Similar procedure is applied to the block means for each subset with obtaining the cluster center coordinate $\hat{I}_{cl_m}, m = 1, \ldots, N_{cl}$.

Having a set of pairs $(\hat{\sigma}^2_{cl_m}, \hat{I}^2_{cl_m})$ it is then possible to fit a curve (straight line, see Fig.10.3) and to obtain the estimates $\hat{\sigma}^2_n$ and $\hat{\sigma}^2_\mu$. If all estimates $\hat{\sigma}^2_{cl_m}$ (points) are characterized by equal accuracy and they have been obtained for equidistant arguments $\hat{I}^2_{cl_m}$, the fitting of a straight line, using LMSE criterion performs, in general, well. But both conditions do not hold for the considered case. Estimates obtained for different clusters are, most probably, characterized by different accuracy, and cluster center arguments are usually not placed equidistantly.

The considered mixed multiplicative and additive noise can be viewed as a particular case of signal dependent noise when dependence of local variance on local mean is expressed as a simple sum of two terms, a constant and a linearly increasing ones. Other models of signal-dependent noise are met in practice. One of them is Poisson like noise $\sigma^2_{loc} = K_p I_{loc}$ where K_p is a proportionality factor or a slightly more complicated model $\sigma^2_{loc} = \sigma^2_n + K_p I_{loc}$. Both are typical for situations when an image is formed using photon counting sensors [PAL 07]. A successful attempt of designing an automatic approach to estimation of parameters of these models is described in [FOI 07].

10.3.4. *Estimation of variance of spatially correlated noise*

The problems dealing with possible but a priori unknown spatial correlation of noise first arise at the stage of noise variance estimation. Even for the simplest case of pure additive noise for which many methods have been designed (see subsection 10.3.1) these problems are serious for both the methods operating in spectral or spatial domain, especially for the former ones. We have already mentioned this, but would like to give here more detailed results. Note that imaging systems are commonly designed in such a manner that no more than $2 \ldots 3$ neighbor pixel values are essentially correlated and there are no far correlations.

Recall that the methods operating in spectral domain rely on certain statistics of spectral coefficient amplitudes that are supposed to have a PDF close to Gaussian for image homogeneous fragments or for high frequency part of spectral coefficients [SEN 02]. This assumption occurs to be not valid anymore if noise is not i.i.d. Due to this, blind estimates of noise variance are characterized by essential bias and this bias is negative in opposite to most practical cases of i.i.d. noise [LUK 08a].

However, we would like to stress that possible spatial correlation of noise influences accuracy of spatial domain methods as well. It results in larger aggregate and relative errors [LUK 08a]. The reason is many of these methods rely on the model of local estimates (10.6) and this model occurs to be not adequate in the

case of spatially correlated noise. Recall that these methods are based on finding a coordinate of distribution maximum. But this coordinate occurs to be shifted with respect to the true value of noise variance σ_n^2 if noise is spatially correlated and block size is too small. Fortunately, for spatial domain methods there exists a rather simple solution to apply a larger block size in order to make this shift smaller. Commonly the selection of the block size equal to 7×7 or 9×9 pixels instead of using 5×5 block size appropriate for i.i.d. noise solves the problem in the sense that the bias of the final estimate becomes appropriate, within required limits. More details concerning simulation results can be found in [LUK 08a]. If noise is not pure additive (e.g., multiplicative or mixed multiplicative and additive), the recommendation is the same – to apply 7×7 or 9×9 blocks for obtaining initial set of local estimates. This allows providing appropriate accuracy for i.i.d. and spatially correlated noise.

Consider now the improved inter-quantile method combined with pre-segmentation and discrimination map obtaining procedures more in detail. The pre-classified image of Fig.10.4a is shown in Fig.10.4b. The obtained discrimination maps for the 7×7 scanning windows are presented in Fig.10.4c . Pixels that belong to the detected homogeneous regions are shown by gray. As seen from joint analysis of images in Fig.10.4, this two-stage procedure of image processing (classification+discrimination) identifies quasi-homogeneous areas quite well. The discrimination map DM_{ij} is then used for obtaining the local estimates $\{\hat{\sigma}_{n_{lock}}^2, k = 1, \ldots, K_{res}\}_{trim}$ calculated for only such blocks positions for which $DM_{ij} = 128$, where K_{res} denotes the sample size after trimming.

(a)   (b)   (c)
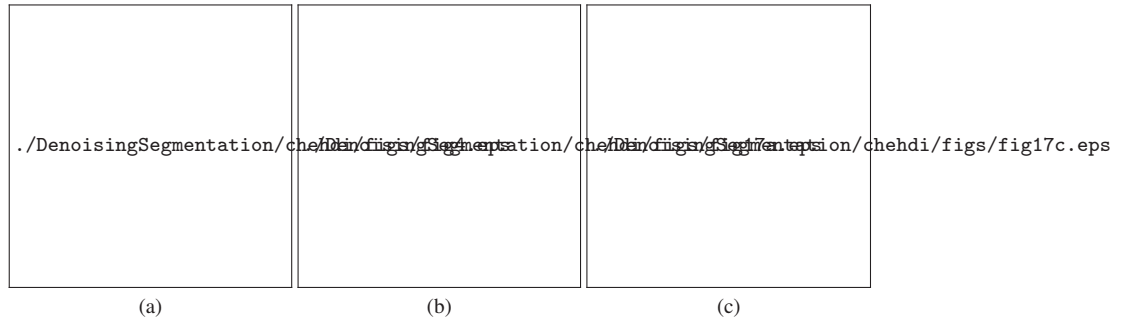
Figure 10.4. *a) A real life optical image b) Segmentation map and c) discrimination maps for the scanning windows* 7×7

The algorithm of processing the obtained estimates $\{\hat{\sigma}_{n_{lock}}^2, k = 1, \ldots, K_{res}\}_{trim}$ is the following :

1) make up sorting the estimates $\{\hat{\sigma}_{n_{lock}}^2, k = 1, \ldots, K_{res}\}_{trim}$ in ascending order with obtaining $X^{(t)}, t = 1, \ldots, K_{res}$ where $X^{(t)}$ denotes the t-th order statistic ;

Figure 10.5. *The histogram (a) of the estimates $\{\hat{\sigma}_k^2, k = 1, \ldots, K_{res}\}_{trim}$ and (b) preliminary curve $X^{(t+s)} - X^{(t)}$ via $X^{t+s/2}$ and its approximation (thick solid line), (c) final curve $X^{(t+s_{q_{opt}})} - X^{(t)}$ via $X^{t+s_{q_{opt}}/2}$ and its approximation (thick solid lines)*

2) calculate $X^{(t+s)} - X^{(t)}, t = 1, \ldots, K_{res} - s$, for $s = [0.1K_{res}]_2$ and find such t_{est}^{initI} for which $X^{(t+s)} - X^{(t)}$ is the smallest;

3) approximate the curve $Y(t) = X^{(t+s)} - X^{(t)}$ by a second order polynomial in the interval $t \in [0.1(t_{est}^{initI} + s/2), 1.9(t_{est}^{initI} + s/2)]$ and find its minimum t_{est}^{fltI} analytically; calculate $t_{est}^{I} = t_{est}^{fltI} - s/2$;

4) calculate quasi-optimal parameter $s^{q_{opt}} = [1.835t_{est}^{I} + 0.079K_{res}]_2$; if $s^{q_{opt}} > K_{res} - 1$, then $s^{q_{opt}} = [K_{res} - 1]_2$; if $s^{q_{opt}} < 0.1K_{res}$, then $s^{q_{opt}} = [0.1K_{res}]_2$;

5) repeat one time the steps 2 and 3 but for $s = s^{q_{opt}}$, determine the initial minimum index t_{est}^{fltII} and the fitted polynomial minimum index t_{est}^{fltII}, calculate $t_{est}^{II} = t_{est}^{fltII} - s^{q_{opt}}/2$;

6) obtain the final estimate $\hat{\sigma}_{fin}^2$ as $X^{t_{est}^{II}+s^{q_{opt}}/2}$.

This algorithm, in fact, obtains the preliminary estimate $\hat{\sigma}^2_{prel} = X^{t^I_{est}+s/2}$ at step 3. Then it determines the quasi-optimal parameter s^{qopt} that depends upon p in the model (10.6), and, finally, provides the final estimate after executing the steps 5 and 6. To illustrate the algorithm operation, Fig.10.5a shows the histogram of the local estimates for the 7×7 window. The curve $X^{(t+s)} - X^{(t)}$ as the function of $X^{t+s/2}$ and the result of its approximation (thick solid line) for the first stage of estimation are represented in Fig.10.5b. Respectively, Fig.10.5c shows the analogous curve $X^{(t+s^{qopt})} - X^{(t)}$ as the function of $X^{t+s^{qopt}/2}$ and the result of its second order polynomial approximation (thick solid line) for the second stage of estimation. These graphs are based on data obtained for the optical image in Fig.10.4. The left and right bounds of approximations are shown.

The preliminary estimate is equal to 10.00 and the final one is 10.01. Note that the preliminary estimate as well as the final estimate correspond well to the mode of histogram of the local estimates (see Fig.10.5a). For the 5×5 block size the obtained $\hat{\sigma}^2_{prel} = 8.89$, $\hat{\sigma}^2_{fin} = 9.11$, i.e., they are slightly smaller due to spatial correlation of noise. Final estimates conform well to the results obtained by analysis of data in the manually selected homogeneous regions (see marked regions in Fig.10.4a). Local variance estimates in these regions were the following : 10.62 ; 10.41 ; 10.82 ; 11.78 ; 11.45 ; 11.07 ; 10.01.

10.4. Conclusion and future directions

Considering the analysis of blind techniques of noise characteristic estimation, carried out in this chapter, we can conclude that there are many methods for blind evaluation of variance for pure additive and multiplicative noise ; good results for these types of noise are provided by spatial domain methods, i.e. [LUK 08a], whenever hypotheses on noise model are satisfied.

Other, more complex models of noise, e.g., mixed noise, are less intensively studied and there are few initial solutions for them. The case of signal-dependent noise needs more attention because of practical importance.

Additional studies concerning application of the obtained noise identification decisions and statistical characteristic estimates to further image processing (edge detection, filtering, compression, etc.) are worth carrying out. Thus, there is a large space for future work in the field of blind estimation of noise characteristics in image.

[ABR]   ABRAMOV S., Methods for Signal and Image Processing in Remote Sensing Systems Based on Myriad Estimate, PhD thesis, Thesis for the Degree of Candidate of Technical Science, National Aerospace University, Kharkov, Ukraine (in Russian)., 216.

[ABR 96]   ABREU E., LIGHTSTONE M., MITRA S., ARAKAWA K., « A new approach for the removal of impulse noise from highly corrupted images », *IEEE Trans. on Image Processing*, vol. 6, n°5, p. 1012–1025, 1996.

[ABR 01] ABRAMOV S., LUKIN V., ZELENSKY A., « Multiplicative Noise Variance Evaluation in MM-band Radar Images Using Myriad Estimation », *Fourth International Kharkov Symposium Physics and Engineering of Millimeter and Sub-Millimeter Waves, Kharkov, Ukraine*, vol. 1, p. 429–431, June 2001.

[ABR 04] ABRAMOV S., LUKIN V., PONOMARENKO N., EGIAZARIAN K., POGREBNYAK O., « Influence of multiplicative noise variance evaluation accuracy on MM-band SLAR image filtering efficiency », *Fifth International Kharkov Symposium Physics and Engineering of Millimeter and Sub-Millimeter Waves, Kharkov, Ukraine*, vol. 1, p. 250–252, June 2004.

[AIA 99] AIAZZI B., ALPARONE L., BARONTI S., « Reliably Estimating the Speckle Noise from SAR Data », *IGARSS*, p. 1546–1548, 1999.

[AIA 06] AIAZZI B., ALPARONE L., BARONTI S., GARZELLI A., SELVA M., « MTF-tailored multiscale fusion of high resolution MS and Pan imagery », *Photogrammetric Engineering and Remote Sensing*, vol. 72, n°5, p. 591–596, May 2006.

[AIS 08] AISENBERG N., PALIY D., ZURADA J. M., ASTOLA J. T., « Blur Identification by Multilayer Neural Network Based on Multivalued Neurons », *IEEE Transactions on Neural Networks*, vol. 19, n°5, p. 883–898, 2008.

[AST 97] ASTOLA J., KUOSMANEN P., *Fundamentals of nonlinear digital filtering*, CRC Press, 1997.

[AUB 02a] AUBERT G., BLANC-FÉRAUD L., MARCH R., Γ-convergence of discrete functionals with non-convex perturbation for image classification, Rapport n°4560, septembre 2002.

[AUB 02b] AUBERT G., KORNPOBST P., *Mathematical problems in image processing (Partial differential equations and the calculus of variations)*, Springer, 2002.

[A.V 03] A.V. SOKOLOV V. L., « Local Recognition of Radar Images by Three-Layer Neural Network Using Spatial Features », *Telecommunications and Radioengineering, Begell House, New-York*, vol. 60, n°1-2, p. 141–150, 2003.

[AYS 07] AYSAL T., BARNER K., « Meridian Filtering for Robust Signal Processing », *IEEE Transactions on Signal Processing*, vol. 55, n°8, p. 3949–3962, Dec. 2007.

[BAR 03] BARNER K., EDS G. A., *Nonlinear Signal and Image Processing : Theory, Methods, and Applications*, CRC Press, 2003.

[BAR 05] BARDUCCI F., GUZZI D., MARCOINNI P., PIPPI I., « CHRIS-Proba performance evaluation : signal-to-noise ratio, instrument efficiency and data quality from acquisitions over San Rossore (Italy) test site », *3rd ESA CHRIS/Proba Workshop, Italy*, March 2005.

[CAR 02] CARTON-VANDECANDELAERE M.-P., VOZEL B., KLAINE L., CHEHDI K., « Application to Multispectral Images of a Blind Identification System for Blur, Additive, Multiplicative and Impulse Noises », *EUSIPCO , Toulouse, France*, vol. III, p. 283–286, 2002.

[CHA 04] CHAN R., HU C., NIKOLOVA M., « An iterative procedure for removing random-valued impulse noise », *Signal Processing Letters, IEEE*, vol. 11, n°12, p. 921-924, Dec. 2004.

[CHE 00a] Chehdi K., B.Vozel, Carton-Vandecandelaere M., Kermad C., « A blind system to identify and filter degradations », *ICSP, Beijing, China*, vol. III, p. 1987–1993, 2000.

[CHE 00b] Chen T., Wu H.-R., « Impulse noise removal by multi-state median filtering », *ICASSP*, vol. 6, p. 2183–2186, 2000.

[DAT 94] Datcu M., Seidel K., « Fractal and Multiresolution Techniques for the Understanding of Geo-Information », *JRC-EARSeL, the Expert Meeting Fractals in Geoscience and Remote Sensing, Ispra, Italy*, 1994.

[DEE 03] Deergha R., Swamy M., Plotkin E., « Adaptive filtering approaches for colour image and video restoration », *IEE Proceedings on Vision, Image and Signal Processing*, vol. 150, n°3, p. 168–177, 2003.

[DIN 87] Ding R., Venetsanopoulos A., « Generalized homomorphic and adaptive order statistic filters for the removal of impulsive and signal-dependent noise », *IEEE Trans. on Circuits and Systems*, , n°8, p. 948–955, 1987.

[DOL 99a] Dolia A., Burian A., Lukin V., Rusu C., Kurekin A., Zelensky A., « Neural network application to primary local recognition and nonlinear adaptive filtering of images », 6-*th IEEE International Conference on Electronics, Circuits and Systems, Pafos, Cyprus*, vol. II, p. 847–850, 1999.

[DOL 99b] Dolia A., Kurekin A., Lukin V., Zelensky A., « Image local recognition via fragment histogram processing using neural network », *International Symposium on Signals, Circuits and Systems, Iasi, Romania*, vol. 2662, p. 167–170, 1999.

[DOL 01] Dolia A., Lukin V., Zelensky A., Anagnastopoulos C., Astola J., « Neural Networks for Local Recognition of Images with Mixed Noise », *IST/SPIE International Conference on Artificial Neural Networks in Image Processing VI, San Jose, CA, USA*, vol. 4305, p. 108–118, 2001.

[ED. 91] Ed. A. K., *Digital Image Restoration*, vol. 23, Springer Verlag, 1991.

[EFR 98] Efron B., Tibshirani R., *An Introduction to the Bootstrap*, CRC Press LLC, 1998.

[EGI 01] Egiazarian K., Melnik V., Lukin V., Astola J., « Local Transform-based Denoising for Radar Image Processing », *IST/SPIE International Conference on Nonlinear Image Processing and Pattern Analysis XII, San Jose, USA*, vol. 4304, p. 170–178, 2001.

[ENG 01] Eng H.-L., Ma K.-K., « Noise adaptive soft switching median filter », *IEEE Trans. on Image Processing*, vol. 10, n°2, p. 242–251, 2001.

[FAL 03] Falconer K., *Fractal Geometry : Mathematical Foundations and Applications*, John Wiley and Sons Ltd., 2003.

[FJO 98] Fjortoft R., Lopes A., Bruniquel J., Marthon P., « Optimal edge detection in SLC images with correlated speckle », vol. 1, p. 342-344 vol.1, Jul 1998.

[FOI 07] Foi A., Pointwise Shape-Adaptive DCT Image Filtering and Signal-Dependent Noise Estimation, PhD thesis, Thesis for the degree of Doctor of Technology, Tampere University of Technology, Tampere, Finland, 2007.

[GON 02] GONZALEZ J. G., ARCE G. R., « Statistically-Efficient Filtering in Impulsive Environments : Weighted Myriad Filters », *EURASIP Journal on Applied Signal Processing*, vol. 2002, n°1, p. 4–20, Dec. 2002.

[HAR 92] HARALICK R., SHAPIRO L., *Computer and Robot Vision*, Addison-Wesley Publishing Company, USA, 1992.

[HOF 01] HOFFMAN R., MARKMAN A., *Interpreting Remote Sensing Imagery : Human Factors*, CRC Press, LLC, 2001.

[HUA 04] HUANG H., LIN X., RAHARDJA S., YU R., « A method for realizing reversible type-IV discrete cosine transform (IntDCT-IV) », 7th *International Conference on Signal Processing*, vol. 1, p. 101–104, 2004.

[J.G 01] J.G. GONZALEZ G. A., « Optimality of the myriad filter in practical impulse noise environments », *IEEE Trans. on Signal Processing*, vol. 49, n°2, p. 438–441, 2001.

[KAA 07] KAARNA A., *Chapter 14 : Compression of Spectral Images*, I-Tech , Vienna, Austria, June 2007.

[KAL 96] KALMYKOV A., TSYMBAL V., MATVEEV A., GAVRILENKO A., IGOLKIN V., « The Two-Frequency Multipolarization L/VHF Airborne SAR for Subsurface Sensing », *European Conference on SAR (EUSAR'96), Konigswinter, Germany*, p. 275–280, 1996.

[KAL 99] KALLURI S., ARCE G., « Adaptive Weighted Myriad Filter Algorithms for Robust Signal Processing in α-stable Noise Environments », *IEEE Trans. on Signal Processing*, vol. 46, n°2, p. 322–334, 1999.

[KLA 05] KLAINE L., VOZEL B., CHEHDI K., « Unsupervised Variational Classification Through Image Multi-Thresholding », *13th EUSIPCO Conference, Antalya, Turkey*, 2005.

[KON 95] KONG H., GUAN L., « Detection and removal of impulse noise by a neural network guided adaptive median filter », *IEEE Int. Conf. on NNs*, vol. 2, p. 845–849, Dec. 1995.

[KUL 04] KULEMIN G., ZELENSKY A., ASTOLA J., LUKIN V., EGIAZARIAN K., KUREKIN A., PONOMARENKO N., ABRAMOV S., TSYMBAL O., GOROSHKO Y., TARNAVSKY Y., Methods and Algorithms for Pre-processing and Classification of Multichannel Radar Remote Sensing Images, Rapport n°TICSP Series 28 - ISBN 952-15-1293-8, Tampere, Finland, TTY Monistamo, Dec 2004.

[KUR 99] KUREKIN A., LUKIN V., ZELENSKY A., O.V. TSYMBAL G. K., ENGMAN E., « Processing multichannel radar images by modified vector sigma filter for soil erosion degree determination », *SPIE/EUROPTO Symposium on Aerosense and Remote Sensing, Florence, Italy*, vol. 3868, p. 412–423, 1999.

[LEE 81] LEE J.-S., « Speckle analysis and smoothing of synthetic aperture radar images », *Comp. Vision, Graphics, Image Processing*, vol. 17, p. 24–32, 1981.

[LEE 98] LEE K.-C., SONG H.-J., SOHN K.-H., « Noise adaptive soft switching median filter », *Electronic Letters*, vol. 34, n°5, p. 449–450, 1998.

[LEE 02] LEE C.-S., HSU C.-Y., KUO Y.-H., « Intelligent fuzzy image filter for impulse noise removal », *Int. Conf. on Fuzzy Systems*, vol. 1, p. 431–436, 2002.

[LOP 07] LOPEZ-MARTINEZ C., POTTIER E., « On the Extension of Multidimensional Spe-
ckle Noise Model from Single-Look to Multilook SAR Imagery », *IEEE Transactions on
Geoscience and Remote Sensing*, vol. 45, n°2, p. 305-320, 2007.

[L.S 02] L. SENDUR I. W. S., « Bivariate Shrinkage Functions for Wavelet Based Denoi-
sing Exploiting Interscale Dependency », *IEEE Trans. on Signal Processing*, vol. 50, n°11,
p. 2744–2756, Nov. 2002.

[LUK 96a] LUKIN V., MELNIK V., POGREBNIAK A., ZELENSKY A., ASTOLA J., SAARI-
NEN K., « Digital adaptive robust algorithms for radar image filtering », *Journal of Electro-
nic Imaging*, vol. 5, n°3, p. 410-421, 1996.

[LUK 96b] LUKIN V., PONOMARENKO N., KUOSMANEN P., ASTOLA J., « Modified Sigma
Filter for Processing Images Corrupted by Multiplicative and Impulsive Noise », *EUSIPCO,
Trieste, Italy*, vol. III, p. 1909–1912, June 1996.

[LUK 02a] LUKIN V., PONOMARENKO N., S.ABRAMOV, K.EGIAZARIAN, ASTOLA J.,
« Automated robust procedure for SAR image filtering and preliminary analysis », *Inter-
national Symp. On Satellite Remote Sensing, Crete (Greece)*, vol. 4883, p. 48–59, Sept
2002.

[LUK 02b] LUKIN V., ABRAMOV S., ZELENSKY A., ASTOLA J., « Blind evaluation of noise
variance in images using myriad operation », *IST/SPIE International Conference on Image
Processing : Algorithms and Systems, San Jose (CA, USA)*, vol. 4667, p. 192-203, 2002.

[LUK 02c] LUKIN V., PONOMARENKO N., ABRAMOV S., ZELENSKY A., EGIAZARIAN K.,
ASTOLA J., « Automatic Robust Procedure for Radar Image Preliminary Analysis and Fil-
tering », *Telecommunication and Radioengineering, Begell House (NY)*, vol. 58, n°5-6,
p. 66–87, 2002.

[LUK 05a] LUKIN V., ABRAMOV S., B.VOZEL, K.CHEHDI, « A method for blind automa-
tic evaluation of noise variance in images based on bootstrap and myriad operations »,
SPIE/EUROPTO Symposium On Satellite Remote Sensing, Bruges, Belgium, vol. 5982,
p. 299–310, 2005.

[LUK 05b] LUKIN V., ABRAMOV S., ZELENSKY A., ASTOLA J., « Myriad based shift pa-
rameter estimation method and its application to image filtering and processing », *SPIE
Conference Mathematical Methods in Pattern and Image Analysis, San Diego, USA*,
vol. 5916, p. 1-12, 2005.

[LUK 05c] LUKIN V., KOIVISTO P., PONOMARENKO N., ABRAMOV S., ASTOLA J., « Two-
stage Methods for Mixed Noise Removal », *EURASIP Workshop on Nonlinear Signal and
Image Processing (NSIP), Japan*, 2005.

[LUK 05d] LUKIN V., KOIVISTO P., PONOMARENKO N., ABRAMOV S., ASTOLA J., « Two-
stage Methods for Mixed Noise Removal », *EURASIP Workshop on Nonlinear Signal and
Image Processing (NSIP), Japan*, 2005.

[LUK 06a] LUKIN V., ABRAMOV S., PONOMARENKO N., VOZEL B., CHEHDI K., « Me-
thods for blind evaluation of noise variance in multichannel optical and radar images »,
Telecommunication and Radioengineering, Begell House (NY), vol. 65, n°6, p. 509-537,
2006.

[LUK 06b] LUKIN V., ABRAMOV S., ZELENSKY A., ASTOLA J., « Use of minimal inter-quantile distance estimation in image processing », *IX SPIE Conference on Mathematics of Data/Image Pattern Recognition, Compression, and Encryption with Applications, San Diego, USA*, vol. 6315, p. 299–310, 2006.

[LUK 06c] LUKIN V., TSYMBAL O., VOZEL B., CHEHDI K., « Processing multichannel radar images by modified vector sigma filter for edge detection enhancement », *ICASSP, 2006*, p. 833–836, 2006.

[LUK 07a] LUKIN V., PONOMARENKO N., ABRAMOV S., VOZEL B., CHEHDI K., « Improved noise parameter estimation and filtering of MM-band SLAR images », *Sixth International Kharkov Symposium Physics and Engineering of Millimeter and Sub-Millimeter Waves, Kharkov, Ukraine*, vol. 1, p. 439–441, 2007.

[LUK 07b] LUKIN V., ABRAMOV S., VOZEL B., CHEHDI K., « Improved minimal inter-quantile distance method for blind estimation of noise variance in images », *SPIE/EUROPTO Symp. on Satellite Remote Sensing, Florence, Italy*, Sept. 2007.

[LUK 07c] LUKIN V., OKTEM R., PONOMARENKO N., EGIAZARIAN K., « Image filtering based on discrete cosine transform », *Telecommunications and Radio Engineering*, vol. 66, n°18, p. 1685–1701, 2007.

[LUK 08a] LUKIN V., ABRAMOV S., VOZEL B., CHEHDI K., ASTOLA J., « Segmentation-based method for blind evaluation of noise variance in images », *SPIE Journal on Applied Remote Sensing*, p. 472–475, Aug. 2008.

[LUK 08b] LUKIN V., PONOMARENKO N., ABRAMOV S., VOZEL B., CHEHDI K., ASTOLA J., « Filtering of radar images based on blind evaluation of noise characteristics », *SPIE Proceedings of Satellite Remote Sensing, Cardiff, UK*, 2008.

[LUK 09] LUKIN V., ABRAMOV S., PONOMARENKO N., VOZEL B., CHEHDI K., « Joint estimation of remote sensing images and mixed noise parameters, Telecommunications and Radio Engineering », *Telecommunication and Radioengineering, Begell House (NY),*, to appear in 2009.

[MAL 98] MALLAT S., *A Wavelet tour of signal processing*, Academic Press, San Diego, 1998.

[MEL 00] MELNIK V., Nonlinear Locally Adaptive Techniques for Image Filtering and Restoration in Mixed Noise Environments, PhD thesis, Thesis for the degree of Doctor of Technology, Tampere University of Technology, Tampere, Finland, 2000, 234 p.

[MEL 01] MELNIK V., LUKIN V., ZELENSKY A., ASTOLA J., KUOSMANEN P., « Local Activity Indicators : Analysis and Application to Hard-Switching Adaptive Filtering of Images », *Optical Engineering Journal*, , n°8, p. 1441–1455, 2001.

[MOT 04] MOTWANI M., GADIYA M., MOTWANI R., HARRIS F. C., « Survey of Image Denoising Techniques », *GSP, Santa Clara, CA*, p. 472–475, Sept. 2004.

[O. 05] O. TSYMBAL, Multistage Robust Adaptive Filtering of Multichannel Remote Sensing Images, PhD thesis, Thesis for the degree of Doctor of Technology, Tampere University of Technology, Tampere, Finland, 2005.

[OKT 07] OKTEM R., EGIAZARIAN K., LUKIN V., PONOMARENKO N., TSYMBAL O., « Lo-
cally adaptive DCT filtering for signal-dependent noise removal », *EURASIP Journal on
Advances in Signal Processing*, , n°Article ID 42472, 2007.

[OLI 04] OLIVER C., QUEGAN S., *Understanding Synthetic Aperture Radar Images.*, SciTech
Publishing, 2004.

[PAL 07] PALIY D., Local Approximations in Demosaicing and Deblurring of Digital Sensor
Data, PhD thesis, Thesis for the degree of Doctor of Technology, Tampere University of
Technology, Tampere, Finland, Dec. 2007.

[PEN 84] PENTLAND P., « Fractal-Based Description of Natural Scenes », *IEEE Transactions
on PAMI*, vol. 6, n°6, p. 661–675, 1984.

[PEN 03] PENG Q., RUN T.-D., « Binarized switching median filter with pre-segmentation for
impulse noise removal », *IEEE Proceedings on Intelligent Transportation Systems*, vol. 1,
p. 69–73, 2003.

[PIT 90] PITAS I., VENETSANOPOULOS A., *Nonlinear Digital Filters : Principles and Appli-
cations*, Kluwer Academic Publisher, Boston, USA, 1990.

[PLA 00] PLATANIOTIS K., VENETSANOPOULOS A., *Color Image Processing and Applica-
tions*, Springer-Verlag, 2000.

[PON 03] PONOMARENKO N., LUKIN V., ABRAMOV S., EGIAZARIAN K., ASTOLA J.,
« Blind evaluation of additive noise variance in textured images by nonlinear processing
of block DCT coefficients », *IST/SPIE International Conference Image Processing : Algo-
rithms and Systems II, Santa Clara, USA*, vol. 5014, p. 178–189, 2003.

[PON 05] PONOMARENKO N., LUKIN V., ZELENSKY A., EGIAZARIAN K., ASTOLA J.,
« Locally adaptive image filtering based on learning with clustering », *SPIE Conference
Image Processing : Algorithms and Systems IV, San Jose, USA*, vol. 5672, p. 94–105, 2005.

[PON 06a] PONOMARENKO N., LUKIN V., EGIAZARIAN K., ASTOLA J., VOZEL B.,
CHEHDI K., « Hybrid Sigma Filter for Processing Images Corrupted by Multiplicative
Noise », *8th International Conference Advanced Concepts for Intelligent Vision Systems
(ACIVS), Springer, LNCS No 4179, Antwerp, Belgium*, p. 46–54, 2006.

[PON 06b] PONOMARENKO N., LUKIN V., ZRIAKHOV M., EGIAZARIAN K., ASTOLA J.,
« Estimation of accesible quality in noisy image compression », *EUSIPCO, Florence, Italy*,
vol. 6365, Sept 2006.

[PON 07] PONOMARENKO N., LUKIN V., ZRIAKHOV M., KAARNA A., ASTOLA J., « An
automatic approach to lossy compression of AVIRIS images », *IGARSS, Barcelona, Spain*,
vol. 6365, p. 472–475, Jul. 2007.

[PON 08] PONOMARENKO N., LUKIN V., ZELENSKY A., ASTOLA J., EGIAZARIAN K.,
« Adaptive DCT-based filtering of images corrupted by spatially correlated noise », *SPIE,
San Jose, USA*, vol. 6812, 2008.

[POR 03] PORTILLA J., STRELA V., WAINWRIGHT M., SIMONCELLI E. P., « Image De-
noising Using Gaussian Scale Mixtures in the Wavelet Domain », *IEEE Trans on Image
Processing*, vol. 12, n°11, p. 1338–1351, 2003.

[QJI 99] Qjidaa H., Radouane L., « Robust line fitting in a noisy image by the method of moments », *IEEE Trans. on Pattern Analysis and Machine Intelligence*, , n°11, p. 1216–1223, 1999.

[RAD 06] Radford D., Kurekin A., Marshall D., Lever K., « A New DCT-based Multiresolutiion Method for Simultaneous Denoising and Fusion of SAR Images », *Fusion*, 2006.

[RAM 99] Ramponi G., D'Alvise R., « Automatic Estimation of the Noise Variance in SAR Images for Use in Speckle Filtering », *IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing, Antalya, Turkey*, vol. 2, p. 835–838, 1999.

[RAN 99] Rank K., Lendl M., Unbehauen R., « Estimation of Image Noise Variance », *IEE Proceedings on Vision, Image and Signal Processing*, vol. 146, n°2, p. 80–84, 1999.

[ROS 97] Rosen E., « Mobile telemedicine arrives », *Telemedicine Today*, vol. 40, n°10, p. 14–42, May 1997.

[SAM 00] Samson C., Blanc-Féraud L., Aubert G., Zerubia J., « A variational model for image classification and restoration », vol. 22, n°5, p. 460–472, mai 2000.

[SCH 97] Schowengerdt R., *Remote Sensing*, Academic Press, 2-nd edition, 522 p., 1997.

[SEN 02] Sendur L., Selesnick I., « Bivariate shrinkage with local variance estimation », *IEEE Signal Processing Letters*, vol. 9, n°12, p. 438–441, 2002.

[SUO 95] Suoranta R., Amplitude domain approach to digital filtering. Theory and applications of L-filters., PhD thesis, Espoo, Technical research centre of Finland, VTT, 1995, 234 p.

[TSY 04] Tsymbal O., Lukin V., Zelensky A., Astola J., « Implementation aspects of multistage adaptive filtering of MM-band SLAR images », *Fifth International Kharkov Symposium Physics and Engineering of Millimeter and Sub-Millimeter Waves, Kharkov, Ukraine*, vol. 1, p. 262–264, June 2004.

[TSY 05] Tsymbal O., Lukin V., Ponomarenko N., Zelensky A., Egiazarian K., Astola J., « Three-state Locally Adaptive Texture Preserving Filter for Radar and Optical Image Processing », *EURASIP Journal on Applied Signal Processing*, vol. 40, n°8, p. 1185–1204, May 2005.

[USS 06] Uss M., Lukin V., Baryshev I., Vozel B., Chehdi K., « Joint Estimation of Additive and Impulsive Noise Parameters in Remote sensing Images with Fractal Structure », *International Conference Modern Problems of Radioengineering, Telecommunications and Computer Science (TCSET), Lviv-Slavsko, Ukraine*, p. 232–235, 2006.

[USS 07] Uss M., Lukin V., Abramov S., Vozel B., Chehdi K., « Joint Estimation of Multiplicative and Impulsive Noise Parameters in Remote Sensing images with Fractal Structure », *ICASSP, Honolulu, USA*, vol. 1, p. 581–584, 2007.

[VOZ 06a] Vozel B., Chehdi K., Klaine L., « Noise identification and estimation of its statistical parameters by using unsupervised variational classification », *ICASSP*, vol. II, p. 841–844, 2006.

[VOZ 06b] VOZEL B., CHEHDI K., KLAINE L., LUKIN V., ABRAMOV S., « Noise identification and estimation of its statistical parameters by using unsupervized variational classification », *ICASSP, Toulouse, France*, vol. II, p. 841–844, 2006.

[V.V 96] V.V. LUKIN N.N. PONOMARENKO J. A. K. S., « Algorithms of Image Nonlinear Adaptive Filtering Using Fragment Recognition by Expert System », *SPIE Conference Image Processing : Algorithms and Systems IV, San Jose, USA*, vol. 2662, p. 179–190, 1996.

[YAR 03] YAROSLAVSKY L., *Digital Holography and Digital Image Processing*, Kluwer Academic Publishers, Boston, 2003.

[ZEL 96] ZELENSKY A., MELNIK V., ET AL V. L., « Airborne Multichannel Remote Sensing Data Processing Techniques and Software », *ERIM, San Francisco, USA*, vol. 3, p. 151–160, 1996.

[ZER 95] ZERVAKIS M., KATSAGGELOS A., « A class of robust entropic functionals for image restoration », *IEEE Trans. on Image Processing*, vol. 4, n°6, p. 752–773, 1995.

[ZHA 00] ZHANG D., SHI Z., WANG H., KOURI D., HOFFMAN D., « Nonlinear filtering impulse noise removal from corrupted images », *Int. Conf. on Image Processing*, vol. 3, p. 285–287, 2000.

Chapitre 11

Multivariate mathematical morphology applied to colour image analysis

11.1. Introduction

The mathematical morphology (MM) theory, founded by G. Matheron [MAT 75] and J. Serra [SER 82], is a powerful image analysis framework, nowadays fully developed for both binary and grey-level images. Its popularity in the image processing community is mainly due to its rigorous mathematical foundation as well as its inherent ability to exploit the spatial relationships of pixels. The morphological framework provides a rich set of tools able to perform from the simplest to the most demanding tasks : noise reduction, edge detection, segmentation, texture and shape analysis, etc. As a methodology, it has been applied to almost all application areas dealing with digital image processing [SOI 03]. Consequently, it was only a matter of time before attempting to extend the same concepts to colour and more generally to multivalued images, i.e. images with more than one channels.

Unfortunately, this extension is not straightforward. Specifically, the morphological framework is based on complete lattices [HEI 94], thus in order to accommodate multivalued images, a way of calculating the extrema of vector data is essential. Yet unlike scalars, there is no unambiguous way of ordering vectors. Even if most of the known vector ordering schemes have already been employed for defining multivariate morphological operators, none of them has yet been widely accepted.

The lexicographical ordering is certainly the most commonly adopted choice due to its positive theoretical properties, totality and anti-symmetry, in combination with its inherent capacity for flexible ordering configurations. However this ordering comes with a severe drawback which is the level of priority attributed to the first vector dimension during lexicographical comparison, hinders highly the effective exploitation of the subsequent image channels.

The existence of balanced lexicographical orderings provides new ways to build theoretically sound and practically relevant multivariate morphological operators. Such operators can then be involved in any mutivariate image processing task, e.g. filtering, segmentation, classification, etc.

Chapter organisation

This chapter initially provides an overview of the theoretical background of multivariate morphology and of the main available colour spaces (Sec. 11.2), and then continues to elaborate on the existing approaches used for extending the fundamental operators of MM to multivariate data (Sec. 11.3) ; and that is why cases limited only to specific operators have been omitted (e.g. colour image segmentation). As one of our primary purposes is to present the state-of-the-art in colour morphology, we focus particularly on cases developed with colour images in mind. Nevertheless, the scope of the survey is not limited to colour specific vector ordering approaches, since theoretically an ordering is independent from the represented context (e.g. a colour, a multispectral pixel value, etc). That is why morphological solutions that have been developed for non-colour data, but can however be trivially applied to this case, are also included. Due to their rich diversity and overwhelming number they are examined according to their methodology of ordering their multivalued input.

Moreover, we then further deal with the problematic aspect of lexicographical ordering (Sec. 11.4), for which we will review in larger detail the existing solutions as well as their practical limits. In particular, new ways of modifying lexicographical ordering for increased symmetry among its components will be presented, including an approach leading to pseudo-morphological operators.

Finally we proceed to an application oriented comparative study of the major approaches (Sec. 11.5). We conduct tests in colour image analysis and more precisely with the aim of texture classification. The goal here is to both compare the performance of different vector ordering schemes as basis for morphological operators, and assert their theoretical properties with the help of experimental results.

11.2. Extension of mathematical morphology to multivariate data

In this section, we recall briefly the theoretical concepts behind the extension of morphological operators to multivariate images. For an in-depth study of the theory behind multivariate mathematical morphology the reader can refer to Refs. [GOU 95, SER 93]. A brief introduction to main colour spaces is also given.

11.2.1. *Orderings and lattices*

As the concept of order plays a central role in this field, we start by recalling the relative definitions. A binary relation $\mathcal{R}$ on a set $\mathcal{S}$ is called :

– *reflexive* if $x \mathcal{R} x, \ \forall x \in \mathcal{S}$

– *anti-symmetric* if $x \mathcal{R} y$ and $y \mathcal{R} x \Rightarrow x = y, \ \forall x, y \in \mathcal{S}$

– *transitive* if $x \mathcal{R} y$ and $y \mathcal{R} w \Rightarrow x \mathcal{R} w, \ \forall x, y, w \in \mathcal{S}$

– *total* if $x \mathcal{R} y$ or $y \mathcal{R} x, \ \forall x, y \in \mathcal{S}$

A binary relation $\leq$ that is reflexive and transitive is called a *pre-ordering* (or *quasi-ordering*) ; if the anti-symmetry constraint is also met, it becomes an *ordering*. If additionally the totality statement holds for $\leq$, it is denoted as *total*, if not *partial*.

The complete lattice theory is widely accepted as the appropriate algebraic basis for mathematical morphology [RON 91]. Besides unifying the approaches previously employed in binary and grey-level morphology, it also makes it possible to generalise the fundamental concepts of morphological operators to a wider variety of image types and situations.

Specifically, a *complete lattice* $\mathcal{L}$ is a non empty set equipped with a partial ordering $\leq$, such that every non-empty subset $\mathcal{P}$ of $\mathcal{L}$ has a greatest lower bound $\bigwedge \mathcal{P}$, called *infimum*, and a least upper bound $\bigvee \mathcal{P}$, called *supremum*. In this context, images are modelled by functions mapping their domain space $\mathcal{E}$, an arbitrary non empty set that is an abelian group with respect to $+$, into a complete lattice $\mathcal{T}$ (with $\top$ and $\bot$ respectively its greatest and least elements), defining the set of possible "grey values". Moreover, if $\mathcal{F}$ represents the set of functions $f : \mathcal{E} \to \mathcal{T}$, then for the partial ordering :

$$f, g : \mathcal{E} \to \mathcal{T}, \ f \leq g \Leftrightarrow \forall x \in \mathcal{E}, \ f(x) \leq g(x) \tag{11.1}$$

$\mathcal{F}$ also forms a complete lattice, where "$f(x) \leq g(x)$" refers to the partial ordering in $\mathcal{T}$. In other words a complete lattice structure is imposed on the pixel intensity range. Usually $\mathcal{E}$ (the space of pixels) is taken to be either $\mathbb{R}^d$ (d-dimensional Euclidean space) or $\mathbb{Z}^d$ (d-dimensional discrete space), hence $\mathcal{F}$ corresponds respectively to the set of continuous or discrete images. Likewise various choices are available for $\mathcal{T}$, such as $\mathcal{T} = \overline{\mathbb{R}}^n$ ($\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty, -\infty\}$) and $\mathcal{T} = \overline{\mathbb{Z}}^n$. The case of $n > 1$

corresponds to the so-called multivalued images [GOU 00], while $n = 3$ includes specifically colour images. Namely, in the case of a multivalued image with n components, $\mathcal{T} = \mathcal{T}_1 \times \ldots \times \mathcal{T}_n$ is considered as the cartesian product of n complete lattices, and each mapping $f_i : \mathcal{E} \rightarrow \mathcal{T}_i, \ i \in \{1, \ldots, n\}$ is called a *channel* or *band* of the multivalued image.

Within this model, morphological operators are represented by mappings between complete lattices (i.e. the input and output images) with some additional properties such as increasingness and translation invariance. They are employed in combination with matching patterns, called *structuring elements* (SE), that are usually subsets of $\mathcal{E}$ (i.e. *flat* SE). Particularly, *erosion* and *dilation* constitute the fundamental blocks of MM, from the combinations of which several sophisticated operators can be derived. More precisely, given two complete lattices $\mathcal{L}$ and $\mathcal{M}$, from an algebraic point of view, an operator $\varepsilon : \mathcal{L} \rightarrow \mathcal{M}$ is called an erosion, if it is distributive over infima, i.e. $\varepsilon(\bigwedge_i P_i) = \bigwedge_i \varepsilon(P_i)$ for every collection $\{P_i\}$ of elements of $\mathcal{L}$. Similarly, $\delta : \mathcal{L} \rightarrow \mathcal{M}$ is called a dilation, if it is distributive over suprema, i.e. $\delta(\bigvee_i P_i) = \bigvee_i \delta(P_i)$ for every collection $\{P_i\}$ of elements of $\mathcal{L}$. As suggested in Ref. [SER 82], dilation and erosion basically rely on three concepts : a ranking scheme, the extrema derived from this ranking and finally the possibility of admitting an infinity of operands. Yet, the first two are missing from multivalued images.

For example, if we apply the preceding notions to the case of continuous multidimensional grey-level images ($f : \mathbb{R}^d \rightarrow \overline{\mathbb{R}}$), it suffices to replace the partial ordering $\leq$ of $\mathcal{T}$ with the usual comparison operator in $\overline{\mathbb{R}}$, in order to induce a complete lattice structure on $\mathcal{T}$ and subsequently on $\mathcal{F}$ by means of Eq. (11.1), which will make the computation of extrema possible during erosion and dilation. Likewise, the inclusion operator "$\subseteq$" can be used with binary images ($f : \mathbb{R}^d \rightarrow \{0, 1\}$). However, if we now consider multivalued images ($f : \mathbb{R}^d \rightarrow \overline{\mathbb{R}}^n, \ n > 1$), it becomes problematic to find an ordering relation for the vectors of $\overline{\mathbb{R}}^n$, due to the fact that there is no universal method for ordering multivariate data.

In order to remedy this inconvenience, Goutsias et al. [GOU 95] proposed to employ an adequate surjective mapping h to transform the image data into a more "suitable" space for morphological operators. More precisely, the idea of using a surjective mapping $h : \mathcal{T} \rightarrow \mathcal{L}$, where $\mathcal{T}$ is a non empty set and $\mathcal{L}$ a complete lattice, constitutes the theoretical support upon which several of the present multivariate morphological frameworks are based. Specifically, its importance lies in the fact that $\mathcal{T}$ is no longer required to be a complete lattice, since the ordering of $\mathcal{L}$ can be induced upon $\mathcal{T}$ by means of h :

$$\forall \, t, t' \in \mathcal{T}, \ \ t \leq_h t' \Leftrightarrow h(t) \leq h(t') \tag{11.2}$$

hence making it possible to construct h-morphological operators on $\mathcal{T}$. Consequently, one can deal with multivalued images, $f : \mathcal{E} \rightarrow \mathbb{R}^n$, through the use of a well chosen mapping $h : \mathbb{R}^n \rightarrow \mathcal{L}$, where $\mathcal{L}$ is a new, more suitable space for lattice based operations [GOU 00, GOU 95].

Besides, given an adequate vector ranking scheme, the vector erosion (ε_b) and dilation (δ_b) of a multivalued image $\mathbf{f}$ by a flat SE b, can be expressed immediately by means of the vector extrema operators $\sup_v$ and $\inf_v$ based on the given ordering :

$$\varepsilon_b(\mathbf{f})(\mathbf{x}) \quad = \quad \inf_{\mathbf{s}\,\in\,b} \{\mathbf{f}(\mathbf{x}+\mathbf{s})\} \tag{11.3}$$

$$\delta_b(\mathbf{f})(\mathbf{x}) \quad = \quad \sup_{\mathbf{s}\,\in\,b} \{\mathbf{f}(\mathbf{x}-\mathbf{s})\} \tag{11.4}$$

Therefore, the main obstacle preventing the extension of morphological operators to multivalued images, consists in defining an ordering relation that will induce a complete lattice structure on the set of vector pixel intensities.

11.2.2. *Colour spaces*

A rich number of colour specification and visualisation methods, or *colour spaces* are nowadays available. While some were designed and developed with a certain family of applications in mind, others were conceived as general purpose colour representations. Each color space has its own set of desirable properties as well as drawbacks for purposes of colour image processing and computer vision. Our goal is not to give here a complete overview of the principal colour spaces, but to focus on those which will be discussed and used throughout this chapter.

Red, Green, Blue

(RGB) is by far the most widely used colour space in imaging environments. It can be found in televisions, computer monitors and digital cameras among others. Theoretically, it is based on the concept of trichromaticity and on the reference colour-matching functions defined by CIE, therefore it represents colours as combinations of the three primaries. Specifically, it is considered as an *additive* colour model. In additive colour reproduction, the mixing of primary colours with the purpose of obtaining secondary ones, is realised on the basis of light emission, conversely to subtractive models where light reflection is taken as reference. Consequently, combining for instance red and green leads to yellow (R = max, G = max, B = 0), whereas red and blue gives magenta (R = max, G = 0, B = max). Black is considered as the absence of light, hence no primary colour is necessary for its production (R = G = B = 0). White on the other hand is equivalent to the presence of all primary colours in maximal amounts (R = G = B = max). Despite its widespread use however, RGB suffers from multiple inconveniences. First, with the exception of the eight cube corners, it is intuitively difficult to specify arbitrary colours by means of the RGB cube. Furthermore, the distances between points residing inside the cube are not proportional to their perceptual differences ; in other words, RGB is perceptually non-uniform. Additionally, the exact

values of the R, G and B components of a given colour depend on the spectral sensitivity functions of the acquisition device under consideration, hence rendering this space device-dependent. And finally, the three components of RGB are highly correlated, specifically, $\rho_{BR} = 0.78$, $\rho_{RG} = 0.98$ and $\rho_{GB} = 0.94$. Due to these drawbacks, RGB has been considered in general as an unsuitable colour space for digital image processing.

Perceptually uniform colour spaces

Perceptual uniformity, the property which for a change of colour value induces a perceptually proportional change, has been a long desired property of colour spaces. As this invaluable property for computing colour differences was not provided by the initial XYZ colour space, in 1976 CIE adopted both CIELAB (or L*a*b*) and CIELUV (or L*u*v*), when no consensus could be achieved over one or the other. Both of these spaces are considered almost perceptually uniform or linear (linear for small values), and derive from XYZ, so given an image produced by an acquisition device in RGB, it is necessary to be able to convert it into and from XYZ :

$$\begin{bmatrix} X & Y & Z \end{bmatrix}^T = A \begin{bmatrix} R & G & B \end{bmatrix}^T \tag{11.5}$$

where the RGB values are normalised to $[0, 1]$, and the matrix A used with this purpose depends on the reference white point corresponding to the illuminant under consideration. For instance, for the widely popular D65 or daylight illuminant, it becomes :

$$A_{D65} = \begin{bmatrix} 0.412 & 0.358 & 0.180 \\ 0.212 & 0.716 & 0.072 \\ 0.019 & 0.119 & 0.950 \end{bmatrix} \tag{11.6}$$

As both L*a*b* and L*u*v* follow the same principles, we focus solely on L*a*b* where the $L^* \in [0, 100]$ component denoting lightness, a^* the red-green and b^* the yellow-blue opposition. Thus, a^* and b^* possess negative values when the colour they represent is close to respectively green and blue, and positive values when it is closer to respectively red and yellow. Given the XYZ triplet, the corresponding L*a*b* representation may be obtained as follows :

$$L^* = 116 f\left(\frac{Y}{Y_n}\right) - 16$$

$$a^* = 500 \left[f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right) \right]$$

$$b^* = 200 \left[f\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right) \right] \tag{11.7}$$

$$f(t) = \begin{cases} t^{1/3} & \text{if } t > 0.008856 \\ 7.787t + \frac{16}{116} & \text{otherwise} \end{cases}$$

where X_n, Y_n, Z_n denote the XYZ triplet representing the white point $(R = G = B = 1)$, which can be obtained from Eq. (11.6). Consequently, L*a*b* inherits the white point reference used in the preceding XYZ expression. The capacity to compute perceptual colour differences is considered to be one of the major advantages of L*a*b*, that has been exploited in various colour image applications, requiring device-independence and/or perceptual uniformity. On the other hand, it has been also recently argued that even though perceptual uniformity is a highly desirable property for colour definition and description, its use in the context of image processing is limited, since images are not processed in terms of "perceived values" [SER 05]. Moreover, one cannot benefit from the property of device independence in cases where no a priori information is available on the acquisition conditions of the images under consideration. Additionally, the transformations to and from RGB become computationally complex with the intermediate role of XYZ, hence the practical interest of L*a*b* has been considerably limited in real-time applications.

Phenomenal colour spaces

Phenomenal or polar colour spaces are those that express colours in intuitive terms as far as human observers are concerned. Colour is here separated into the following notions :

– *hue*, which represents the dominant wavelength of a colour, described as "green","blue" etc. It is an angular, modulo 2π value $H \in [0, 2\pi] = [0°, 360°]$, with $0°$ corresponding to red.

– *saturation* ($S \in [0, 1]$), that describes the purity of a colour. A fully saturated colour appears vivid, whereas zero saturation transforms it into a shade of grey. Thus saturation can be also considered as the distance from the achromatic axis.

– *brightness*, *luminance* or *value* (*L*), that denotes the component describing the sensation associated with the amount of emitted light, or in other words the "amplitude" of the colour under consideration. Likewise to saturation, it is often normalised to $[0, 1]$.

Polar spaces in their majority are obtained through direct transformations from RGB, hence being perceptually non-uniform and device dependent. An additional disadvantage that further hinders their use with a purpose besides colour specification, is the angular nature of hue, which through modulo based operations leads to visual discontinuities. Moreover, the most widely used representatives of this category, HSV and HLS, suffer from major structural inconveniences [HAN 08] and for these reasons phenomenal spaces have long been viewed as unsuitable for colour image processing. Indeed, they lack the following properties : independence of saturation from brightness, existence of norms for the notions of brightness and saturation, inversibility of

the space back to RGB. Thus some recent original phenomenal spaces have been proposed, e.g. LSH [ANG 05] which is defined from RGB as :

$$L_{LSH} = \frac{1}{3}(\max + \text{med} + \min)$$

$$S_{LSH} = \begin{cases} \frac{3}{2}(\max - L_{LSH}) & \text{if } L_{LSH} \geq \text{med} \\ \frac{3}{2}(L_{LSH} - \min) & \text{if } L_{LSH} \leq \text{med} \end{cases} \qquad (11.8)$$

$$H_{LSH} = k\left[\lambda + \frac{1}{2} - (-1)^{\lambda}\left(\frac{\max + \min - 2\text{med}}{2S_{LSH}}\right)\right]$$

where max, med and min denote respectively the maximum, median and minimum value of the transformed RGB triplet. k is the angle unit ($\pi/3$ for radians and 42 for 256 levels), whereas λ is set as 0 if $r > g \geq b$, 1 if $g \geq r > b$, 2 if $g > b \geq r$, 3 if $b \geq g > r$, 4 if $b > r \geq g$, and 5 if $r \geq b > g$.

Let us remark that the "optimal" colour space choice clearly depends on the needs of the application under consideration. In this chapter, we will not deal further with this issue and rather focus on the ways mathematical morphology can be applied on multivariate images. When dealing with colour images. we will mainly illustrate our proposals using the LSH space [ANG 05]. Nevertheless the proposed solutions for multivariate mathematical morphology are of course not limited to this colour space and can be used with any available data representation model.

11.3. Taxonomy of vector orderings

When applied to mutivariate images, there are two general methods for morphological processing : *marginal* (or *componentwise/scalar*) and *vector*. While marginal processing consists in dealing separately with each channel of the image (thus ignoring completely the inter-channel correlation), vector processing considers all available channels globally and simultaneously but requires to adapt the existing algorithms in order to accommodate vector data. Thus a lot of effort has been put in engineering a way of ordering vectors, especially in the last few decades. We review in this section the main solutions proposed in the literature and discuss their pros and cons. For a more complete survey of this topic, the reader is refered to [APT 07].

Several taxonomies can be given for presenting vectorial orderings, the most famous one being introduced by Barnett [BAR 76] where four different groups are defined. Marginal orderings (M-orderings) correspond to univariate orderings realised on every component of the given vectors, where data is ordered along each one of its channels independently from others, hence also the name componentwise ordering. In conditional orderings (C-orderings), vectors are ordered by means of some of their marginal components, selected sequentially according to different conditions.

Whereas the components not participating in the comparison process are listed according to the position of their ranked counterparts. Hence, the ordering of the vectors is conditioned upon the particular marginal set of ranked components. Lexicographical ordering constitutes a widely known example of C-ordering employing potentially all the available components of the given vectors and will be discussed further in this chapter. Let us observe that C-orderings are most suitable for cases where one can establish a priority among the image channels. Partial orderings (P-orderings) consist of approaches that partition the given vectors into equivalence classes with respect to order, rank or extremeness. They are generally geometric in nature and account well for the inter-relations between components. Let us remark however that "partial" is an abuse of terminology (Sec. 11.2), because not only there are total pre-orderings belonging to this class, but algebraically partial orderings do not necessarily belong to this class either. That is why we will rather use the term P-ordering. Finally, reduced orderings (R-orderings) aim at first reduce vectors to scalar values and then rank vectors according to their natural scalar order. A further categorisation of R-orderings consists in classifying them as distance orderings and projection orderings. For instance, a R-ordering on $\mathbb{R}^n$ could consist in defining first a transformation $h : \mathbb{R}^n \to \mathbb{R}$, and then ordering the vectors of $\mathbb{R}^n$ with respect to the scalar order of their projection on $\mathbb{R}$ by h. According to the chosen transformation it is possible to obtain a total pre-ordering (h non-injective) or even a total ordering (h injective) [CHA 98]. An additional advantage of R-orderings lies in the fact that with an adequately chosen h, they can attribute equal priority to all components, unlike C-orderings.

In this section, we rather adopt the scheme employed by Chanussot [CHA 98] where vector orderings are classified according to their algebraic properties (anti-symmetry, totality, etc) and distinguish between partial orderings, total pre-orderings and total orderings. This choice was made with the aim of underlining the effect that these basic properties have on the end result of processing. Additionally, the vectors of the ordering relations that are mentioned in the sequel, are considered in $\mathbb{R}^3$, unless otherwise specified.

11.3.1. *Partial orderings*

The so-called *marginal* processing strategy, despite being presented usually as an alternative to *vector*, is as a matter of fact no more than just its variant, as it employs the partial ordering defined as

$$\forall\, \mathbf{v}, \mathbf{v}' \in \mathbb{R}^n, \ \ \mathbf{v} \leq \mathbf{v}' \Leftrightarrow \forall\, i \in \{1, \ldots, n\}, \ \ v_i \leq v_i'. \tag{11.9}$$

Obviously, there can be vectors that may not be comparable under this ordering relation, for instance $\mathbf{a} = [7, 2]^T$ and $\mathbf{b} = [3, 4]^T$. Nevertheless, this does not prevent the definition of valid morphological operators based on extrema computed by means of this ordering [SER 93]. Furthermore, it makes it possible to employ all tools offered

by grey-level morphology with no need for special adaptation steps. For instance, the erosion and dilation expressions given in Sec. 11.2 become equivalent to :

$$\varepsilon_b(\mathbf{f})(\mathbf{x}) \quad = \quad [\varepsilon_b(f_1)(\mathbf{x}), \dots, \varepsilon_b(f_n)(\mathbf{x})]^T \tag{11.10}$$

$$\boldsymbol{\delta}_b(\mathbf{f})(\mathbf{x}) \quad = \quad [\delta_b(f_1)(\mathbf{x}), \dots, \delta_b(f_n)(\mathbf{x})]^T \tag{11.11}$$

where ε_b and δ_b denote respectively the scalar erosion and dilation operators with a SE b. Let us notice that a generalisation of the marginal approach was given by the matrix morphology theory of Wilson [WIL 92], which was based on the work of Heijmans and Ronse [HEI 90].

Despite its implementational simplicity, marginal ordering suffers mainly from two disadvantages : not accounting for inter-channel information as well as the risk of altering the spectral composition of its input. More precisely, as each component is processed independently, any eventual correlation among them is totally ignored, hence rendering this approach unsuitable for images with highly correlated components (e.g. RGB colour images) [AST 90]. A possible solution to this problem, as proposed in Ref. [GOU 95], consists in applying a decorrelating transformation (e.g. maximum noise fraction transform (MNF), principal component analysis (PCA), discrete cosine transform, a colour space with independent components, etc) prior to ordering. Nevertheless, these decorrelating transformations also introduce an additional computational burden.

Furthermore, there is absolutely no guarantee that marginally processed vectors belong to the input image. The lack of vector preservation constitutes an undesirable effect for several applications. For example, in the case of colour image processing this would lead to the appearance of new hues (also known as *false colours*) and thus deteriorate the visual quality of the result, and in particular the colour balance and the object boundaries ; whereas the effect near the spatial edges, leads to the so-called *edge jitter*. This effect is illustrated in Fig. 11.1, where the application of a marginal median filter leads to the appearance of yellow, as the result of additive combination of red and green. According to [TAL 98], the only way to use morphological operators without generating new vectors is to impose an order on the vector space by means of an (pre-)ordering verifying the totality constraint ; besides, this is why the majority of the published articles on multivalued morphology deal with total (pre-)orderings. Yet, vector preserving approaches are in the same time limited by this property, due to the restriction imposed on their output (i.e. the output must be a vector from the input set) ; whereas the marginal approach has access to a much broader range of output values, an advantage which may become valuable during noise reduction [COM 99]. Further attempts motivated by the intuitiveness of the marginal approach have resulted in improvements, and we can cite for instance the work of Al-Otum [Al- 03] who

proposes the corrected componentwise morphological algorithm with the goal of preventing the appearance of new vectors. It consists in replacing each new vector of the output with its closest vector from the original image, chosen with the help of a Mahalanobis distance based error function.
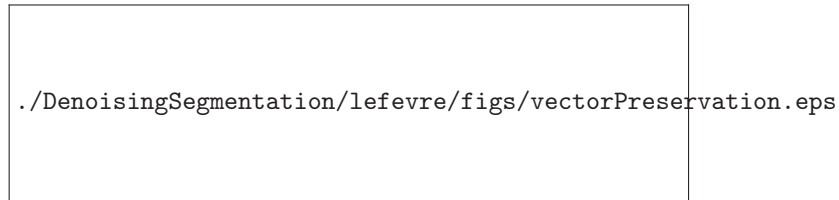
Figure 11.1. *Example of false colours. From left to right, the original image, the result of a marginal and of a vector (lexicographical) median filter.*

In brief, either accompanied by additional transformations or not, the marginal processing strategy uses conventional grey-level operators and pixels are still treated as scalar values, whereas its potential with colour data is severely limited by its lack of vector (or colour) preservation.

11.3.2. *Total pre-orderings*

Contrary to the marginal approach, by means of the additional property of totality, all vectors become comparable and as a result it is possible to construct a totally ordered lattice structure. Hence pixels can be manipulated as whole vectors, and consequently the risk of introducing new vectors is completely eliminated [TAL 98].

All pre-orderings however share a common drawback, which is the relaxation of the anti-symmetry constraint. Thus distinct vectors can eventually end up being equivalent. That is why additional measures become necessary, in order to resolve the ambiguity of eventually multiple extrema. For instance an useful solution, though partial, proposed by Comer and Delp [COM 99] for this problem, consists in selecting the output vector according to its position in the SE. Besides, the combination of an adequately chosen pre-ordering with an ordering for tie-breaking purposes has also been successfully used in practice [ANG 07].

Reduced total pre-orderings

Total pre-orderings can be obtained with R-orderings employing a non-injective reduction transformation [CHA 98]. Distance measures are typical examples of such transformations ; in fact they account for the majority of the proposed R-orderings. Colour distances however do not necessarily represent the corresponding perceptual differences, unless they are chosen in combination with a perceptually linear colour

space. That is why distance based colour orderings may lead to unexpected results, if used with non perceptually uniform spaces.

This first variant ranks vectors according to their distance from a reference vector $\mathbf{v}_{ref}$:

$$\mathbf{v} \leq \mathbf{v}' \Leftrightarrow d(\mathbf{v}, \mathbf{v}_{ref}) \leq d(\mathbf{v}', \mathbf{v}_{ref}) \qquad (11.12)$$

where $d(\cdot, \cdot)$ represents a distance measure. This particular case of distance based colour ordering has been studied in particular by Angulo [ANG 07], where the pitfall of multiple extrema is avoided by combining the distance computation with the lexicographical cascade, Eq. (11.17) :

$$\mathbf{v} \leq \mathbf{v}' \Leftrightarrow [d(\mathbf{v}, \mathbf{v}_{ref}), v_1, \ldots, v_n]^T \leq_L [d(\mathbf{v}', \mathbf{v}_{ref}), v'_1, \ldots, v'_n]^T \qquad (11.13)$$

In case the reference vector is the origin, Eq. (11.12) becomes equivalent to using the norms of the vectors. A variant of Eq. (11.12), proposed for the morphological processing of multispectral remote sensing data [PLA 04], which eliminates the need for a reference vector, consists in associating each vector with the sum of its distances from the other vectors of a family $\{\mathbf{v}_j\}$:

$$\forall \, \mathbf{v}_k, \mathbf{v}_l \in \{\mathbf{v}_j\}, \quad \mathbf{v}_k \leq \mathbf{v}_l \Leftrightarrow \sum_j d(\mathbf{v}_k, \mathbf{v}_j) \leq \sum_j d(\mathbf{v}_l, \mathbf{v}_j) \qquad (11.14)$$

This additional advantage however comes at an elevated computational cost. Furthermore, although this approach associates each vector with a scalar value denoting its "spectral purity", as far as colour images are concerned, the infimum of a family of vectors calculated in this way corresponds to the notion of "median vector" and consequently does not carry the significance of a "minimum" in the numerical sense. Moreover, if the reference vector is chosen to be "in the middle" (e.g. the median or the average), the supremum computed by means of these orderings will be very unstable. Indeed, despite the remarkable stability of the infimum under the same conditions, even a slightly varied input can radically change the least upper bound. Consequently, the dilation operator given in Eq. (11.4), that makes use of the supremum becomes unsound from a practical point of view.

The choice of the distance measure is of course another key topic. Theoretically, any kind of metric can be used, and so it is in practice, with Mahalanobis and Euclidean distances being the most common choices, but let us observe that any other non-injective transformation can be involved.

In summary, distance based R-orderings hold the potential of accounting for all dimensions without privileging any of them, a property which becomes particularly useful in case there is no predefined order of importance among the available channels (e.g. RGB colour images). Otherwise any transformation capable of realising the necessary reduction may be used, e.g. a linear weighted combination in [COM 99].

Conditional total pre-orderings

As previously mentioned C-orderings restrict the ordering process to only one or more components of the given vectors, while the others are conditioned upon them. That is why C-orderings are suitable for situations where certain channels are more "privileged" than others. Besides, unless all vector components participate in the ordering process, the resulting C-ordering is bound to be a total pre-ordering, thus sharing their aforementioned inconveniences. For example, in the case where only the first component is employed [HAR 91] :

$$\mathbf{v} \leq \mathbf{v}' \Leftrightarrow v_1 \leq v_1' \tag{11.15}$$

Hence, the two distinct vectors $\mathbf{a} = [7, 2]^T$ and $\mathbf{b} = [7, 3]^T$ would be considered equivalent according to Eq. (11.15). The main problem of C-orderings concerns of course the choice of the ordered components. Obviously, ordering vectors along only some of their components is practically justifiable only if the given components represent sufficiently the vectors.

In the general case, this approach attributes far too much significance to the selected components, disregarding all others. While usually this constitutes a severe problem, there are cases where it becomes invaluable, For instance when it is a priori known that certain channels are noise free. In fact, C-orderings are almost always used in combination with a suitable domain space change, so that the data of interest, in its majority, will lie in only some of the channels.

An example can be found in Ref. [VAR 02], where a C-ordering on the HSV colour space is employed, ignoring the hue component, hence avoiding the problem of false colours, despite the use of a non-total ordering. Specifically, given two HSV triplets $\mathbf{v} = (h, s, v)$ and $\mathbf{v}' = (h', s', v')$:

$$\mathbf{v} \leq \mathbf{v}' \Leftrightarrow \begin{cases} v \leq v', & \text{or} \\ v = v' & \text{and } s \geq s' \end{cases} \tag{11.16}$$

In fact, this also constitutes a fine example of a combined ordering ; more precisely, the result can also be considered as a P-ordering where the ordered groups contain the colours of equal value and saturation with no internal distinction whatsoever as hue is not taken into account.

11.3.3. *Total orderings*

Total orderings, from a theoretical point of view, have two main advantages that render them more suitable for vector ordering, as far as multivariate MM is concerned. First, thanks to their totality, they are vector preserving, and contrary to pre-orderings, as they verify the anti-symmetry constraint the computed extrema are unique. That is

why the majority of the attempts concentrated on extending morphological operators to multivalued images are based on total orderings. In particular, the lexicographical ordering (C-ordering) along with its variants is among the most implemented choices in colour morphology.

However, the uniqueness of extrema takes a serious toll, since the prioritisation of certain vector components becomes inevitable [CHA 98]. That is why they are almost always used in combination with a suitable domain transformation (e.g. PCA, a colour space with independent components, etc) that will place the "interesting" part of the data in the first few channels. Nevertheless, as it will be subsequently presented some implementations tend to be more "symmetric" than others.

Lexicographical ordering

Lexicographical ordering is undoubtedly the most widely employed total ordering within this context, and is defined as

$$\forall\, \mathbf{v}, \mathbf{v}' \in \mathbb{R}^n,\ \ \mathbf{v} <_L \mathbf{v}' \Leftrightarrow \exists\, i \in \{1, \ldots, n\},\ \ (\forall\, j < i, v_j = v'_j) \wedge (v_i < v'_i) \quad (11.17)$$

and we will adopt the notation $V \rightarrow W \rightarrow X$ for vectors in a 3D space VWX where we compare first dimension V, then W and finally X. As a conditional ordering, it is most suitable to situations where an order of "importance" exists on the available channels, either inherently or artificially created by means of an appropriate transformation. This prioritisation of certain vector components with respect to others, constitutes also the main drawback of lexicographical ordering and will be discussed in Sec. 11.4. To explain, Fig. 11.2 provides an example of the priority attributed to the first component during lexicographical ordering. More precisely, a vector dilation is applied on a RGB colour image (Fig. 11.2, left) and as red is the head component, it dominates visibly over green (Fig. 11.2, middle). Whereas if we permute the channels as GRB, the effect is reversed in favour of green (Fig. 11.2, right).

./DenoisingSegmentation/lefevre/figs/lexicoDifferentFirst.eps

Figure 11.2. *original image (left), results of applying a vector dilation based on a lexicographical ordering (RGB - middle) (GRB - right) with a 21×21 square SE*

As far as colour MM is concerned, the inherent separation of chromatic and achromatic information in colour spaces such as L*a*b*, HSV, and LSH, makes lexicographical ordering a prime ordering choice, since achromatic information alone is

often sufficient for the recognition of most objects. That is why, unless the application demands it, most colour MM implementations are based on brightness first lexicographical configurations. A first example on the HSV colour space can be found in [LOU 02] which however does not take into consideration the 2π periodicity of the hue component [HAN 01, APT 09]. The polar representations of the almost perceptually uniform spaces (e.g. $L^*a^*b^*$ in [HAN 02]) have also been used together with a lexicographical ordering. Besides, following the variety and number of lexicographical approaches developed for phenomenal spaces, Angulo [ANG 05] proposed a "unified framework" consisting of using the lexicographical ordering in the LSH colour space.

Bit mixing based ordering

Bit interlacing (or mixing) constitutes an innovative R-ordering, originally developed for RGB colour images, aiming mainly to eliminate the unavoidable asymmetry which results from the application of total orderings. Specifically it employs an injective transformation exploiting the binary representation of each component in order to impose a total order on the vector space [CHA 98].

Given a vector $\mathbf{v}$, with each component coded in k bits, the corresponding reduction transformation $h : \mathbb{Z}^n \to \mathbb{Z}$ is formulated as :

$$h(\mathbf{v}) = \sum_{m=1}^{k} \left\{ 2^{n \cdot (k-m)} \cdot \sum_{i=1}^{n} 2^{n-i} \cdot v_{i,m} \right\} \tag{11.18}$$

where $v_{i,m}$ denotes the mth bit of the ith component of $\mathbf{v}$. Hence, the resulting binary representation of $h(\mathbf{v})$ becomes :

$$v_{1,1}v_{2,1} \ldots v_{n,1} v_{1,2} v_{2,2} \ldots v_{n,2} \quad \ldots \quad v_{1,k}v_{2,k} \ldots v_{n,k} \tag{11.19}$$

Besides being endowed with all the qualities of a total ordering, bit mixing provides a more symmetrical approach than its lexicographical counterpart as dimensions are mixed in bit level. Of course some dimensions continue to be more privileged than others, with the degree of that privilege being proportional to the significance of the bit position that they occupy. Furthermore, a finer grained symmetry can be obtained by modifying the original mix order [CHA 98]. Then again, there are several situations where data channels need to be processed with a certain priority. Bit mixing can easily respond to this requirement by placing the important vector components to more significant bit positions.

From a theoretical point of view, this approach aims to fill a given multi-dimensional space using a "balanced" *space filling curve* (SFC) with respect to the available dimensions. In the case of total orderings, these curves (e.g. Peano curve) pass through all vector coordinates of the space under consideration, hence vectors can be ordered

according to their position on it. More details on SFC will be given in the next section. On the other hand, according to Ref. [SOI 02] the main inconvenience of a total ordering obtained in this way is its lack of physical interpretation.

Synopsis

This section presented the main approaches that have appeared so far in the literature, with the goal of extending morphological operators to multivalued images. Please note that we have omitted some rather "unconventional" orderings or extremum calculation approaches, in the sense that they either do not consist in simply using a standard vector ordering scheme, or they are developed for a particular form of image data, or even combine additional theories with the goal of achieving an efficient solution. A more complete description of multivariate mathematical morphology is given in [APT 07].

Unfortunately, each of the main solutions reviewed here comes with some drawback which are worth being recalled. Partial orderings (or marginal approaches) prevent to ensure vector preserving, thus may introduce new (and false) colours or spectral signatures in the resulting image. Pre-orderings do not guarantee extrema uniqueness, which can be a very awkward issue when dealing with morphological operations. Finally, total orderings seem the best solution from a theoretical point-of-view by avoiding these two drawbacks. However, when coming to practical use, it can be observed that they give very much importance to the first channel of an image, thus preventing to perform a real multivariate image processing. In the following section, we will deal with this problem and present existing and new solutions to ensure symmetry within the most representative total ordering, i.e. lexicographical ordering.

11.4. Balanced total orderings

Lexicographical ordering, defined in Eq. (11.17), as explained in previous section, possesses multiple properties than render it a more adequate choice than its competitors, as a way of ordering vectors in the context of multivariate MM. Specifically, the first property is its totality. Since all vectors are lexicographically comparable, according to Ref. [TAL 98] it preserves the original vectors of the input image, hence preventing the appearance of false colours. Moreover, as it satisfies the anti-symmetry constraint, a property lacking from most reduced orderings, it makes it possible to compute unique vector extrema, thus effectively avoiding ambiguities during vector ordering. Additionally, it represents an excellent choice in combination with spaces possessing an either inherent or artificially induced assymetric distribution of information ; which is asserted by the number of authors that have considered it, due to its capacity to attribute a certain amount of priority to the first vector or colour dimension. That is why it has been popular with phenomenal spaces, that separate effectively

the different notions of higher level colour perception. Besides, through the control of the channels' order during comparison (e.g. luminance first, saturation second, etc), it provides a certain degree of customisation flexibility, a notion of crucial importance as far as general purpose colour morphological operators are concerned.

However, notwithstanding its desirable properties, lexicographical ordering also suffers from a serious drawback. More precisely, the outcome of the vast majority of lexicographical comparisons, is decided based only on the first few vector components that are compared, while the contribution of its remaining dimensions is typically negligible [HAN 02]. This property is illustrated in Table 11.1, where the percentages of comparisons determined by the three channels of the three RGB colour images of Fig. 11.3, during a vector dilation based on a lexicographical ordering are shown. The channel occupying the first position of the lexicographical cascade (i.e. red) is obviously responsible for the vast majority of comparisons outcomes.
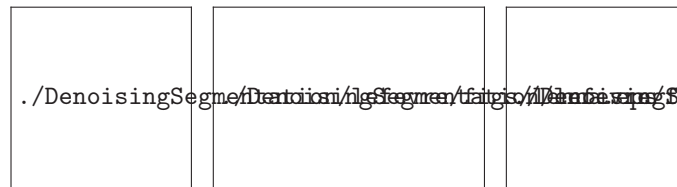
Figure 11.3. *The test images, from left to right, Lenna, Macaws and Mandrill.*

Images	Equalities (%)	Red (%)	Green (%)	Blue (%)
Lenna	0.12	93.02	6.3	0.56
Macaws	8.89	83.3	5.71	2.1
Mandrill	0.54	95.28	3.51	0.67

Tableau 11.1. *The percentages of comparisons that have led to equalities, and those that have been determined at each channel, during the dilation of the images of Fig. 11.3, with a lexicographical ordering ($R \rightarrow G \rightarrow B$) using a square shaped SE of size 5×5 pixels.*

Of course this might be a desired behaviour in cases where the fist image channel contains the majority of the total variational information, for instance after applying a PCA transform. Nevertheless, most often it leads to an insufficient exploitation of the image channels and inter-channel relations. This effect is most aggravated in the case of hyperspectral images, where despite the availability of hundreds of channels, only at most the first few participate in the overall process. That is why, variations of Eq. (11.17) have been proposed, with the purpose of better tuning the priority as well as degree of influence of each vector component on the comparison outcome.

11.4.1. *Basic solutions to limit assymmetry*

As the majority of lexicographical comparisons are determined by the first components, variants of the classical lexicographical ordering have been proposed, with the purpose of better tuning the priority as well as degree of influence of each component. A first group of variants is based on preceding the lexicographical cascade by a component representative of the entire vector. This principle is used for example along with an Euclidean norm in Refs. [RIV 04] :

$$\mathbf{v} \leq \mathbf{v}' \Leftrightarrow [\|\mathbf{v}\|, v_1, \ldots, v_n]^T \leq_L [\|\mathbf{v}'\|, v_1', \ldots, v_n']^T \tag{11.20}$$

and similarly in in Ref. [ANG 07] where a distance with respect to a reference vector is involved, Eq. (11.13). Of course, there is no limit to the number or type of functions that can be used according to this principle. Other examples include the use of the maximum and minimum of the compared components in the case of RGB colour images, as well as their weighted combinations [ANG 03]. However these extensions do not provide any way for the user to specify the prioritisation among bands.

Another type of extension to the classical lexicographical ordering consists in using a user defined parameter α, in such a way that it can modify the degree of influence of the first component. The first attempt in this context was made by Ortiz et al. [ORT 01], that proposed the α-*lexicographical ordering* :

$$\forall \mathbf{v}, \mathbf{v}' \in \mathbb{R}^n, \quad \mathbf{v} < \mathbf{v}' \Leftrightarrow \begin{cases} v_1 + \alpha < v_1', \text{ or} \\ v_1 + \alpha \geq v_1' \text{ and } [v_2, \ldots, v_n]^T <_L [v_2', \ldots, v_n']^T \end{cases} \tag{11.21}$$

where $\alpha \in \mathbb{R}^+$. The α argument is thus used to the end of increasing the occurrence of equivalences within the first vector dimension, since a scalar value v_1 becomes "equal" to all values contained in the interval $[v_1 - \alpha, v_1 + \alpha]$, hence allowing comparisons to reach more frequently the second dimension. Nevertheless, Eq. (11.21) is not transitive, and consequently does not represent an ordering from an algebraic point of view.

A theoretically sounder approach was proposed by Angulo and Serra [ANG 03] and was named α-*modulus lexicographical ordering* :

$$\forall \mathbf{v}, \mathbf{v}' \in \mathbb{Z}^n, \quad \mathbf{v} < \mathbf{v}' \Leftrightarrow [\lceil v_1/\alpha \rceil, v_2, \ldots, v_n]^T <_L [\lceil v_1'/\alpha \rceil, v_2', \ldots, v_n']^T \tag{11.22}$$

which aims to create equivalence groups within the first dimension. It relies on a quantisation through division by a constant α followed by a rounding off, which reduces the dynamic margin of the first dimension, thus allowing a greater number of comparisons to reach the second. On the contrary of the aforementioned approach, it is reflexive and transitive, hence a pre-ordering.

The effect of the operation realised on the first dimension, becomes more clear by studying the *space filling curves* (SFC), that travel through points of multi-dimensional
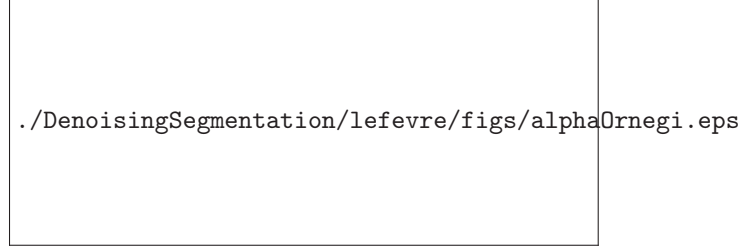
Figure 11.4. *Space filling curves in a bi-dimensional space for a lexicographical ordering (left), where the arrows denote the direction of increasing vector coordinates, and for the α-modulus lexicographical ordering with $\alpha = 4$ (right).*

space. Since the search for a total vector ordering can be formulated as a search for an injective function mapping all the points of a multi-dimensional space onto a uni-dimensional space, SFC satisfy this requirement and make it possible to model different solutions, where vectors are ordered according to the position of their coordinates on the SFC. Since lexicographical ordering corresponds to a bijection [CHA 98], its SFC will pass once on all points of a bi-dimensional discrete space $D1 \times D2 = [0, 16]^2$ as illustrated in Fig. 11.4 (left), where the high priority attributed to the first dimension $(D1)$ is represented by the high frequency of horizontal curves. Fig. 11.4 (right) shows the quantised form $D1'$ of dimension $D1$ with $\alpha = 4$, which leads to the creation of equivalence groups. For instance the points $\{5, 6, 7, 8\}$ of $D1$ now belong to the same group 2 of $D1'$, and thus given two coordinates $(5, 1)$ and $(7, 0)$ in $D1 \times D2$, the first components are considered equal in $D1' \times D2$ and the outcome of the comparison is determined by the second dimension, in other words $(5, 1) > (7, 0)$.

Nevertheless, special attention is required with the use of this approach, since the resulting equivalence groups obviously eliminate the anti-symmetry property of lexicographical ordering. One way of countering this problem, that renders this approach as well as any other similar to it anti-symmetric and thus an ordering, is to continue the lexicographical cascade with the un-quantised vector dimensions. To explain :

$$\forall \, \mathbf{v}, \mathbf{v}' \in \mathbb{Z}^n, \;\; \mathbf{v} < \mathbf{v}' \Leftrightarrow$$

$$\begin{cases} [j_1(v_1), \ldots, j_{n-1}(v_{n-1}), v_n]^T <_L [j_1(v_1'), \ldots, j_{n-1}(v_{n-1}'), v_n']^T, \;\; \text{or} \\ [j_1(v_1), \ldots, j_{n-1}(v_{n-1}), v_n]^T = [j_1(v_1'), \ldots, j_{n-1}(v_{n-1}'), v_n']^T \\ \text{and} \;\; \mathbf{v} <_L \mathbf{v}' \end{cases} \tag{11.23}$$

where $j_i(\cdot)$, $i \in [1, n-1]$ denotes some function used to reduce the dynamic margin of the ith dimension. Thus only equal vectors are considered equivalent.

In conclusion, among the solutions that have been so far reported with the purpose of fine-tuning the dimension prioritisation of lexicographical ordering, only Eq. (11.22) in combination with Eq. (11.23) satisfies the theoretical requirements of an ordering and hence leads to valid morphological operators. The practical use of α-modulus lexicographical ordering however is limited, as it relies on an implicit assumption on the dimension to be quantised. In the next section are presented some advanced quantisation models which have been proposed recently by the authors [APT 08b, APT 08c].

11.4.2. *Spectral quantisation*

As far as its theoretical properties are concerned, α-modulus lexicographical ordering appears to be the most pertinent among the two variations, as it provides an effective means of shifting priority away from the first vector dimension, while preserving the desirable characteristics of lexicographical ordering that have made it popular within the colour morphological context.

However, its practical use relies on an important implicit assumption. More precisely, as shown in Fig. 11.5a, the dimension under consideration is processed uniformly, based on the assumption that all of its subsets are equally "unimportant" with respect to the second dimension in the lexicographical cascade. In practice, often complicated relations are present among the image channels, and the need to shift priority to the next dimensions arises only partially. Consider for instance the case of colour morphology in a phenomenal colour space of type LSH, where the importance of saturation is minimised for extreme levels of luminance.

./DenoisingSegmentation/left/Deno/fsiigg/SelphenStatiitCi/lgefeepse/figs/alphaEkspCizge.eps

a)   b)
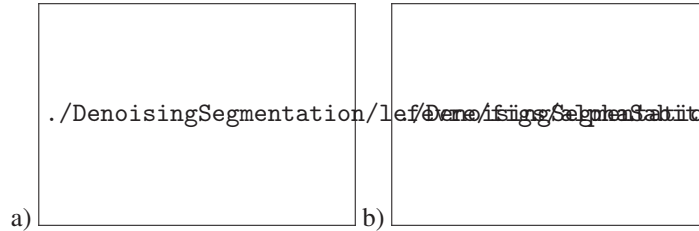
Figure 11.5. *Example of quantisation applied on a dimension D1 using* $\alpha = 10$ *and Eq.* (11.22) *(a), and using the proposed approach with the same* α *and an exponential distribution (b).*

In the light of these remarks, we propose the generalisation described in Algorithm 1, as a replacement for the division followed by rounding off, of Eq. (11.22). Based on a simple principle, this algorithm realises a quantisation of a given discrete "pixel range" interval, in practice often $I = [0, 255]$, by associating to each value, a group of equivalence, the size of which is computed using a user defined function f, and is

limited by the value of $\alpha \in \mathbb{N}^*$. Hence, it is possible to reformulate Eq. (11.22) as :

$$\forall\, \mathbf{v}, \mathbf{v}' \in \mathbb{Z}^n, \ \ \mathbf{v} < \mathbf{v}' \Leftrightarrow [w_1, v_2, \ldots, v_n]^T <_L [w_1', v_2', \ldots, v_n']^T \qquad (11.24)$$

where w_1 and w_1' represent respectively the equivalence group of v_1 and v_1', obtained through Algorithm 1. Of course one is by no means limited to applying this procedure only to the first dimension. In fact, complicated inter-channel relations often require the repeated use of such an approach in more than one dimensions in order to be effectively modelled.

Algorithm 1 The algorithm for computing a quantised discrete dimension based on an arbitrary priority distribution.

Input : $I = [a, \ldots, b] \subseteq \mathbb{Z}$, an array containing
the discrete dimension to have its dynamic margin reduced
$\alpha \in \mathbb{N}^*$, a parameter setting the maximum allowed size
of equivalence groups within I
$f : I \rightarrow [0, 1]$, a function modelling the desired priority distribution within I
Output : $J \subseteq \mathbb{N}$, the array containing the new quantised dimension
$tmp \leftarrow 0$
for $i \leftarrow a$ to b **do**
   $k \leftarrow \lceil \alpha \times f(I[i - a]) \rceil$
   for $j \leftarrow i$ to $i + k$ **do**
     $J[j - a] \leftarrow tmp$
   end for
   $tmp \leftarrow tmp + 1$
   $i \leftarrow i + k$
end for

Similar to Eq. (11.22), Eq. (11.24) also represents a relation lacking the anti-symmetry property for the same reasons. That is why we will implicitly consider the combination of Eq. (11.24) with the standard lexicographical ordering, as shown in Eq. (11.23), for tie-breaking purposes.

An example of this algorithm is given in Fig. 11.5b, where it is applied on the integer interval $[0, 100]$ using an exponential priority distribution. In other words, values close to zero are of high importance and need to be processed with a fine precision whereas the second dimension may be used with values approaching 100. The increase in the size of equivalence groups can be easily observed as the values approach the upper interval bound. Consequently, in order to obtain the uniform distribution of Fig. 11.5a, corresponding to the quantisation of α-modulus lexicographical ordering, with this approach, it is sufficient to employ the constant function $\forall n \in \mathbb{N}, \ f(n) = 1$ within Algorithm 1. By means of the function f, once can thus model arbitrary priority distributions within image channels. Furthermore, image-specific ordering approaches

may be developed using the histogram of the dimension under consideration, as shown in Fig. 11.6, hence leading to adaptive vector orderings.

./DenoisingSegmentation/lefevre/figs/adaptiveAlphaYeni.eps

Figure 11.6. *The probability distribution of dimension D1 (bottom) for an arbitrary image, general trend of the space filling curve frequency corresponding to its lexicographical ordering (A), α-modulus lexicographical ordering with α = 20 (B), and quantisation based α-lexicographical ordering using the dimension histogram as priority distribution function (C).*

11.4.3. *Spatial quantisation*

Creating artificial equivalence groups within the dimension to have its lexicographical priority reduced, is an effective means of shifting priority to the remaining vector dimensions. However, this is an operation realised independently from the images to be processed (except if the priority distribution is image specific, e.g. based on its histogram), and no a priori information is available on where the, eventually abrupt, equivalence group limits correspond within the input image. This undesirable situation is illustrated in Fig. 11.7 where the equivalence groups of a uni-dimensional discrete signal, obtained with Eq. (11.22), lead to artificial edges/value variations during comparison.

An alternative way of achieving a priority shift that avoids this inconvenience, consists in assuming an image specific approach and forming these equivalence groups based on the content of the image to be processed. Specifically, one can preprocess the input image's channel that is to have its priority reduced, so that it is "flattened", or in other words heavily smoothed. If we call this image a "marker" m, then the ordering of vectors of an image g can be realised as follows :

$$\forall x, y, s, t \in \mathbb{Z}, \ g(x,y) < g(s,t) \Leftrightarrow$$

$$[m(x,y), g_2(x,y), \ldots, g_n(x,y)]^T <_L [m(s,t), g_2(s,t), \ldots, g_n(s,t)]^T \quad (11.25)$$
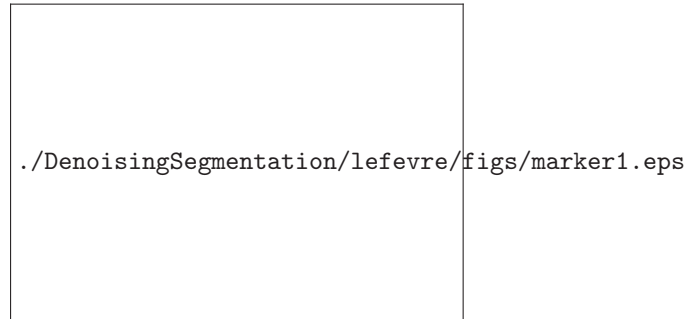
Figure 11.7. *Example of a uni-dimensional discrete signal with its corresponding equivalence groups obtained with Eq.* (11.22) *and* $\alpha = 11$.

In other words, the formation of equivalence groups is now controlled only by the marker image. If a couple of pixels have the same value in the marker image, then they are considered equal for the dimension that this marker image represents, and the comparison outcome is determined by the following dimensions. Hence equivalence groups are now formed at the flat regions of m (Fig. 11.8). The "flattening" process may be achieved using a variety of filters such as large median filters, alternating sequential filters [SOI 03], morphological levelings [MEY 04], etc. Consequently, although the creation of equivalence groups is image specific, and needs to be realised independently for each image to be processed, marker based ordering provides nonetheless a means of avoiding the artificial edge related pitfalls of the aforementioned approaches, by taking into account the spatial relationships of pixels, thus making it possible to accommodate topological restrictions during ordering.
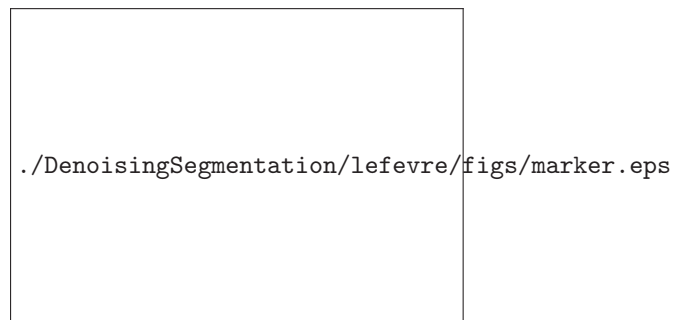
Figure 11.8. *Example of a uni-dimensional signal and its marker, obtained with an alternating sequential filter using line shaped SEs of length 5 pixels.*

11.4.4. *Iterative quantisation*

The third alternative presented here greatly differs from the previous schemes. Instead of considering the problem of colour/vector ordering as a means to compute colour extrema, that are to be used for defining the dilation and erosion operators, we concentrate directly on the computation of colour extrema, given a set of colours. Thus we can exploit the distribution of the vector/colour set in the multi-dimensional space, in order to better control the lexicographical comparison cascade. Naturally, since there is no underlying ordering, there is no complete lattice structure, hence the resulting operators are theoretically invalid. Nevertheless, as it will be shown in Sec. 11.5, they still have a certain practical interest.

Definitions

The α-trimming principle has long been used in filters such as the α-trimmed mean filter (αMF) and its variants [OTE 04] against impulsive noise. Given a vector $\mathbf{v} \in \mathbb{R}^n$, containing the sorted scalar pixels under the filtering window, the underlying idea of α-trimming consists in computing their mean by ignoring the 2α extreme :

$$\alpha MF(\mathbf{v}) = \frac{1}{n - 2\alpha} \sum_{i=\alpha+1}^{n-\alpha} v_i \tag{11.26}$$

where $\alpha \in [0, n/2]$. Similarly, in the case of multidimensional vectors we can apply a likewise principle to each dimension in an iterative mode. Specifically, in the case of the maximum, starting from the first dimension, we can sort all k vectors according to this dimension, and then keep the $\lceil \alpha \times k \rceil$ greatest. By repeating this simple process for each dimension, at each step the initial set of vectors will get smaller. In the eventual case where more than one vector remain at the end of this procedure, the last dimension is used for determining the sought extremum. A more formal description for computing the maximum based on this procedure, is given in Algorithm 2.

As to the minimum, it can be obtained in a likewise mode by simply sorting in a decreasing order. Consequently, the resulting extrema can be used in order to define the operators in Eqs (11.3) and (11.4) for multivariate pixels. An illustration of this approach on a three dimensional space $D1 \times D2 \times D3$ is given in Fig. 11.9. Moreover, the advantage of a collective extremum calculation comes however with an increased computational burden. To explain, assuming an optimal sorting procedure is used, in the worst scenario, where all n dimensions would need to be sorted, with k vectors the complexity would be in the order of $O(n \times k \times \log k)$. On the other hand, with the standard lexicographical ordering and the same scenario, the complexity would be $O(n \times k)$. Note that as k represents in practice the size of the SE in use, its value is relatively small.

Algorithm 2 The α-trimmed lexicographical maximum computation algorithm

---

   Input : a set $V = \{v_j\}$ of k, n-dimensional vectors; $\alpha \in]0, 1]$
   Output : max V
   for $i \leftarrow 1$ to $n - 1$ **do**
      Sort in increasing order the vectors of V with respect to their i^{th} dimension
      $k \leftarrow \lceil \alpha \times k \rceil$
      $V \leftarrow$ the greatest k vectors in V as well as those equal to the kth vector
      if $(|V| = 1)$ **then**
         return $v \in V$
      end if
      $i \leftarrow i + 1$
   end for
   return the greatest vector within V with respect to the nth dimension

---

Figure 11.9. *A set of vectors in a three dimensional space $D1 \times D2 \times D3$, and the three iterations of the α-trimmed lexicographical maximum computation with $\alpha = 0.5$ (left to right). According to the proposed approach, the maximum is the greatest with respect to the third dimension of the remaining three vectors.*

Setting α

The extrema obtained in this way, depend of course directly on the value of α. As a matter of fact, with an α approaching zero, from each dimension i, only the extreme (i.e. maximum or minimum) vector is kept along with those equal to it with respect to the ith dimension. In other words the procedure becomes identical to the standard lexicographical ordering. On the other hand, when α approaches one, priority is shifted gradually to the last dimension, and almost all initial vectors reach the last dimension where the comparison is finally decided.

This transition of priority is illustrated in Fig. 11.10. When keeping only a single value at each dimension, identically to a standard lexicographical ordering, the first dimension decides the outcome of the majority of comparisons. The final decision is

shifted rapidly to the last dimension by slightly increasing the number of kept vectors. The main difference with respect to other lexicographical ordering approaches, is the collective extremum calculation, conversely to a binary one. Specifically, even if the final extremum choice is made in the last dimension, unless $\alpha = 1.0$, all previous dimensions contribute to this choice by trimming the set of vectors accordingly. Additionally, more complicated priority relations among the available channels can be established by means of different α values for each vector dimension.

./DenoisingSegmentation/lefevre/figs/fig3.eps
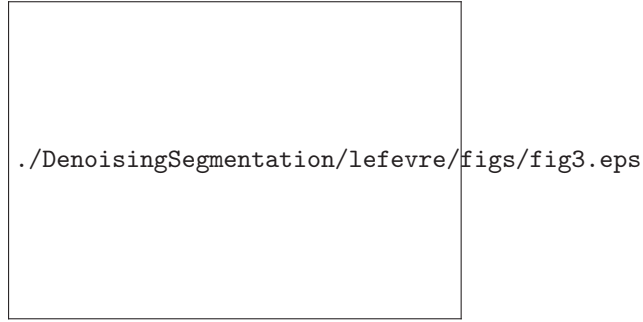
Figure 11.10. *The plot of comparisons decided by each dimension of the image Lenna (Fig. 11.3 left) in the RGB colour space, during a dilation with a square shaped SE of size 5×5 pixels, for various numbers of vectors kept at each iteration.*

In practice however, it is often necessary to set these arguments in an unsupervised way. Here we introduce a simple parameter setting model based on the standard deviation (σ) of each dimension. More precisely, if the data are relatively concentrated with respect to their i^{th} dimension, or in other words if this image channel does not contain much of the total variational information, we consider using a large α to be more pertinent, thus decreasing the influence of this dimension by carrying the majority of the input to the next dimension with minor trimming. On the other hand, if the data are highly dispersed with respect to the other dimensions, meaning that this channel represents relatively important variational information, a small α would be used leading to major trimming. Given n dimensions, one way of obtaining the corresponding α_i value of dimension i would be :

$$\forall\, i \in \{1, \ldots, n\}\,,\ \ \alpha_i = 1 - \frac{\sigma_i}{\sum_{j=1}^{n} \sigma_j} \tag{11.27}$$

where σ_j denotes the standard deviation of dimension j. A more effective way of employing this principle would be to first apply a principal components transformation on the image data and then use the variances of each new dimension to this end.

Variations

Using the α argument as a cardinality limit at each iteration of the algorithm can lead to unexpected situations, such as the one shown in Fig. 11.11. Indeed, by selecting the vectors to be used in the next stage of the cascade based only on the cardinality of the vector set (i.e. $\lceil \alpha \times k \rceil$), in other words irrespectively to their relative distances, one may end up with extremely scattered vector clusters. This situation can be easily countered by using α as a distance measurement. Specifically, instead of keeping the $\lceil \alpha \times k \rceil$ greatest vectors with respect to the ith dimension, one can keep only the vectors whose i^{th} dimension is at most at distance α with respect to the greatest vector of the same dimension.

./DenoisingSegmentation/lefevre/figs/fig4.eps
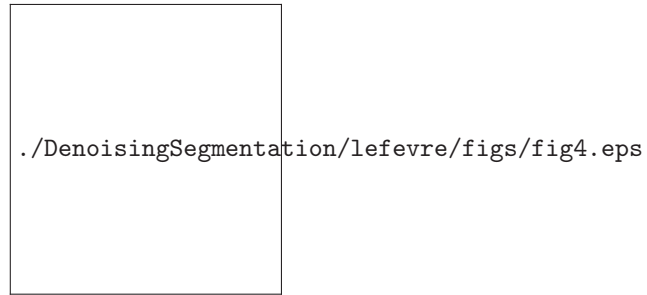
Figure 11.11. *Example of problematic vector distribution. The vector **a** is chosen as maximum with $\alpha = 0.4$, while being positioned at a far distance withg respect to the upper vector group.*

As previously mentioned, a more important "disadvantage" however of α-trimmed lexicographical extrema is the lack of an underlying binary ordering relation. Nevertheless, given any extremum computation method, it is possible to construct from it an ordering. In particular, considering a discrete multi-dimensional space, one can employ the collective extremum computation method at hand, in order to calculate the maximum or minimum of this space, and then repeat for the remaining points, up until the entire space is ordered.

Moreover, since α-trimmed lexicographical extrema are unique, this approach leads to a total ordering. The exact algebraic properties of the result however depend strongly both on the extremum computation method used (e.g. α-trimmed lexicographical, cumulative distances, etc) as well as on the extremum employed (i.e. minimum or maximum) during its construction.

11.5. Application to colour image analysis : texture classification

Having provided an insight into several morphological frameworks, and in particular some recent improvements over lexicographical ordering, in this section we carry out a brief series of comparative tests with the aim of better determining their properties and relative performances. As a representative application for which effective means of quantitative performance assessment are available, we propose to focus on the problem of colour texture classification and employ the colour textures of Outex13 (Fig. 11.12) [OJA 02]. This set contains 68 different textures, with 20 images from each obtained with 100 dpi resolution under an incandescent CIE A light source. The total amount of 1360 images have been evenly divided as training and test sets.

Figure 11.12. *Examples of the 68 textures of Outex 13 [MAE 04].*

The question whether colour should be processed separately or jointly from texture is still an open problem [MAE 04], and vector morphological feature extraction operators represent in their majority the latter case [HAN 05]. As colour texture descriptor we employ the morphological version of the autocorrelation operator, namely morphological covariance, which was initially proposed [MAT 75, SER 82] as the equivalent in MM of the autocorrelation operator. The morphological covariance K of an image f, is defined as the volume Vol (i.e. sum of pixel values for grey-level images, sum of the Euclidean norm of pixels in RGB for colour images, etc), of the image, eroded by a pair of points $P_{2,v}$ separated by a vector v :

$$K(f; P_{2,v}) = \text{Vol}\left(\varepsilon_{P_{2,v}}(f)\right) \tag{11.28}$$

In practice, K is computed for varying lengths of **v**, and most often the normalised version is used for measurements :

$$K^n(f) = \text{Vol}\left(\varepsilon_{P_{2,v}}(f)\right) / \text{Vol}(f) \tag{11.29}$$

Given the resulting uni-dimensional covariance series, one can gain insight into the structure of a given image [SOI 03]. In particular, the periodic nature of covariance is strongly related to that of its input. Furthermore, the period of periodic textures can easily be determined by the distance between the repeated peaks, that appear at multiples of the sought period; whereas the size of the periodic pattern can be quantified by means of the width of the peaks. In other words, their sharpness is directly proportional to the thinness of the texture patterns appearing in the input. Likewise, the initial slope at the origin provides an indication of the coarseness, with quick drop-off corresponding to coarse textures. In order to obtain additional information on the directionality of f, one can plot against not only different lengths of $\mathbf{v}$, but orientations as well.

The covariance based feature vectors have been calculated using four directions for the point pairs ($0°$, $45°$, $90°$, $135°$), each along with distances ranging from 1 to 49 pixels in steps of size two. Consequently 25 values are available for each direction, making a total of 100 values for every image channel after concatenation. The classification process is realised using a kNN classifier with $k = 1$ [MAE 04]. The colour space under consideration is LSH [ANG 05]. In order to carry out the various quantisation steps the luminance channel is manipulated with integer precision in $[0, 255]$. In particular, the extrema employed for defining the basic morphological operators, are computed using only the luminance channel (Lum), only saturation (Sat), only hue (Hue), lexicographical ordering, Eq. (11.17) (Lex), α-modulus lexicographical ($\alpha = 10$), Eq. (11.22) (α-modLex), quantisation based α-lexicographical ($\alpha = 10$), Eq. (11.24) (QuantaLex), with the LSH prioritisation function f described in [APT 08c], the marker based lexicographical ordering, Eq. (11.25) (MarkerLex), where the marker image is obtained by applying a leveling [MEY 04] along with a leveling marker image provided by median filtering with a square-shaped SE of size 7×7, the α-trimmed lexicographical extrema (α-trimmedLex), according to Algorithm 2 and $\alpha = 0.45$, as well as its variations : the one using adaptive α values, Eq. (11.27) (α-trimmedLex-adaptive), the total ordering resulting from the application of collective extremum computation scheme with $\alpha = 0.45$ (α-trimmedLex-ordering), and finally the version based on distance computations with $\alpha = 0.3$. The comparisons are realised in the order $L \rightarrow S \rightarrow H$, and the hue component is processed according to the principle of weighted multiple references which is not described here for the sake of conciseness [APT 09].

The results obtained from this texture classification application are given in Table 11.2. The overall performances appear to be relatively close. Specifically, the pertinence of luminance over both saturation and hue is observed, showing that colour information is in this case just an auxiliary component. As expected, Lum and Lex exhibit almost identical performances, a result showing the level of priority attributed to the first component, ignoring largely the remaining two dimensions. A minor improvement over Lex is obtained through α-modLex, as it allows saturation to participate at a slightly higher degree in the process of computing covariance. However, hue remains

	Accuracy (%)
Lum	73.82
Sat	66.91
Hue	55.29
Lex	74.26
α-modLex	76.62
QuantaLex	77.17
MarkerLex	79.86
α-trimmedLex	**81.47**
α-trimmedLex-adaptive	79.89
α-trimmedLex-ordering	54.35
α-trimmedLex-distance	73.74

Tableau 11.2. *Classification accuracies for the textures of Outex13, using vector erosion based covariance.*

largely unexploited. With the use of α-trimmed extrema, the chrominance dimensions increase their contribution to the final result. As a matter of fact, the empirically set α-trimmedLex leads to an overall best, while the unsupervised model provides a relatively sufficient approximation. Nevertheless, no improvements are observed with the last two variations. Moreover, QuantaLex leads to an improvement over α-modLex, as it controls the transitions among the image channels in a finer way, the difference being however almost negligible. The marker based approach improves the spectral quantisation-based ordering substantially and outperforms α-modLex. Apparently, the use of saturation with relatively "flat" regions, and that of luminance with transitions within the image, lead to more pertinent descriptors by better combining colour and textural information.

11.6. Conclusion

This chapter has introduced the theoretical concepts behind the multivariate morphological framework, as well as the various approaches that have been proposed with this purpose. Given the complete lattice theory based structure of morphology, it becomes easy to observe that a vector ordering is sufficient for the definition of valid multivariate morphological operators. Among the numerous ordering possibilities, total orderings are the most theoretically sound approaches since they ensure extrema uniqueness and vector preservation.

In this context, the lexicographical ordering is certainly the most widely used total ordering since besides the aforementioned theoretical advantages, it also provides a means of effective customisation through the configuration of the lexicographical cascade's order. However its standard definition introduces an extreme prioritisation

of the first vector dimension and thus limits its practical interest by ignoring the other available image channels.

We address this problem and present three different variations of lexicographical ordering with the purpose of resolving this issue. The first, quantisation based α-lexicographical ordering, consists of a generalisation of the α-modulus lexicographical ordering. By sub-quantising the vector dimensions according to arbitrary models, it hence makes possible for the user to incorporate into the ordering any a priori information available on the channels under consideration. The second variation of lexicographical ordering, marker based lexicographical ordering, uses an image specific principle, where the first channel is processed by means of grey-level operators, in order to form flat zones, that determine the areas where priority is shifted to the subsequent dimensions. Additionally, a third method for computing α-trimmed lexicographical extrema has been introduced, which also succeeds in manipulating the contribution of each channel to this process through a user defined argument. Nevertheless, as there is no underlying ordering it leads to pseudo-morphological colour operators.

A comparison between the different vector orderings presented in this chapter has also been given, with the purpose of texture classification in colour images by means of morphological covariance. Nevertheless it should be noted that the proposed ordering variations are usable in the general case of multivariate image data, e.g. hyperspectral remote sensing images, multispectral astronomical images, etc.

Further readings related to multivariate mathematical morphology applied to colour image analysis include the PhD work of the first author [APT 08a] and its related publications [APT 07, APT 08b, APT 08c, APT 09], as well as the numerous work made in the leading research centre within the field of mathematical morphology, namely the *Centre de Morphologie Mathématique* of the Mines ParisTech School, France (see http ://cmm.ensmp.fr), e.g. doctoral theses of Angulo and Hanbury.

[Al- 03] AL-OTUM H. M., « Morphological operators for color image processing based on Mahalanobis distance measure », *Optical Engineering*, vol. 42, n°9, p. 2595-2606, September 2003.

[ANG 03] ANGULO J., SERRA J., « Morphological coding of color images by vector connected filters », *IEEE Proceedings of the 7th International Symposium on Signal Processing and Its Applications (ISSPA'2003)*, vol. 1, Paris, France, p. 69-72, July 2003.

[ANG 05] ANGULO J., « Unified morphological color processing framework in a lum/sat/hue representation », RONSE C., NAJMAN L., DECENCIÈRE E., Eds., *Mathematical Morphology : 40 Years On*, vol. 30 de *Computational Imaging and Vision*, p. 387-396, Springer-Verlag, Dordrecht, 2005.

[ANG 07] ANGULO J., « Morphological colour operators in totally ordered lattices based on distances : Application to image filtering, enhancement and analysis », *Computer Vision and Image Understanding*, vol. 107, n°1-2, p. 56-73, July 2007.

[APT 07] APTOULA E., LEFÈVRE S., « A comparative study on multivariate mathematical morphology », *Pattern Recognition*, vol. 40, n°11, p. 2914–2929, November 2007.

[APT 08a] APTOULA E., Analyse d'Images Couleur par Morphologie Mathématique. Application à la Description, l'Annotation, et la Recherche d'Images — Colour Image Analysis with Mathematical Morphology. Application to Image Description, Annotation, and Retrieval, PhD thesis, Université Louis Pasteur, Strasbourg, France, 2008.

[APT 08b] APTOULA E., LEFÈVRE S., « alpha-trimmed lexicographical extrema for pseudo-morphological image analysis », *Journal of Visual Communication and Image Representation*, vol. 19, n°3, p. 165–174, April 2008.

[APT 08c] APTOULA E., LEFÈVRE S., « On lexicographical ordering in multivariate mathematical morphology », *Pattern Recognition Letters*, vol. 29, n°2, p. 109–118, January 2008.

[APT 09] APTOULA E., LEFÈVRE S., « On the morphological processing of hue », *Image and Vision Computing*, 2009, to appear, 20 pages, doi :10.1016/j.imavis.2008.12.007.

[AST 90] ASTOLA J., HAAVISTO P., NEUVO Y., « Vector Median Filters », *IEEE Proceedings*, vol. 78, n°4, p. 678-689, April 1990.

[BAR 76] BARNETT V., « The ordering of multivariate data », *Journal of the Statistical Society A*, vol. 139, n°3, p. 318-355, 1976.

[CHA 98] CHANUSSOT J., Approches Vectorielles ou Marginales pour le Traitement d'Images Multi-composantes, PhD thesis, Université de Savoie, France, 1998.

[COM 99] COMER M., DELP E., « Morphological operations for color image processing », *Journal of Electronic Imaging*, vol. 8, n°3, p. 279-289, July 1999.

[GOU 95] GOUTSIAS J., HEIJMANS H. J. A. M., SIVAKUMAR K., « Morphological operators for image sequences », *Computer Vision and Image Understanding*, vol. 62, n°3, p. 326-346, November 1995.

[GOU 00] GOUTSIAS J., HEIJMANS H. J. A. M., « Fundamenta Morphologicae Mathematicae », *Fundamenta Informaticae*, vol. 41, n°1-2, p. 1-31, January 2000.

[HAN 01] HANBURY A., SERRA J., « Morphological operators on the unit circle », *IEEE Transactions on Image Processing*, vol. 10, n°12, p. 1842-1850, December 2001.

[HAN 02] HANBURY A., SERRA J., « Mathematical Morphology in the CIELAB Space », *Image Analysis and Stereology*, vol. 21, n°3, p. 201-206, March 2002.

[HAN 05] HANBURY A., KANDASWAMY U., ADJEROH D. A., « Illumination-invariant Morphological Texture Classification », RONSE C., NAJMAN L., DECENCIÈRE E., Eds., *Mathematical Morphology : 40 Years On*, vol. 30 de *Computational Imaging and Vision*, p. 377-386, Springer-Verlag, Dordrecht, 2005.

[HAN 08] HANBURY A., « Constructing cylindrical coordinate colour models », *Pattern Recognition Letters*, vol. 4, n°29, p. 494-500, March 2008.

[HAR 91] HARDIE R. C., ARCE G. R., « Ranking in R^p and its use in multivariate image estimation », *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 1, n°2, p. 197-209, June 1991.

[HEI 90] HEIJMANS H. J. A. M., RONSE C., « The algebraic basis of mathematical morpho-logy, part I : dilations and erosions », *Computer Vision, Graphics and Image Processing*, vol. 50, n°3, p. 245–295, June 1990.

[HEI 94] HEIJMANS H. J. A. M., *Morphological Image Operators*, Advances in Electronics and Electron Physics Series, Academic Press, Boston, 1994.

[LOU 02] LOUVERDIS G., VARDAVOULIA M., ANDREADIS I., TSALIDES P., « A new ap-proach to morphological color image processing », *Pattern Recognition*, vol. 35, n°8, p. 1733-1741, September 2002.

[MAE 04] MAENPAA T., PIETIKAINEN M., « Classification with color and texture : jointly or separately ? », *Pattern Recognition*, vol. 37, n°8, p. 1629-1640, August 2004.

[MAT 75] MATHERON G., *Random Sets and Integral Geometry*, Wiley, New York, 1975.

[MEY 04] MEYER F., « Levelings, Image Simplification Filters for Segmentation », *Journal of Mathematical Imaging and Vision*, vol. 20, p. 59-72, 2004.

[OJA 02] OJALA T., MAENPAA T., PIETIKAINEN M., VIERTOLA J., KYLLONEN J., HUOVI-NEN S., « Outex : New framework for empirical evaluation of texture analysis algorithms », *Proceedings of the 16th ICPR*, vol. 1, Quebec City, Canada, p. 701-706, August 2002.

[ORT 01] ORTIZ F., TORRES F., ANGULO J., PUENTE S., « Comparative study of vectorial morphological operations in different color spaces », *Proceedings of Intelligent Robots and Computer Vision XX : Algorithms, Techniques, and Active Vision*, vol. 4572, p. 259-268, November 2001.

[OTE 04] OTEN R., DE FIGUEIREDO R. J. P., « Adaptive alpha-trimmed mean filters under deviations from assumed noise model », *IEEE Transactions on Image Processing*, vol. 13, n°5, p. 627-639, May 2004.

[PLA 04] PLAZA A., MARTINEZ P., PEREZ R., PLAZA J., « A new approach to mixed pixel classification of hyperspectral imagery based on extended morphological profiles », *Pattern Recognition*, vol. 37, n°6, p. 1097-1116, June 2004.

[RIV 04] RIVEST J., « Morphological operators on complex signals », *Signal Processing*, vol. 84, n°1, p. 133-139, January 2004.

[RON 91] RONSE C., HEIJMANS H. J. A. M., « The algebraic basis of mathematical morpho-logy, part II : openings and closings », *Computer Vision, Graphics and Image Processing*, vol. 54, n°1, p. 74–97, July 1991.

[SER 82] SERRA J., *Image Analysis and Mathematical Morphology Vol I*, Academic Press, London, 1982.

[SER 93] SERRA J., « Anamorphoses and function lattices », DOUGHERTY E. R., Ed., *Ma-thematical Morphology in Image Processing*, Chapitre 13, p. 483-523, Marcel Dekker, New York, 1993.

[SER 05] SERRA J., « Morphological Segmentations of Colour Images », RONSE C., NA-JMAN L., DECENCIÈRE E., Eds., *Mathematical Morphology : 40 Years On*, vol. 30 de *Computational Imaging and Vision*, p. 151-176, Springer-Verlag, Dordrecht, 2005.

[SOI 02] SOILLE P., PESARESI M., « Advances in mathematical morphology applied to geoscience and remote sensing », *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, n°9, p. 2042-2055, September 2002.

[SOI 03] SOILLE P., *Morphological Image Analysis : Principles and Applications*, Springer-Verlag, Berlin, 2nd édition, 2003.

[TAL 98] TALBOT H., EVANS C., JONES R., « Complete ordering and multivariate mathematical morphology », HEIJMANS H., ROERDINK J., Eds., *Mathematical Morphology and Its Applications to Image and Signal Processing*, Amsterdam, Kluwer Academic Press, p. 27-34, 1998.

[VAR 02] VARDAVOULIA M. I., ANDREADIS I., TSALIDES P., « Vector ordering and morphological operations for colour image processing : fundamentals and applications », *Pattern Analysis and Applications*, vol. 5, n°3, p. 271-287, 2002.

[WIL 92] WILSON S. S., « Theory of matrix morphology », *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, n°6, p. 636-652, June 1992.

New challenge for massive multicomponent image analysis

Chapitre 12

Spectral-Spatial Classification of Hyperspectral Images Using Segmentation-Derived Adaptive Neighborhoods

12.1. Introduction

The advent and growing availability of hyperspectral imagery, which records a detailed spectrum of light for each pixel, has opened new possibilities in image analysis and classification. Examples of hyperspecral imaging systems are AVIRIS [GRE 98], HYDICE [RIC 93], ARCHER [Ste 05], HyMap [COC 98], Hyperion [PEA 00]. They cover a range of 126 to 512 spectral channels, with the spatial resolution 3-30 m per pixel (Figure 12.1 depicts the structure of a hyperspectral image). The fine spectral resolution of the data provides an invaluable source of information about physical materials, increasing the capability to distinguish structures and objects in the image scene [CHA 03].

However, a large number of spectral channels implies high dimensionality of the data and presents challenges to image analysis and classification. Most of the commonly used methods designed for the analysis of grey level, color or multispectral images are not appropriate for hyperspectral images. As a matter of fact, very limited reference data are usually available (the ratio of the number of referenced samples to the number of spectral channels quickly drops), which limits the performances of supervised classification techniques. Furthermore, for analysis of hyperspectral

---

Chapitre rédigé par Y. TARABALKA.

Figure 12.1. *Structure of a hyperspectral image*

images the well known curse of dimensionality prevents robust statistical estimations, usual vector norms become meaningless and so on (e.g., the Hughes phenomenon [HUG 68]). Therefore, to take full advantage of the rich information provided by the spectral dimension, the development of new algorithms is required.

The first attempts to classify hyperspectral images were designed to assign each pixel to one of the classes based on its spectrum only [LAN 02]. A wide range of *pixel-wise* classification techniques were investigated for this purpose, such as maximum-likelihood or Bayesian estimation techniques [LAN 03, CHE 08, BAL 08], neural networks [SUB 97, YAN 99, HER 04], decision trees [GOE 03, ZHO 05], genetic algorithms [VAI 03] and kernel-based methods [CAM 05, FAU 07, FAU 06, GUO 08]. In particular, Support Vector Machines (SVM) and other kernel-based methods have shown good classification results, because they tend to be robust when a limited number of training samples is available.

Further modification to improve classification results consists in the integration of spatial and spectral information during the image analysis. *Spectral-spatial classification* aims at assigning each image pixel to one class using a feature vector based on 1) its own spectral value (the spectral information) and 2) information extracted from its neighborhood (referred to as the spatial information in the following). One of the approaches consists in including the information from the closest neighborhood when classifying each pixel. For instance, morphological filters [FAU 07], Markov random fields [FAR 05] and Hidden Markov Quadrees [**?**] were used for this purpose. These methods have shown improvements in classification accuracies compared to the pixel-wise methods, when applied to hyperspectral images. However, their use raises the problem of scale selection, especially when small or complex structures are present in the image.

Another approach to include spatial information in classification consists in performing image segmentation. *Segmentation* can be defined as an exhaustive partitioning of the input image into regions, each of which is considered to be homogeneous

with respect to some criterion of interest (homogeneity criterion, e.g., intensity or texture) [JAI 95]. Each of these regions from a segmentation map can be further used as an adaptive neighborhood for all the pixels within this region. In this chapter we will describe several techniques for segmentation of hyperspectral images.

When adaptive neighborhoods are defined, the obtained spatial information must be combined with the available spectral information into a classifier. Here, three major approaches can be considered : 1) for each region, a set of features can be defined and then the whole region is classified ; 2) a new set of features for every pixel is defined, which combines spectral information of this pixel and contextual information from its neighborhood ; then each pixel is classified using these features ; 3) each pixel is classified independently using only its spectral information, and then for each region from a segmentation map, all its pixels are classified to one class, based on the results of a previous step and a defined rule for fusion of spectral and spatial information. In this chapter, different approaches are discussed, and then the experimental results are present using the third approach. For that, pixel-wise Support Vector Machines classification and segmentation are performed independently, and then the results are combined, using the *majority vote* approach.

This chapter is organized as follows. In the next section, two sets of a remote sensing hyperspectral data used for the illustration and validation of segmentation and classification techniques are briefly described. In Section 12.3, techniques for segmentation of hyperspectral images are discussed. Section 12.4 presents approaches for fusion of spectral and spatial information into a classifier. Experimental results for spectral-spatial classification are presented in Section 12.5. Finally, conclusions are drawn.

12.2. Data Set

Two different hyperspectral airborne images were used for the experiments, with different contexts (one urban area and one agricultural area), different spatial resolutions (1.3 m and 20 m per pixel, respectively) and different number of bands (103 and 220 bands, respectively). These two data sets are detailed in the two next subsections.

12.2.1. *"University of Pavia" Image*

The image of an urban area was recorded by the ROSIS-03 optical sensor, over the University of Pavia, Italy. The image has spatial dimensions of 610 by 340 pixels, with a spatial resolution of 1.3m per pixel. The number of bands of the ROSIS-03 sensor is 115 with a spectral coverage ranging from 0.43 to 0.86μm. The 12 most noisy channels have been removed, and the experiments are conducted on the 103-band image.

(a)   (b)
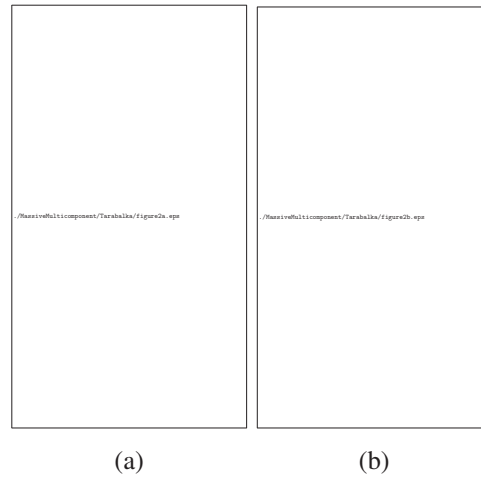
Figure 12.2. University of Pavia *image : (a) 3-band color composite ; (b)*
Reference data : Asphalt, meadows, gravel, trees, metal sheets, bare soil,
bitumen, bricks and shadows.

Nine classes of interest are considered, namely : Asphalt, meadows, gravel, trees, metal sheets, bare soil, bitumen, bricks, shadows. The number of test and training samples for each class can be found in [TAR a]. A 3-band false color image and the reference data are presented in Figure 12.2.

12.2.2. *"Indiana" Image*

The image of a vegetation area was captured by the AVIRIS sensor over the Indian Pines test site in Northwestern Indiana. The image is 145 by 145 pixels, with a spatial resolution of 20m per pixel. It is composed of 220 spectral channels, and the full spectral range was used for experiments. The data contains 16 classes, namely : Corn-no till, corn-min till, corn, soybeans-no till, soybeans-min till, soybeans-clean till, alfalfa, grass/pasture, grass/trees, grass/pasture-mowed, hay-windrowed, oats, wheat, woods, bldg-grass-tree-drives, stone-steel towers. The number of samples for each class in the available reference data can be found in [TAD 98]. Figure 12.3 depicts false color composite and the reference data. Ten percent of samples for each class are chosen randomly from the reference data, in order to be used as training samples.

(a)   (b)
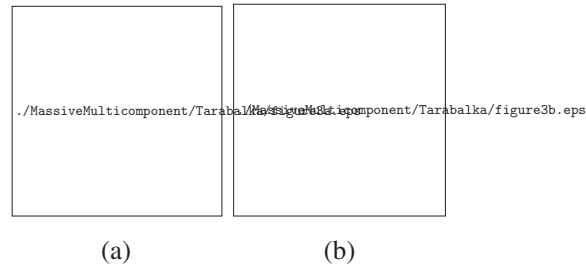
Figure 12.3. Indiana *image : (a) 3-band color composite ; (b) Reference data :*
Corn-no till, Corn-min till, Corn, Soybeans-no till, Soybeans-min till,
Soybeans-clean till, Alfalfa, Grass/pasture, Grass/trees, Grass/pasture-mowed,
Hay-windrowed, Oats, Wheat, Woods, Bldg-Grass-Tree-Drives, Stone-steel
towers.

12.3. Segmentation of Hyperspectral Images

Fu and Mui [FU 81] identified three classes of image segmentation techniques :
Edge-based, *region-based* and *characteristic feature thresholding or clustering*. Methods from the first two classes operate in the spatial domain, while those from the last class work in the spectral domain. Finally, a combination of spatial-based and spectral-based segmentation is possible.

12.3.1. *Spatial-Based Segmentation*

Segmentation techniques working in the spatial domain search for groups of spatially adjacent pixels (regions) which are similar accordingly to the defined criterion. *Edge-based* techniques search for discontinuities in the image, while *region-based* techniques search for similarities between image regions.

In previous studies, several methods for multispectral image segmentation have been investigated. Numerous works exploit region merging techniques, where neighboring image segments are merged iteratively based mostly on their spectral similarity [DAR 03, TIL 03]. For instance, the eCognition software performs multiresolution segmentation, based on a bottom-up region merging [DAR 03]. Initially, each pixel is considered as a separate region, and subsequently pairs or regions are merged, based on a homogeneity criterion, which is a combination of spectral and shape properties. The main drawback of applying region merging for image segmentation is that the homogeneity criterion or convergence criterion must be chosen (in other words, the level of details for a segmentation map must be defined). For accurate segmentation,

these techniques usually produce a pyramid of segmentation maps, using a range of thresholds. Then, manual interpretation of the results is needed.

Other studies exploit mathematical morphology based segmentation approaches [PES 01, SOI 96, FLO 98, SCH 01, LI 04, NOY 07, AK08, TAR 08], which mostly use granulometries or watershed transformation. The extension of morphological operators to the case of multispectral images is not straightforward, because there is no natural means for total ordering of multivariate pixels, which is a requirement in mathematical morphology. An extension of a watershed segmentation to the case of hyperspectral images has been investigated by Noyel *et al.* [NOY 07] and Tarabalka *et al.* [TAR b]. In the following, the watershed segmentation technique is described, and then its extensions for hyperspectral images are presented.

12.3.1.1. *Watershed segmentation*

Watershed transformation is a powerful mathematical morphology technique for image segmentation [BEU 79, SOI 03]. It considers a two-dimensional one-band image as a topographic relief. The value of a pixel h stands for its elevation. The watershed lines divide the image into catchment basins, so that each basin is associated with one *minimum* in the image. The watershed transformation is usually applied to the gradient function of the image. Gradient defines transitions between regions, so that it has high values on the borders between objects and *minima* in the homogeneous regions. And in this case, if the crest lines in the gradient image correspond to the edges of image objects, watershed transformation partitions this image into meaningful regions.

A wealth of literature describes techniques for computing the watershed transformation (see for instance [VIN 91, MEY 94, STO 00, OSM 07]). Vincent and Soille [VIN 91] have proposed an efficient watershed algorithm using flooding simulations, which has become one of the classical algorithms to compute watersheds.

The output of the watershed transform is a partition of the image composed of regions (sets of pixels connected to the same local *minimum*) and of *watershed pixels* (the borders between the regions). Figure 12.4 shows an example of watershed transformation in one dimension, where three regions, associated with the three *minima*, are defined. The two *maxima* correspond to the borders between regions and are not assigned to any region (watershed pixels).

Typically, the result of watershed segmentation on the gradient image without any additional processing is a severe oversegmentation (every single local *minimum* of the gradient leads to one region). Common ways to reduce the number of local *minima* are to filter the original image or the gradient function (*e.g.*, area filtering) or to use markers [MEY 90]. The oversegmentation effect can be also corrected using some post-processing, such as merging of similar neighboring regions.

Figure 12.4. *Example of watershed transformation in one dimension.*

In the next section, different approaches of the extension of a watershed segmentation to the case of hyperspectral images are presented.

12.3.1.2. *Watershed segmentation of hyperspectral images*

As previously mentioned, the watershed transformation requires as input a one-band image. Our objective is to apply watershed to a B-band hyperspectral image. Let us consider this image as a set of n pixel vectors $\mathbf{X} = \{\mathbf{x}_j \in \mathbb{R}^B, j = 1, 2, ..., n\}$ (each pixel is characterized by its spatial location and a vector of spectral values ; see Fig. 12.1). We denote also the image of every spectral band as $X_\lambda, \lambda = 1, 2, ..., B$.

Before computing a gradient, *feature extraction* on the original image can be performed, applying one of the transformations such as the Principal Component Analysis (PCA) [RIC 99, GUP 06], Maximum Noise Fraction (MNF) [GRE 88] and Independent Component Analysis (ICA) [WAN 06a, MOU 08]. The aim of this step is to obtain either a one-band image or a multi-band image which would contain enough information to distinguish between spatial structures in the image.

If a one-band image with a good distinguishing capability between structures can be obtained, the algorithm for computing a gradient and watershed is straightforward. For instance, for a one-band image Y, a basic morphological gradient (also called the Beucher gradient) can be applied. It is defined as the arithmetic difference between the dilation and the erosion of Y by the structuring element E (definitions of the dilation and the erosion can be found in [SOI 03]) :

$$\rho_E(Y) = \delta_E(Y) - \varepsilon_E(Y). \qquad (12.1)$$

If, however, at the input of the *gradient* step we still have a multi-band image, we can proceed in different ways, that can be grouped into the following three categories :

– to compute a vectorial gradient ;

– to compute a multidimensional gradient ;

– to combine watershed segmentation maps *a posteriori*.

These three options are discussed in the next three sections.

12.3.1.3. *Computation of a vectorial gradient*

Vectorial gradients are based on the distance between pixel vectors, and produce from the B-band image one-band gradient [EVA 06, NOY 07]. Noyel *et al.* [NOY 07] proposed to use a metric-based gradient for hyperspectral images, which is described as follows : For each pixel vector $\mathbf{x}_p$, let $\psi = [\mathbf{x}_p^1, \mathbf{x}_p^2, ..., \mathbf{x}_p^e]$ be a set of e vectors in the neighborhood of the vector $\mathbf{x}_p$ (set ψ does not contain $\mathbf{x}_p$). The *metric-based gradient* is defined as a difference between the *supremum* and the *infimum* of the defined distances between $\mathbf{x}_p$ and vectors from the set ψ :

$$\nabla_{\psi,d}^{MB}(\mathbf{x}_p) = \sup_{i \in \psi}\{d(\mathbf{x}_p, \mathbf{x}_p^i)\} - \inf_{j \in \psi}\{d(\mathbf{x}_p, \mathbf{x}_p^j)\}. \tag{12.2}$$

Various distances can be used to compute gradient from (12.2) such as Euclidean, Mahalabobis, chi-squared distances [NOY 07].

Another type of vectorial gradient is the *Robust Color Morphological Gradient* (RCMG). It was developed for color images by Evans and Liu [EVA 06] and proposed to be used for hyperspectral images by Tarabalka *et al.* [TAR 08]. For each pixel vector $\mathbf{x}_p$, let $\chi = [\mathbf{x}_p^1, \mathbf{x}_p^2, ..., \mathbf{x}_p^e]$ be a set of e vectors within a structuring element E, which defines the neighborhood of the vector $\mathbf{x}_p$, and the set χ contains $\mathbf{x}_p$. The *Color Morphological Gradient (CMG)*, using the Euclidean distance, is computed as

$$\nabla_{\chi,d}^{CM}(\mathbf{x}_p) = \max_{i,j \in \chi}\{\|\mathbf{x}_p^i - \mathbf{x}_p^j\|_2\}, \tag{12.3}$$

i.e., the *maximum* of the distances between all pairs of vectors in the set χ. One of the drawbacks of the CMG is that it is very sensitive to noise. To overcome the problem of outliers, the RCMG has been proposed [EVA 06]. The scheme to make the CMG robust consists in removing the two pixels that are the furthest apart and then finding the CMG of the remaining pixels. This process is repeated several times until a good estimate of the gradient is obtained. Thus, the *RCMG*, using the Euclidean distance, can be defined as

$$\nabla_{\chi,d}^{RCM}(\mathbf{x}_p) = \max_{i,j \in [\chi - REM_r]}\{\|\mathbf{x}_p^i - \mathbf{x}_p^j\|_2\}, \tag{12.4}$$

where REM_r is the set of the r vector pairs removed. The appropriate value of the parameter r in (12.4) depends on the chosen structuring element E and the amount of noise present in the image, as discussed in [EVA 06]. When a one-band vectorial gradient is computed, it is used as the input of the watershed algorithm.

(a)   (b)
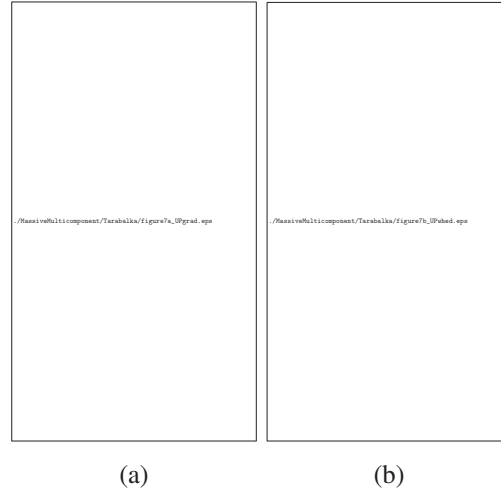
Figure 12.5. *Result of watershed segmentation for the* University of Pavia
image : (a) RCMG using the Euclidean distance, $r = 1$; (b) Watershed on the
RCMG.

Example 1 : RCMG on the *University of Pavia* image is computed by (12.4), with $r = 1$. Then watershed transformation is applied to the obtained gradient, using the algorithm of Vincent and Soille [VIN 91] (based on 8-neighborhood connectivity). Figure 12.5 shows the results of watershed segmentation (for a segmentation map, the colors of each region correspond to the label of this region, scaled in order to obtain a grey-scale 8-bit image). The segmentation map contains 11802 regions. As expected, the obtained segmentation results are severely oversegmented (each local *minimum* of the gradient results in one region). However, from visual observation undersegmentation is not present (or not significant). Therefore, the obtained regions can be further used as adaptive neighborhoods in spectral-spatial classification.

12.3.1.4. *Multidimensional gradient methods*

Another approach to compute a one-band gradient from the multi-band image consists in considering the B-band image as a set of B one-band images. In this case, the gradients of every spectral band can be computed, using for instance a morphological gradient (see (12.1)). Then the obtained B gradient images $\rho_E(X_\lambda), \lambda = 1, 2, ..., B$ are combined into one image using linear or non-linear operators.

As an example of the linear operators, the weighted sum of gradients can be computed by

(a)   (b)
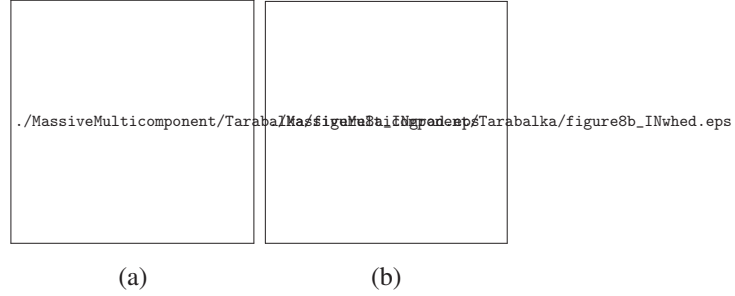
Figure 12.6. *Result of watershed segmentation for the* Indiana *image : (a) Sum of gradients over all bands ; (b) Watershed on the sum of gradients over all bands.*

$$\nabla_E^+(\mathbf{X}) = \sum_{\lambda=1}^{B} \omega_\lambda \rho_E(X_\lambda), \tag{12.5}$$

where ω_λ denotes the weight of the gradient of the band λ. If $\omega_\lambda = 1, \lambda = 1, 2, ..., B$, all the bands are supposed to have an equal importance in defining the gradient. Modifying the weight coefficients, the gradient estimation can be improved. For instance, coefficients that are inversely proportional to the estimated noise of each spectral band can be used as the weights in (12.5).

Examples of non-linear operators are the *supremum* and the median operators. The gradient *supremum* over morphological gradients of every band is defined as follows : for every pixel the *supremum* over all gradient images is taken as the output value of this pixel.

Example 2 : First, a morphological gradient of every band of the *Indiana* image is computed. Then, the sum of gradients is obtained by (12.5), assuming that $\omega_\lambda = 1, \lambda = 1, 2, ..., B$. Finally, a watershed segmentation is applied, using the algorithm of Vincent and Soille. The obtained gradient and segmentation results are shown in Figure 12.6. The segmentation map consists of 1215 regions. As in *Example 1*, the segmentation results are oversegmented, but undersegmentation is not observed.

12.3.1.5. *Combination of watershed segmentation maps*

First, B standard gradients are computed, one for each spectral band. Then each gradient image is used to compute a watershed transformation. This gives B segmentation maps that can be further combined to provide a single output segmentation map.

One of the ways to combine the B segmentation maps, in order to define relevant edges, consists in summing watershed lines. Here, for each segmentation map obtained from the gradient of band λ, a binary image W_λ of watershed lines is produced. Thus, W_λ is an image in which watershed pixels are equal to 1 and all other pixels are equal to 0. The sum of watershed lines is computed by

$$\mathbf{W} = \sum_{\lambda=1}^{B} W_\lambda. \tag{12.6}$$

The obtained image $\mathbf{W}$ can be further thresholded, in order to define the border pixels that were presented in most of the segmentation maps, hence ensuring a reliable edge detection. However, when summing the watershed lines, we do not have information about regions anymore, but only about edges. Furthermore, some edges can become open after thresholding. Therefore, closing of edges and image region labeling must be performed after the procedure described above.

12.3.1.6. *Segmentation map without watershed pixels*

When one of the watershed segmentation techniques described above is applied to a hyperspectral image, in the resulting segmentation map each pixel contains a label of the region it belongs to, or a watershed pixel label (see Figure 12.4).

It is often desirable to produce a segmented image where each pixel belongs to some region, without border pixels between regions. In this case, each watershed pixel can be assigned to one of the regions in its neighborhood. For this purpose, a standard vector median [AST 90] for every region $\mathbf{S}$ ($\mathbf{S} = \{\mathbf{s}_j \in \mathbb{R}^B, j = 1, 2, ..., l\}$, $\mathbf{S} \subseteq \mathbf{X}$, with l equal to the number of pixels in the region) can be computed. A standard vector median $\mathbf{s}_{VM}$ for a set of pixel vectors is a vector, which fulfills the condition that the sum of the distances between this vector and all the other vectors in the set is minimal (for instance, L_1 norm is used to compute distances) :

$$\mathbf{s}_{VM} = \arg \min_{\mathbf{s} \in \mathbf{S}} \left\{ \sum_{j=1}^{l} \|\mathbf{s} - \mathbf{s}_j\|_1 \right\}. \tag{12.7}$$

Every watershed pixel is assigned to the neighboring region with the "closest" median, *i.e.*, the distance between the vector median of this region and the watershed pixel vector is minimal.

12.3.2. *Spectral-Based Segmentation*

Spectral-based segmentation techniques search for similarities between image pixels and clusters of pixels, not taking into consideration the spatial location of these pixels.

Lambert and Macaire [LAM 00] have grouped these techniques into two classes : *Histogram-based* and *cluster-based* methods. The *histogram-based* methods relate the modes of a spectral histogram to homogeneous regions in the image [**?**, PAC 03]. With a high dimensionality, these methods become memory-consuming and produce less accurate results. The *cluster-based* segmentation techniques aim at finding distinct structures in the spectral feature space. Thus, *clustering* is an exhaustive partitioning of a set of pixels from the input image into homogeneous groups of pixels.

A taxonomy and survey of clustering techniques can be found in [JAI 99]. Two principal groups of clustering methods can be distinguished : Hierarchical and partitional approaches. Hierarchical methods usually produce a dendrogram, where at the lowest level each cluster contains only one pixel (*i.e.*, each pixel forms a cluster), and with the increase of levels the most similar clusters are merged (and the number of clusters decreases). Then, the result with the desired number of clusters can be chosen. Lee and Crawford [LEE 05] have applied the hierarchical clustering approach for unsupervised classification of hyperspectral images. Hierarchical clustering is a versatile technique for image segmentation that can produce a series of segmentation results. However, its application to high-dimensional data leads to significant time and memory requirements.

The alternative approach is partitional clustering. The algorithms of this approach produce a single partition of the data and have advantages when applying for large data sets, as they are computationally and memory less demanding than hierarchical clustering methods. However, the number of desired clusters must be chosen. The other problem accompanying the use of these techniques is that the clustering results depend on the initialization. In the next section, we present two approaches of partitional clustering and discuss their use for hyperspectral images.

12.3.2.1. *Partitional clustering of hyperspectral images*

The three principal stages of the clustering technique are :

1) *Feature selection/extraction* : *Feature selection* consists in identifying a subset of the original features. *Feature extraction* consists in applying one or more transformations of the input features to produce new salient features. As a pixel vector from hyperspectral image contains hundreds of spectral values, feature extraction/selection is often a required step, to reduce the dimensionality of the clustering/classification problem. The most common transformations are PCA [RIC 99], MNF [GRE 88], ICA [HYV 01, WAN 06a, MOU 08].

2) *Similarity measure* : Clustering aims at grouping pixels, so that pixels belonging to the same cluster are spectrally similar. To quantify this relationship, a similarity measure must be chosen. Proximity between pixels is usually measured by a distance function defined on pairs of spectral values. A simple distance measure like the Euclidean distance is often used to measure similarity between vectors. Description of various distance measures can be found in [DID 76, JAI 99, NOY 07].

3) *Grouping* : In this step, pixels are grouped into clusters. Partitional clustering algorithms identify the partition that optimizes a clustering criterion.

Tarabalka *et al.* [TAR a] have investigated two partitional clustering algorithms for segmentation of hyperspectral images : ISODATA (squared error clustering method) [BAL 65] and Expectation Maximization (EM) for the Gaussian mixture resolving [DEM 77]. Both ISODATA and EM are iterative optimization techniques. Thus, on each iteration i a partition $\mathbf{Q}_1^i, \mathbf{Q}_2^i, ..., \mathbf{Q}_C^i$ of the set $\mathbf{X}$ into C clusters is computed, so that $\mathbf{Q}_c^i = \{\mathbf{x}_{j,c}^i \in \mathbf{R}^B, j = 1, 2, ..., m_c^i\}$ contains the pixels belonging to the component c on the iteration i, and m_c^i is the number of pixels in $\mathbf{Q}_c^i$.

12.3.2.2. *ISODATA algorithm*

The simplest and most frequently used criterion in partitional clustering is the squared error criterion, which is the most suitable in the case of isolated and compact clusters. The squared error for a clustering Υ of a set $\mathbf{X}$ into C clusters is defined as :

$$e^2(\mathbf{X}, \Upsilon) = \sum_{c=1}^{C} \sum_{j=1}^{m_c} \|\mathbf{x}_{j,c} - \boldsymbol{\mu}_c\|^2 \qquad (12.8)$$

where $\boldsymbol{\mu}_c$ is the centroid of the cluster c.

ISODATA clustering is a well-known algorithm [BAL 65] which uses the squared error criterion. It starts with a random initial partition of the pixel vectors into candidate clusters and then reassigns these vectors to clusters in such a way that the squared error (12.8) is reduced at each iteration, until a convergence criterion is achieved. The algorithm permits splitting, merging and deleting of clusters, in order to produce more accurate results and to mitigate dependence of results on the initialization.

The ISODATA algorithm is implemented in the ENVI software [env], where its application for hyperspectral images is straightforward. A vector of spectral values can be used as a feature vector for every pixel. When we have a reference map for the images, we can define a *minimum* number of clusters C_{min} equal to the number of classes in the reference map, and choose a *maximum* number of clusters C_{max} superior to this value.

12.3.2.3. *EM algorithm*

While ISODATA is a deterministic clustering approach, the EM algorithm belongs to the group of statistical algorithms. The underlying assumption for the mixture resolving approach to cluster analysis (that includes the EM algorithm) is that the pixel vectors are drawn from one or several distributions. Most often the individual components of the mixture density are assumed to be Gaussian. In this case, the parameters of a Gaussian mixture model have to be estimated. The EM algorithm was proposed

by Dempster *et al.* [DEM 77] to obtain iteratively a maximum likelihood estimate of the parameters of component densities from the pixel vectors.

To cluster a hyperspectral image by the EM technique, it is assumed that pixels belonging to the same cluster are drawn from a multivariate Gaussian probability distribution. Each image pixel can be statistically modelled by the following probability density function :

$$p(\mathbf{x}) = \sum_{c=1}^{C} \omega_c \phi_c(\mathbf{x}; \boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c) \tag{12.9}$$

where $\omega_c \in [0, 1]$ is the mixing proportion (weight) of cluster c with $\sum_{c=1}^{C} \omega_c = 1$, and $\phi(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the multivariate Gaussian density with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$:

$$\phi_c(\mathbf{x}; \boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c) = \frac{1}{(2\pi)^{B/2}} \frac{1}{|\boldsymbol{\Sigma}_c|^{1/2}} \exp\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_c)^T \boldsymbol{\Sigma}_c^{-1}(\mathbf{x} - \boldsymbol{\mu}_c)\}. \tag{12.10}$$

The parameters of the distributions $\boldsymbol{\psi} = \{C, \omega_c, \boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c; c = 1, 2, ..., C\}$ can be estimated by an iterative method similar to the Classification Expectation Maximization (CEM) algorithm [CEL 92]. The detailed algorithm of parameter estimation for the considered method is given in [TAR a]. During the procedure of parameter estimation, pixels are assigned to the C clusters. Therefore, when the algorithm converges, the partitioning of the set of pixel vectors into C clusters is obtained.

The total number of parameters to be estimated is $P = (B(B+1)/2+B+1)C+1$, where B is a dimensionality of feature vectors. If the value of B is large, P may be quite a large number. This may cause the problem of the covariance matrix singularity or inaccurate parameter estimation results. To avoid these problems, the spectral dimension of pixel vectors in hyperspectral data must be reduced. A simple method of feature extraction applied in [TAR a] consists in averaging every AW neighboring bands, so that :

$$x_{j,b}^{av} = \frac{\sum_{i=1}^{AW} x_{j,[(b-1)AW+i]}}{AW}, \tag{12.11}$$

where $x_{j,i}$ is a value of pixel x_j in the input band i and $x_{j,b}^{av}$ is a value of pixel x_j in the output band b ; $j = 1, ..., n$, $b = 1, ..., B_{av}$, where $B_{av} = n/AW$.

Although being simple, this method has proved its efficiency when dealing with hyperspectral data. A similar method was applied for feature reduction in [JEN 07], where averaging of contiguous spectral bands has been effectuated by applying piecewise constant functions. The method has shown to outperform most of the feature reduction methods proposed in the literature. Furthermore, more complex methods for feature extraction can be considered, such as the PCA, the ICA, the ISOMAP or clustering-based band selection [RIC 99, WAN 06a, WAN 06b, MAR 07].

12.3.2.4. *Segmentation using clustering*

The partitional clustering algorithm produces an exhaustive partitioning of the set of image pixels $\mathbf{X}$ into C clusters. Thus, each pixel has a numerical label of the cluster it belongs to. But as no spatial information is used during the clustering procedure, pixels with the same cluster label can be connected in the image plane, thus forming a spatial region, or they can belong to disjoint regions within the spatial coordinates. Therefore, in order to obtain a segmentation map (where each connected spatial region has a unique label), a connected components labeling algorithm must be applied to the output image partitioning obtained by the clustering algorithm [JAI 99, DRI 08, SHA 02]. This algorithm allocates different labels for disjoint regions in the image plane that were placed in the same cluster.

The obtained segmentation map can be oversegmented, as reported for example in [ACI 03]. However, the final goal is not to obtain the segmentation result, but to define adaptive neighborhoods in order to incorporate them into a spectral-spatial classifier. Therefore, oversegmentation is not a crucial problem, and undersegmentation is not desired. As oversegmentation is preferable to undersegmentation, a 4-neighborhood connectivity is preferable to use while performing the labeling of connected components.

Example 3 : In this example, two described clustering techniques are applied for segmentation of the *University of Pavia* image. For the ISODATA algorithm, considering that image pixels belong to one of the nine classes (based on reference data), we chose the number of clusters as $C_{min} = 9$ and $C_{max} = 10$. The algorithm (ENVI implementation) splits all the pixels into 9 clusters. The EM clustering algorithm is performed with the *maximum* number of clusters $C_{max} = 10$. As explained before, the spectral dimension needs to be reduced before applying the EM algorithm. For this purpose, a 10-band image is obtained by averaging over every 10 neighboring bands[1]. At the output of the EM algorithm, the grouping of the image pixels into 10 clusters is obtained.

Figures 12.7(a) and 12.7(b) show the unsupervised classification maps obtained by the ISODATA and the EM algorithms, respectively. In each of these figures, different colors correspond to different clusters. As seen from the figures, the main spatial structures in the scene are well defined. Based on a visual inspection, the two obtained segmentation results are of comparable accuracies. The obtained unsupervised maps are clearly oversegmented, *i.e.*, there are cases where the regions of pixels belonging

––––––––––––––––

1. The 103-band image was split into 10 groups of 10 bands, the three remaining bands were omitted.

(a)   (b)
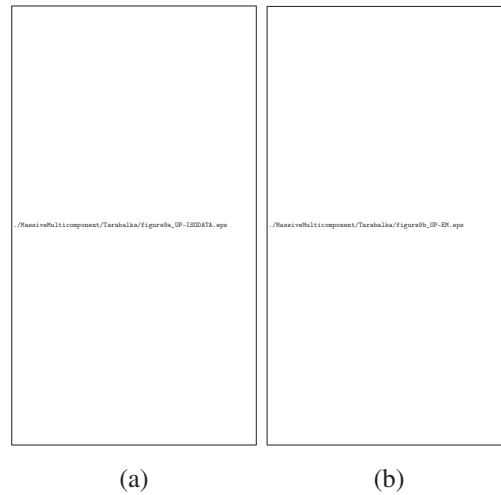
Figure 12.7. *Unsupervised classification maps for the* University of Pavia
image obtained by (a) ISODATA and (b) Expectation Maximization ; figures
are reprinted from [TAR a] ©2009, IEEE.

to the same object are classified to different clusters. The classical connected components algorithm using the union-find data structure [SHA 02] is applied to these two unsupervised classification maps. The resulting segmentation maps contain 20952 and 21450 regions for the ISODATA and the EM techniques, respectively. In both cases, some regions contain a whole single physical object. At the same time, a lot of small, up to one-pixel regions, are present, which explains the large number of regions in the obtained segmentation maps.

12.3.3. *Segmentation Both in the Spatial and Spectral Domain*

In this section, the hierarchical image segmentation algorithm (HSEG) is described, as an example of a segmentation technique working both in the spatial and spectral domain. This technique developed by J. Tilton [TIL 98, TIL 03, TIL 08] is based on iterative hierarchical step-wise optimization (HSWO) region growing method [BEA 89]. Furthermore, it provides a possibility of merging spatially non-adjacent regions by spectral clustering.

The following description of the **HSEG algorithm** is based on the description given in [TIL 03, TIL 08] :

1) Initialize the segmentation by assigning for each pixel a region label. If a pre-segmentation is provided, label each pixel according to the pre-segmentation. Otherwise, label each pixel as a separate region.

2) Calculate the dissimilarity criterion value between all pairs of spatially adjacent regions.

3) Find the smallest dissimilarity criterion value $dissim\_val$ and set $thresh\_val$ equal to it. Then merge all pairs of spatially adjacent regions with $dissim\_val = thresh\_val$.

4) If a parameter $S_{wght} > 0.0$, merge all pairs of spatially non-adjacent regions with $dissim\_val \leq S_{wght} \cdot thresh\_val$.

5) If convergence is not achieved, go to step 2.

When determining most similar pairs of regions, a dissimilarity criterion must be chosen. As described in [TIL 08], a criterion of dissimilarity can be based on vector norms, mean-squared error, entropy, Spectral Angle Mapper (SAM), spectral information divergence.

The optional parameter S_{wght} tunes the relative importance of spectral clustering *versus* region growing. If $S_{wght} = 0.0$, only merging of spatially adjacent regions is performed. If $0.0 < S_{wght} \leq 1.0$, merges between spatially adjacent regions are favored versus merges of spatially non-adjacent regions by a factor of $1.0/S_{wght}$. If $S_{wght} > 0.0$, labeling of connected components has to be applied after HSEG, in order to obtain a segmentation map where each spatially connected component has a unique label.

HSEG gives as output a hierarchical sequence of image partitions. In this structure, a particular object can be represented by several regions at finer levels of details, and can be assimilated with other objects in one region at coarser levels of details. This gives a flexibility to choose an appropriate level of details in the segmentation map. However, the merging of spatially non-adjacent regions is computationally demanding. In order to reduce computational demands, a recursive, divide-and-conquer, approximation of HSEG, called RHSEG, has been developed. The NASA Goddard's RHSEG software provides an efficient parallel implementation of the RHSEG algorithm.

The use of the RHSEG algorithm for segmentation of hyperspectral images has been investigated by Tarabalka *et al.* [TAR 09]. The authors used a SAM between the region mean vectors as the dissimilarity criterion between regions. Good segmentation results has been achieved by using this method.

Example 4 : The RHSEG algorithm is applied to the University of Pavia image, with a SAM between the region mean vectors as the dissimilarity criterion and $S_{wght} =$

(a)   (b)   (c)
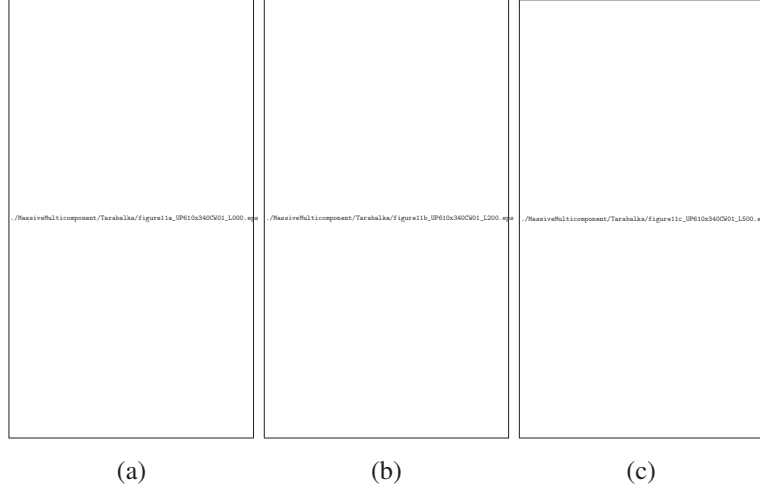
Figure 12.8. *RHSEG segmentation results for the* University of Pavia *image at different levels of details ; $S_{wght} = 0.1$; (a) $L = 0$; (b) $L = 200$; (c)*
$$L = 500.$$

0.1. Figure 12.8 shows segmentation results at different levels of details L (level of hierarchy in the hierarchical sequence of image partitions). Levels of segmentation details L are chosen interactively with the program HSEGViewer [TIL 08]. At each next level of details, regions are merged and the number of regions decreases. Therefore, with the increase of L the risk of undersegmentation increases. Although the oversegmentation is still present at level $L = 500$ (the segmentation map contains 7575 spatially adjacent regions), these segmentation results are accurate enough, and the regions from the segmentation map can further be used as adaptive neighborhoods into a classifier.

12.4. Spectral-Spatial Classification

In order to incorporate spatial information defined by regions from a segmentation map into a classifier, several approaches can be considered. Three major approaches have been used in previous works for spectral-spatial classification of hyperspectral data and we present them in the following.

12.4.1. *Object-Based Classification*

This approach consists in defining for every region a feature/set of features, and then classifying the whole region to one of the classes using this set of features.

Linden *et al.* [LIN 07] applied this approach for a classification of remotely sense hyperspectral images. The authors used the vector mean as a feature for each region in a spectral-spatial classifier. First, they performed the segmentation of a hyperspectral image, based on region growing (using the eCognition software [DAR 03]). Then, a vector mean was computed for each region, such that the value in each band represented the average spectral information of the pixels in this region in the respective band. Afterwards, the regions were classified by an SVM classifier.

Widayati *et al.* [WID p] used the same spectral-spatial classification approach for a multispectral (4-band IKONOS) image. A segmentation map was obtained using the *Merge Using Moments* algorithm [COO 94], and then each region from the segmentation map was classified using its vector mean as a feature and a maximum likelihood classifier.

In both works, the obtained results were mostly not an improvement over those obtained by the pixel-wise classification, *i.e.*, when only spectral information was used.

12.4.2. *Combination of the Spectral and Spatial Information Within a Pixel Feature Vector*

Another approach of spectral-spatial classification consists in combining within a feature vector of each pixel both the spectral and spatial information, and then classifying each pixel using these feature vectors.

Numerous studies explored the use of composite kernels for integration of spectral and spatial information within feature vectors of pixels [LAF 05, CAM 06, CAM 08, FAU 08a, FAU 08b]. The framework of composite kernels, with its application for classification of hyperspectral images, is described in [CAM 06]. The authors present four different kernel approaches for the joint consideration of spectral and spatial information into classification : stacked feature approach ; direct summation kernel ; weighted summation kernel ; cross-information kernel (see [CAM 06] for details). The first two moments were taken into account when building kernels. Furthermore, SVM classification was applied. The use of composite kernels led to the improvement of the classification accuracies when compared to the SVM classification using spectral information only.

Benediktsson *et al.* [BEN 05, PLA] proposed to use extended morphological profiles (EMPs), obtained from the first principal components (PCs) of the hyperspectral

data, as a collection of features for classification. In the resulting EMP, each pixel contains spectral and spatial information. The image was further classified by a neural network [BEN 05] or a SVM classifier [PLA]. Good classification accuracies were reported for this method, when compared to pixel-wise or other spectral-spatial classification techniques.

12.4.3. *Combination of Segmentation and Pixel-Wise Classification Maps*

Accordingly to this approach, a pixel-wise classification, based on the spectral information of pixels, and a segmentation are performed independently. Then, the results of a segmentation and classification are combined, using a defined fusion rule.

Li and Xiao [LI 04] used spectral and spatial information for classification of a multispectral (SPOT 5) image. A watershed segmentation and a maximum likelihood classification were independently performed. Then, pixels of the whole region were assigned to one class if more than 50% of pixels in this region were categorized into one class by a pixel-wise classifier. The classification results were substantially improved with the spectral-spatial approach compared to the pixel-wise classification.

Another, simpler approach, called *majority vote*, is described in the following. Segmentation and pixel-wise classification of an image are independently performed. Then, for every region in the segmentation map, all the pixels are assigned to the most frequent class within this region. Widayati *et al.* [WID p] and Tarabalka *et al.* [TAR 08, TAR a, TAR 09] used this approach for spectral-spatial classification of multispectral and hyperspectral images, respectively. Despite its simplicity, in both cases this approach has shown the best classification performances, when compared with other methods of fusion of spectral and spatial information. In the next section, we use this approach in order to classify hyperspectral images. The use of several segmentation techniques described above is investigated for this purpose. The SVM classifier is used for a pixel-wise classification of the images.

12.5. Experimental results

12.5.1. *Spectral-Spatial Classification of the University Of Pavia Image*

In this section we assess the accuracy of spectral-spatial classification using different segmentation methods in order to define adaptive neighborhoods and majority vote rule for the fusion of segmentation and pixel-wise SVM classification results.

First, segmentation of the *University of Pavia* image is performed, by using the following techniques :

1) Watershed segmentation is performed, using the algorithm of Vincent and Soille [VIN 91]. Two different gradients are used for this purpose :

    - RCMG is computed by (12.4), with $r = 1$ (see Example 1).

    - Sum of morphological gradients of every band (*SumBands*) is computed.

2) Segmentation by partitional clustering is applied, by using the ISODATA and the EM for the Gaussian mixture resolving algorithms, as described in Example 3.

3) The RHSEG is applied, with $S_{wght} = 0.0 \, ; 0.1 \, ; 0.5$. A SAM between the region mean vectors is chosen as the criterion of dissimilarity. Segmentation maps for several levels of hierarchy are chosen interactively (with the program HSEGViewer) for each value of S_{wght}, in order to be used for spectral-spatial classification. In the following, the results leading to the best classification accuracies are shown (that correspond to the hierarchical levels 400, 500 and 200 for $S_{wght} = 0.0 \, ; 0.1 \, ; 0.5$, respectively).

Independently of the segmentation procedure, the pixel-wise classification is performed using the multi-class pairwise (one *versus* one) SVM classifier, with the Gaussian Radial Basis Function (RBF) kernel (by means of the LIBSVM library [CHA 08]). The optimal parameters C and γ were chosen by 5-fold cross validation : $C = 128$, $\gamma = 0.125$. Figure 12.9(a) shows the obtained classification map. Furthermore, the results of pixel-wise classification are combined with the segmentation results, using a *majority vote* rule. Table 12.1 gives the global and class-specific classification accuracies for the pixel-wise SVM and spectral-spatial classification. The following measures of accuracy are used :

– *Overall Accuracy* (OA) is the percentage of correctly classified pixels ;

– *Average Accuracy* (AA) is the mean of class-specific accuracies, *i.e.*, the mean of the percentage of correctly classified pixels for each class ;

– *kappa coefficient* (κ) is the percentage of agreement (correctly classified pixels) corrected by the number of agreements that would be expected purely by chance [RIC 99].

Figures 12.9(b) and 12.9(c) show the classification maps for the spectral-spatial classification, using the segmentation maps obtained by the EM clustering technique and the RHSEG algorithm with $S_{wght} = 0.1$ (these two classification maps correspond to the best accuracies achieved by spectral-spatial classification). In order to compare these results with other works that used an SVM and spatial information for hyperspectral image classification, we have included in Table 12.1 accuracies of mathematical morphology-based classification of the *University of Pavia* image using an SVM, PCs and EMPs (*EMP*) [PLA]. This method was recently proposed by Benediktsson *et al.* [BEN 05] and is considered as one of the most advanced methods for spectral-spatial classification of a multi-band data. Other results of joint spectral-spatial classification of the *University of Pavia* image can be found in [AKS 06, FAU 07, FAU 08b].

(a)   (b)   (c)
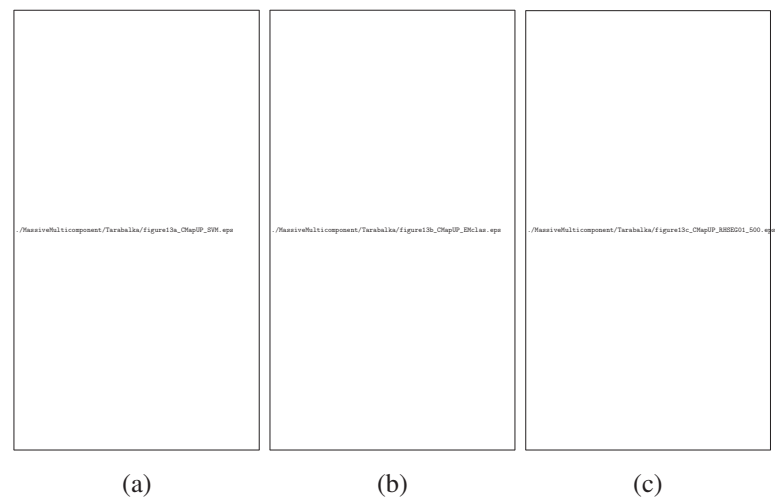
Figure 12.9. *Classification maps for the* University of Pavia *image : (a) Pixel-wise SVM ; (b) SVM + EM ; (c) SVM + RHSEG, with $S_{wght} = 0.1$.*

Tableau 12.1. *Classification Accuracies in Percentage for the* University of Pavia *Image.*

	SVM	+Watershed		+Part.Clustering		+RHSEG			EMP
		RCMG	SumBands	ISODATA	EM	0.0	0.1	0.5	
OA	81.01	85.42	85.04	90.57	93.59	90.00	**93.85**	86.26	85.22
AA	88.25	91.31	90.73	92.48	94.39	94.15	**97.07**	91.00	90.76
κ	75.86	81.30	80.83	87.62	91.48	86.86	**91.89**	82.25	80.86
asphalt	84.93	93.64	93.62	92.16	90.72	73.33	94.77	89.28	**95.36**
meadows	70.79	75.09	74.49	87.07	**92.73**	88.73	89.32	79.09	80.33
gravel	67.16	66.12	64.79	61.43	82.09	**97.47**	96.14	68.37	87.61
trees	97.77	98.56	97.18	98.59	**99.21**	98.45	98.08	98.56	98.37
metal sheets	99.46	99.91	99.64	99.46	**100**	99.10	99.82	99.28	99.48
bare soil	92.83	97.35	97.83	97.38	96.78	98.43	**99.76**	95.23	63.72
bitumen	90.42	96.23	94.29	99.69	92.46	95.92	**100**	94.50	98.87
bricks	92.78	97.92	98.78	98.69	97.80	98.81	**99.29**	96.58	95.41
shadows	**98.11**	96.98	95.97	97.86	97.74	97.11	96.48	**98.11**	97.68

As can be seen from Table 12.1 and Figure 12.9, the SVM classifier gives high classification accuracies. Furthermore, the incorporation of spatial information into a classifier improves the classification accuracies and provides classification maps with more homogeneous regions, compared to pixel-wise classification. All segmentation maps lead to higher classification accuracies. The best global accuracies are achieved when performing spectral-spatial classification using the RHSEG algorithm for image segmentation with $S_{wght} = 0.1$ (*i.e.*, performing segmentation both in the spatial and spectral domain). In this case, the overall accuracy is improved by 12.8 percentage points and the average accuracy improved by 8.8 percentage points compared to the pixel-wise SVM classification. The RHSEG with $S_{wght} = 0.0$, *i.e.*, when the merging of only spatially adjacent regions is possible, leads to slightly lower accuracies. When the merging of spatially non-adjacent regions becomes more favored (for $S_{wght} = 0.5$, classification accuracies are also lower then for $S_{wght} = 0.1$. From here, a conclusion can be made that the best performances are achieved when performing segmentation both in spatial and spectral domain, but the merging of adjacent regions is favored.

Classification using the segmentation maps obtained by partitional clustering techniques also gives very high accuracies, though slightly lower than the best accuracies achieved with the RHSEG technique. Watershed segmentation performs worse than other segmentation techniques for the given data set. From here, the conclusion can be made that the considered image contains classes with mostly dissimilar spectral responses. Therefore, segmentation using cluster-based methods or combination of spatial-based and cluster-based techniques gives accurate results.

Spectral-spatial classification significantly improves accuracies for almost all the classes. Only for the class *shadows* the accuracy of classification is non-significantly reduced or is not changed. This class presents small spatial structures that are in danger to be assimilated with the larger structures in their neighborhood. However, when comparing the classification accuracies for this class obtained with the segmentation techniques working mostly in the spectral domain (the ISODATA, the EM, the RHSEG with $S_{wght} = 0.5$) with those working in the spatial domain, it can be concluded that spectral-based segmentation techniques lead to lower risk of removing small spatial structures than spatial-based techniques. Furthermore, as the HSEG gives a hierarchy of segmentation maps at different level of details, it is interesting to investigate how multiscale analysis can help to preserve small structures in the image.

Finally, when the obtained results are compared with the recent results of spectral-spatial classification, using the SVM and EMPs (see Table 12.1), the majority vote within adaptive neighborhoods gives significantly higher global accuracies and higher class-specific accuracies for most of the classes. Thus, the described spectral-spatial classification method appears to be appropriate for classification of hyperspectral images of an urban area.

(a)   (b)   (c)   (d)
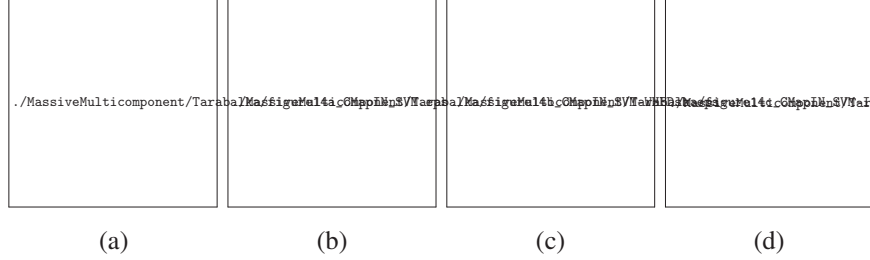
Figure 12.10. *Classification maps for the* Indiana *image : (a) Pixel-wise SVM ;*
(b) SVM + Watershed on the sum of gradients over all bands ; (c) SVM +
ISODATA ; (d) SVM + RHSEG, with $S_{wght} = 0.0$.

12.5.2. *Spectral-Spatial Classification of the Indiana Image*

In the second case study, spectral-spatial classification of the *Indiana* image is performed. This image of a rural area contains more bands (number of bands $B = 220$) and a lower spatial resolution as compared to the *University of Pavia* image.

First, the image is segmented, by using different methods :

1) Watershed segmentation on the RCMG and on the sum of gradients of every band is applied, in the same way as for the *University of Pavia* image.

2) The ISODATA and the EM algorithms are applied to perform a segmentation by partitional clustering. The ISODATA is performed with $C_{min} = 16$ and $C_{max} = 19$. The EM clustering is applied to the 22-band image (obtained by averaging over every 10 neighboring bands of the original data), with $C_{max} = 17$. The grouping of image pixels into 17 clusters is obtained in both cases.

3) The RHSEG is performed, with $S_{wght} = 0.0\,;0.1\,;0.5$ and a SAM between the region mean vectors as the dissimilarity criterion. The results obtained by using the level of details leading to the best classification performances are shown hereinafter ($L = 150\,;100\,;120$ for $S_{wght} = 0.0\,;0.1\,;0.5$, respectively).

Multi-class, one *versus* one, SVM classification is performed on the original image using the Gaussian RBF kernel. The parameters C and γ are determined by 5-fold cross validation giving : $C = 1024$, $\gamma = 2^{-7}$. The resulting classification map is given in Figure 12.10(a). Finally, majority voting within the regions from each of the segmentation maps is performed. Table 12.2 gives classification accuracies for the pixel-wise SVM and spectral spatial classification. Classification maps for the spectral-spatial classification, using watershed on the sum of gradients over all bands, the ISODATA and the RHSEG segmentation techniques are shown in Figure 12.10.

Tableau 12.2. *Classification Accuracies in Percentage for the* Indiana *Image.*

	SVM	+Watershed		+Part.Clustering		+RHSEG		
		RCMG	SumBands	ISODATA	EM	0.0	0.1	0.5
OA	78.76	92.48	**93.78**	88.53	87.25	92.20	92.09	90.73
AA	69.66	77.26	80.53	79.01	70.34	80.49	**80.54**	79.63
κ	75.73	91.39	**92.88**	86.90	85.43	91.06	90.94	89.40
corn-no till	74.59	87.45	**90.24**	79.32	71.65	85.82	85.67	81.80
corn-min till	64.58	88.42	90.68	84.95	84.15	84.82	85.75	**93.08**
corn	58.77	81.04	86.73	75.83	60.66	**91.00**	87.68	81.04
soybeans-no till	69.76	83.26	**89.56**	83.85	82.02	87.86	87.29	80.28
soybeans-min till	79.21	98.47	**99.33**	93.16	95.05	98.69	98.65	97.03
soybeans-clean till	75.41	94.39	**95.48**	85.17	90.05	89.87	88.61	89.87
alfalfa	32.65	**44.90**	**44.90**	12.24	0	2.04	2.04	2.04
grass/pasture	87.05	94.42	93.53	93.08	93.97	94.42	**94.64**	91.96
grass/trees	92.72	**99.11**	**99.11**	94.80	**99.11**	98.37	98.37	98.37
grass/pasture-mowed	29.17	0	45.83	91.67	0	**95.83**	**95.83**	91.67
hay-windrowed	96.37	98.41	99.32	97.51	99.09	99.32	99.32	**99.55**
oats	**22.22**	0	0	16.67	0	0	0	0
wheat	90.58	**99.48**	**99.48**	93.19	98.95	98.43	98.43	95.81
woods	91.07	97.68	97.51	97.17	95.36	**98.45**	96.74	96.48
bldg-grass-tree-drives	65.50	**84.21**	75.44	79.53	69.30	75.73	82.46	77.49
stone-steel towers	84.88	84.88	81.40	86.05	86.05	87.21	87.21	**97.67**

Despite the low spatial resolution of the image, 78.76% of pixels from the test set are correctly classified by the SVM classifier. The classification accuracies for the classes *alfalfa*, *grass/pasture-mowed* and *oats*, which are represented by only a few samples in the training set (further called as *small classes*), are low (less than 33%). Consequently, the average classification accuracy is only 69.66%. Furthermore, incorporation of spatial information from any of the segmentation maps into a classifier improves the classification accuracies. The classification maps obtained by the spectral-spatial classification (see Figure 12.10) are seen to be much less noisy than the one obtained by the pixel-wise classification.

The use of the watershed segmentation map based on the sum of gradients over all bands leads to the the highest overall accuracy (an improvement by 15.0 percentage points when compared to the pixel-wise classification) and kappa coefficient, while the highest average accuracy is achieved by applying the RHSEG algorithm with $S_{wght} =$

0.1 (it is 10.9 percentage points higher than for the pixel-wise classification). The RHSEG with $S_{wght} = 0.0$ also gives very high classification accuracies.

The watershed based on the RCMG gives lower classification accuracies than the one based on the sum of gradients over all bands. This can be explained by the fact that the *Indiana* image contains more bands (220 *versus* 103 bands for the *University of Pavia* image). Furthermore, the vectorial RCMG, based on the Euclidean distance between pixel vectors does not give accurate results due to the curse of dimensionality. In this case, it appears to be more appropriate to compute marginal gradients of every band and to sum them together or to reduce the spectral dimension by performing feature extraction first, and then compute a gradient (as discussed in [TAR b]).

Segmentation maps obtained by partitional clustering techniques lead to lower classification accuracies than the watershed and the RHSEG segmentation. From here, the conclusion can be made that for the *Indiana* image the segmentation methods working in the spatial domain (watershed, the RHSEG with $S_{wght} = 0.0$) appear to give better accuracies than those working in the spectral domain (partitional clustering, the RHSEG with $S_{wght} > 0.0$). This can be explained by the fact that in the *Indiana* image some classes have very similar spectral responses (for instance, three classes of corn, three classes of soybeans). Therefore, for the considered image the methods related on the spectral information (clustering) behave worse while discriminating regions corresponding to different classes than the methods working in the spatial domain. As shown in the previous section, for the *University of Pavia* image cluster-based methods or combination of spatial-based and cluster-based techniques are preferable for segmentation. One of the possible explanations is that the classes in the *University of Pavia* image are more spectrally different than in the *Indiana* image. From the other side, an urban area contains complex and small structures, such as shadows. The clustering techniques differentiate them better, while the techniques working in the spatial domain tend to assimilate them with neighboring structures.

Class-specific accuracies are improved by spectral-spatial classification for almost all the classes, except for the class *oats*. For this class, none of the pixels from the test set is identified correctly by most of the spectral-spatial classifiers. The *oats* class is represented by only two samples in the train set, and a pixel-wise SVM classifier correctly classified four *oats* pixels from the test set. Therefore, when performing segmentation following by majority voting, these four correctly classified pixels risk to be easily assimilated with their neighboring regions. Reasons for the potential misclassification of *small classes* are discussed in [TAR a].

12.6. Conclusion

Hyperspectral imaging provides rich spectral information per pixel, increasing the capability to distinguish physical structures in a scene. However, a large number of

spectral channels presents challenges to image classification. While pixel-wise classi-fication techniques process each pixel independently without considering information about spatial structures, further improvement of classification performances can be achieved by the incorporation of spatial information into classifier, especially in areas where structural information is important to distinguish between classes.

One of the recently proposed approaches of spectral-spatial classification of hy-perspectral imagery consists in performing an image segmentation, in order to use every region from the segmentation map as an adaptive spatial neighborhood for all the pixels within this region. Furthermore, this spatial information must be combined with the available spectral information into a classifier. There are different approaches for fusion of spectral and spatial information for classification. A simple approach leading to good classification performances consists in performing pixel-wise classi-fication and segmentation independently, and then assigning all the pixels of every region to the majority class within this region (majority vote rule). An SVM classifier have shown to be very well suited to classify hyperspectral data. Therefore, it can be used as a pixel-wise classifier in the described spectral-spatial classification scheme.

The choice of an appropriate segmentation technique is important for achieving the best segmentation and classification performances. As was discussed in this chapter and validated by experimental results on two hyperspectral data sets, the spatial-based segmentation techniques appear to be preferable when an image contains classes with similar spectral responses (images of a vegetation area are often a case, when an ac-curate distinction between different species of vegetation is required). If an image contains classes with dissimilar spectral responses, cluster-based segmentation tech-niques give good segmentation results. Furthermore, cluster-based methods can be helpful if an image contains a lot of small and complex spatial structures that face a risk of being assimilated with larger neighboring structures (for instance, remote sensing images of an urban area). The segmentation methods working both in the spa-tial and spectral domain appear to be promising for obtaining accurate segmentation results on different images. The RHSEG segmentation method provides a possibility for choosing the relative importance of spatial-based *versus* spectral-based proces-sing. Furthermore, this technique gives good segmentation results and leads to the best classification accuracies. The RHSEG gives as output a hierarchical sequence of image partitions. In order to achieve the highest classification accuracies, an appro-priate level of details must be chosen, or multiscale analysis can be conducted. The manual selection of an appropriate level is time-consuming. Therefore, the automatic selection is desired. The possibility of the automated selection of results for the RH-SEG was previously explored in [PLA 05]. Furthermore, it is interesting to investigate how multiscale analysis can help in preserving small structures in the image.

12.7. Acknowledgment

[ACI 03] ACITO N., CORSINI G., DIANI M., « An unsupervised algorithm for hyperspectral image segmentation based on the Gaussian mixture model », *Proc. of IGARSS '03*, vol. 6, p. 3745-3747, July 2003.

[AKÕ8] AKÇAY H. G., AKSOY S., « Automatic Detection of Geospatial Objects Using Multiple Hierarchical Segmentations », *IEEE Trans. Geos. and Remote Sens.*, vol. 46, n°7, p. 2097–2111, July 2008.

[AKS 06] AKSOY S., « Spatial Techniques for Image Classification », *C.H. Chen, ed., Signal and Image Processing for Remote Sensing, Taylor & Francis*, p. 491-513, 2006.

[AST 90] ASTOLA J., HAAVISTO P., NEUVO Y., « Vector median filters », *Proc. of the IEEE*, vol. 78, n°4, p. 678–689, 1990.

[BAL 65] BALL G., HALL D., ISODATA, a novel method of data analysis and classification, Rapport, Technical Report AD-699616, Stanford University, Stanford, CA, 1965.

[BAL 08] BALI N., MOHAMMAD-DJAFARI A., « Bayesian Approach With Hidden Markov Modeling and Mean Field Approximation for Hyperspectral Data Analysis », *IEEE Trans. Image Processing*, vol. 17, n°2, p. 217–225, Feb. 2008.

[BEA 89] BEAULIEU J.-M., GOLDBERG M., « Hierarchy in picture segmentation : a stepwise optimization approach », *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 11, n°2, p. 150–163, 1989.

[BEN 05] BENEDIKTSSON J. A., PALMASON J. A., SVEINSSON J. R., « Classification of hyperspectral data from urban areas based on extended morphological profiles », *IEEE Trans. Geos. and Remote Sens.*, vol. 43, n°3, p. 480-491, March 2005.

[BEU 79] BEUCHER S., LANTUEJOUL C., « Use of watersheds in contour detection », in Int. Workshop Image Processing, Real-Time Edge and Motion Detection/Estimation. Rennes, France : CCETT/IRISA, Sept. 1979.

[CAM 05] CAMPS-VALLS G., BRUZZONE L., « Kernel-based methods for hyperspectral image classification », *IEEE Trans. Geos. and Remote Sens.*, vol. 43, n°6, p. 1351-1362, June 2005.

[CAM 06] CAMPS-VALLS G., GOMEZ-CHOVA L., MUNOZ-MARI J., VILA-FRANCES J., CALPE-MARAVILLA J., « Composite kernels for hyperspectral image classification », *IEEE Geos. and Remote Sens. Letters*, vol. 3, n°1, p. 93–97, Jan. 2006.

[CAM 08] CAMPS-VALLS G., GOMEZ-CHOVA L., MUNOZ-MARI J., ROJO-ALVAREZ J. L., MARTINEZ-RAMON M., « Kernel-Based Framework for Multitemporal and Multisource Remote Sensing Data Classification and Change Detection », *IEEE Trans. Geos. and Remote Sens.*, vol. 46, n°6, p. 1822–1835, June 2008.

[CEL 92] CELEUX G., GOVAERT G., « A Classification EM algorithm for clustering and two stochastic versions », *Comput. Stat. Data Anal.*, vol. 14, n°3, p. 315–332, 1992.

[CHA 03] CHANG C.-I., *Hyperspectral Imaging : Techniques for Spectral Detection and Classification*, Kluwer Academic Publishers, Dordrecht, 2003.

[CHA 08] CHANG C., LIN C., « LIBSVM - A Library For Support Vector Machines », Software available at http ://www.csie.ntu.edu.tw/∼cjlin/libsvm, 2008.

[CHE 08] CHEN C. H., PETER HO P.-G., « Statistical pattern recognition in remote sensing », *Pattern Recognition*, vol. 41, n°9, p. 2731–2741, 2008.

[COC 98] COCKS T., JENSSEN R., STEWART A., WILSON I., SHIELDS T., « The HyMap airborne hyperspectral sensor : The system, calibration and performance », *Proc. of First EARSEL Workshop on Imaging Spectroscopy*, October 1998.

[COO 94] COOK R., MCCONNELL I., OLIVER C. J., « MUM (Merge Using Moments) segmentation for SAR images », *Proc. of SPIE.*, vol. 2316, p. 92–103, 1994.

[DAR 03] DARWISH A., LEUKERT K., REINHARDT W., « Image segmentation for the purpose of object-based classification », *Proc. of IGARSS '03*, vol. 3, p. 2039-2041, July 2003.

[DEM 77] DEMPSTER A. P., LAIRD N. M., RUBIN D. B., « Maximum Likelihood from Incomplete Data via the EM Algorithm », *Journal of the Royal Statistical Society*, vol. B 39, n°1, p. 1–38, 1977.

[DID 76] DIDAY E., SIMON J. C., « Clustering analysis », *In Digital Pattern Recognition*, p. 47–94, K. S. Fu, Ed. Springer-Verlag, Secaucus, NJ, 1976.

[DRI 08] DRIESEN J., SCHEUNDERS P., « A Multicomponent Image Segmentation Framework », *Proc. of ACIVS*, 2008.

[env] ITT Corporation, For further information about the ENVI software : http ://rsinc.com/envi/.

[EVA 06] EVANS A., LIU X., « A morphological gradient approach to color edge detection », *IEEE Trans. Image Processing*, vol. 15, n°6, p. 1454-1463, June 2006.

[FAR 05] FARAG A., MOHAMED R., EL-BAZ A., « A unified framework for MAP estimation in remote sensing image segmentation », *IEEE Trans. Geos. and Remote Sens.*, vol. 43, n°7, p. 1617-1634, July 2005.

[FAU 06] FAUVEL M., CHANUSSOT J., BENEDIKTSSON J. A., « Evaluation of Kernels for Multiclass Classification of Hyperspectral Remote Sensing Data », *Proc. of ICASSP'06*, p. II-813 – II-816, May 2006.

[FAU 07] FAUVEL M., Spectral and Spatial Methods for the Classification of Urban Remote Sensing Data, PhD thesis, Grenoble Institute of Technology, 2007.

[FAU 08a] FAUVEL M., CHANUSSOT J., BENEDIKTSSON J. A., « Adaptive pixel neighborhood definition for the classification of hyperspectral images with Support Vector Machines and composite kernel », *Proc. of ICIP'08*, San Diego, USA, 2008.

[FAU 08b] FAUVEL M., CHANUSSOT J., BENEDIKTSSON J. A., SVEINSSON J. R., « Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles », *IEEE Trans. Geos. and Remote Sens.*, vol. 46, n°10, Oct 2008.

[FLO 98] FLOUZAT G., AMRAM O., CHERCHALI S., « Spatial and spectral segmentation of satellite remote sensing imagery using processing graphs by mathematical morphology », *Proc. of IGARSS '98*, vol. 4, p. 1769-1771 vol.4, 6-10 Jul 1998.

[FU 81] FU K., MUI J., « A survey on Image Segmentation », *Pattern Recognition*, vol. 13, n°1, p. 3–16, 1981.

[GOE 03] GOEL P. K., PRASHER S. O., PATEL R. M., LANDRY J. A., BONNELL R. B., VIAU A. A., « Classification of hyperspectral data by decision trees and artificial neural networks to identify weed stress and nitrogen status of corn », *Comput. Electron. Agricult.*, vol. 39, p. 67–93, 2003.

[GRE 88] GREEN A. A., BERMAN M., SWITZER P., CRAIG M. D., « A transformation for ordering multispectral data in terms of image quality with implications for noise removal », *IEEE Trans. Geos. and Remote Sens.*, vol. 26, n°1, p. 65-74, Jan 1988.

[GRE 98] GREEN R. O., EASTWOOD M. L., SARTURE C. M., CHRIEN T. G., ARONSSON M., CHIPPENDALE B. J., FAUST J. A., PAVRI B. E., CHOVIT C. J., SOLIS M. S., OLAH M. R., WILLIAMS O., « Imaging Spectroscopy and the Airborne Visible Infrared Imaging Spectrometer (AVIRIS) », *Remote Sens. Environ.*, vol. 65, n°3, p. 227-248, September 1998.

[GUO 08] GUO B., GUNN S. R., DAMPER R. I., NELSON J. D. B., « Customizing Kernel Functions for SVM-Based Hyperspectral Image Classification », *IEEE Trans. Image Processing*, vol. 17, n°4, p. 622–629, April 2008.

[GUP 06] GUPTA M. R., JACOBSON N. P., « Wavelet Principal Component Analysis and its Application to Hyperspectral Images », *Proc. of IEEE ICIP'06*, p. 1585–1588, Oct. 2006.

[HER 04] HERNÁNDEZ-ESPINOSA C., FERNÁNDEZ-REDONDO M., TORRES-SOSPEDRA J., « Some Experiments with Ensembles of Neural Networks for Classification of Hyperspectral Images », *ISNN (1)*, p. 912-917, 2004.

[HUG 68] HUGHES G., « On the mean accuracy of statistical pattern recognizers », *IEEE Trans. Information Theory*, vol. 14, n°1, p. 55-63, Jan 1968.

[HYV 01] HYVÄRINEN A., KARHUNEN J., OJA E., *Independent Component Analysis*, New York : Wiley, 2001.

[JAI 95] JAIN R., KASTURI R., SCHUNCK B. G., *Machine Vision*, McGraw-Hill series in Computer Science. McGraw-Hill, Inc., 1995.

[JAI 99] JAIN A. K., MURTY M. N., FLYNN P. J., « Data Clustering : A Review », *ACM Computing Surveys*, vol. 31, n°3, p. 264–323, 1999.

[JEN 07] JENSEN A., SOLBERG A., « Fast Hyperspectral Feature Reduction Using Piecewise Constant Function Approximations », *IEEE Geos. and Remote Sens.*, vol. 4, n°4, p. 547–551, Oct. 2007.

[LAF 05] LAFARGE F., DESCOMBES X., ZERUBIA J., « Textural kernel for SVM classification in remote sensing : application to forest fire detection and urban area extraction », *Proc. of ICIP 2005*, vol. 3, p. 1096–1099, Sept. 2005.

[LAM 00] LAMBERT P., MACAIRE L., « Filtering and Segmentation : the specifity of color images », *Int. Conf. Color Graph. Image Process.*, 2000.

[LAN 02] LANDGREBE D., « Hyperspectral image data analysis », *IEEE Signal Proces. Mag.*, vol. 1053-5888, p. 17–28, January 2002.

[LAN 03] LANDGREBE D. A., *Signal Theory Methods in Multispectral Remote Sensing*, John Wiley & Sons, Inc., 2003.

[LEE 05] LEE S., CRAWFORD M., « Unsupervised multistage image classification using hierarchical clustering with a Bayesian similarity measure », *IEEE Trans. Image Processing*, vol. 14, n°3, p. 312-320, March 2005.

[LI 04] LI P., XIAO X., « Evaluation of multiscale morphological segmentation of multispectral imagery for land cover classification », *Proc. of IGARSS'04*, vol. 4, p. 2676-2679, 20-24 Sept. 2004.

[LIN 07] VAN DER LINDEN S., JANZ A., WASKE B., EIDEN M., HOSTERT P., « Classifying segmented hyperspectral data from a heterogeneous urban environment using Support Vector Machines », *Journal of Applied Remote Sensing*, vol. 1, n°1, 013543, SPIE, 2007.

[MAR 07] MARTINEZ-USO A., PLA F., SOTOCA J. M., GARCIA-SEVILLA P., « Clustering-Based Hyperspectral Band Selection Using Information Measures », *IEEE Trans. Geoscience and Remote Sens.*, vol. 45, n°12, p. 4158-4171, Dec. 2007.

[MEY 90] MEYER F., BEUCHER S., « Morphological segmentation », *Journal of Visual Communication and Image Representation*, vol. 1, p. 21–46, 1990.

[MEY 94] MEYER F., « Topographic distance and watershed lines », *Signal Process.*, vol. 38, n°1, p. 113–125, Elsevier North-Holland, Inc., 1994.

[MOU 08] MOUSSAOUI S., HAUKSDOTTIR H., SCHMIDT F., JUTTEN C., CHANUSSOT J., BRIE D., DOUTÉ S., BENEDIKTSSON J. A., « On the decomposition of Mars hyperspectral data by ICA and Bayesian positive source separation », *Neurocomputing*, vol. 71, p. 2194–2208, 2008.

[NOY 07] NOYEL G., ANGULO J., JEULIN D., « Morphological segmentation of hyperspectral images », *Image Analysis & Stereology*, vol. 26, p. 101–109, International Society for Stereology, 2007.

[OSM 07] OSMA-RUIZ V., GODINO-LLORENTE J. I., SÁENZ-LECHÓN N., GÓMEZ-VILDA P., « An improved watershed algorithm based on efficient computation of shortest paths », *Pattern Recognition*, vol. 40, n°3, p. 1078–1090, 2007.

[PAC 03] PACLÍK P., DUIN R., KEMPEN G. V., KOHLUS R., « Segmentation of multi-spectral images using the combined classifier approach », *Image and Vision Computing*, vol. 21, p. 473–482, 2003.

[PEA 00] PEARLMAN J., SEGAL C., LIAO L., CARMAN S., FOLKMAN M., BROWNE B., ONG L., UNGAR S., « Development and Operations of the EO-1 Hyperion Imaging Spectrometer », *Proc. of Earth Observing Systems V, SPIE*, vol. 4135, p. 243–253, 2000.

[PES 01] PESARESI M., BENEDIKTSSON J. A., « A new approach for the morphological segmentation of high-resolution satellite imagery », *IEEE Trans. Geos. and Remote Sens.*, vol. 39, n°2, p. 309–320, Feb. 2001.

[PLA]   PLAZA A., BENEDIKTSSON J. A., BOARDMAN J., BRAZILE J., BRUZZONE L., CAMPS-VALLS G., CHANUSSOT J., FAUVEL M., GAMBA P., GUALTIERI J. A., MARCONCINI M., TILTON J. C., TRIANNI G., « Recent Advances in Techniques for Hyperspectral Image Processing », *accepted for publication, Remote Sensing of Environment*.

[PLA 05]   PLAZA A. J., TILTON J. C., « Automated selection of results in hierarchical segmentations of remotely sensed hyperspectral images », *Proc. of IGARSS '05*, vol. 7, p. 4946–4949, July 2005.

[RIC 93]   RICKARD L., BASEDOW W., ZALEWSKI E., SILVERGLATE P., LANDERS M., « HYDICE : An Airborne System for Hyperspectral Imaging », *Proc. of Imaging Spectrometry of the Terrestrial Environment, SPIE*, vol. 1937, p. 173–179, 1993.

[RIC 99]   RICHARDS J. A., JIA X., *Remote Sensing Digital Image Analysis : An Introduction*, Springer-Verlag New York, Inc., 1999.

[SCH 01]   SCHEUNDERS P., « Multivalued image segmentation based on first fundamental form », *Proc. of the 11th International Conference on Image Analysis and Processing*, p. 185-190, September 2001.

[SHA 02]   SHAPIRO L., STOCKMAN G., *Computer Vision*, Prentice Hall, 2002.

[SOI 96]   SOILLE P., « Morphological partitioning of multispectral images », *Journal of Electronic Imaging*, vol. 5, n°3, p. 252-265, SPIE, 1996.

[SOI 03]   SOILLE P., *Morphological Image Analysis*, Springer-Verlag, 2 édition, 2003.

[Ste 05]   STEVENSON B., O'CONNOR R., KENDALL W., STOCKER A., SCHAFF W., HOLASEK R., EVEN D., ALEXA D., SALVADOR J., EISMANN M., MACK R., KEE P., HARRIS S., KARCH B., KERSHENSTEIN J., « The Civil Air Patrol ARCHER hyperspectral sensor system », *Proc. of Airborne ISR Systems and Applications* II*, SPIE*, vol. 5787, p. 17–28, 2005.

[STO 00]   STOEV S., « Rafsi - a Fast Watershed Algorithm Based on Rainfalling Simulation », *Proc. of the 8th International Conference on Computer Graphics, Visualization, and Interactive Digital Media*, 2000.

[SUB 97]   SUBRAMANIAN S., GAT N., SHEFFIELD M., BARHEN J., TOOMARIAN N., « Methodology for hyperspectral image classification using novel neural network », *Proc. of SPIE.*, vol. 3071, Algorithms for Multispectral and Hyperspectral Imagery III, A. Evan Iverson ; Sylvia S. Shen ; Eds., p. 128–137, August 1997.

[TAD 98]   TADJUDIN S., LANDGREBE D., Classification of High Dimensional Data with Limited Training Samples, PhD thesis, School of Electrical and Computer Engineering, Purdue University, May 1998.

[TAR a]   TARABALKA Y., BENEDIKTSSON J. A., CHANUSSOT J., « Spectral-Spatial Classification of Hyperspectral Imagery Based on Partitional Clustering Techniques », *accepted for publication, IEEE Trans. Geos. and Remote Sens.*

[TAR b]   TARABALKA Y., CHANUSSOT J., BENEDIKTSSON J., « Segmentation and classification of hyperspectral images using watershed transformation », *under review, Pattern Recognition*.

[TAR 08] TARABALKA Y., CHANUSSOT J., BENEDIKTSSON J. A., ANGULO J., FAUVEL M., « Segmentation and Classification of Hyperspectral Data Using Watershed », *Proc. of IGARSS '08*, Boston, USA, 2008.

[TAR 09] TARABALKA Y., BENEDIKTSSON J. A., CHANUSSOT J., « Classification of hyperspectral data using Support Vector Machines and adaptive neighborhoods », *Proc. of the 6th EARSeL SIG IS workshop*, Tel Aviv, Israel, 2009.

[TIL 98] TILTON J., « Image segmentation by region growing and spectral clustering with a natural convergence criterion », *Proc. of IGARSS '08*, vol. 4, p. 1766–1768, 1998.

[TIL 03] TILTON J., « Analysis of hierarchically related image segmentations », *IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data*, p. 60–69, 2003.

[TIL 08] TILTON J. C., « HSEG/RHSEG, HSEGViewer and HSEGReader user's manual (Version 1.40) », *Provided with the evaluation version of* RHSEG *available from : http ://ipp.gsfc.nasa.gov/RHSEG*, 2008.

[VAI 03] VAIPHASA C., « Innovative genetic algorithm for hyperspectral image classification », *Proc. of the International Conference Map Asia*, vol. 20, 2003.

[VIN 91] VINCENT L., SOILLE P., « Watersheds in digital spaces : an efficient algorithm based on immersion simulations », *IEEE Trans. Pattern Analysis and Machine Intel.*, vol. 13, n°6, p. 583–598, June 1991.

[WAN 06a] WANG J., CHANG C.-I., « Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis », *IEEE Trans. Geos. and Remote Sens.*, vol. 44, n°6, p. 1586-1600, June 2006.

[WAN 06b] WANG X. R., KUMAR S., RAMOS F., KAUPP T., « Probabilistic Classification of Hyperspectral Images by Learning Nonlinear Dimensionality Reduction Mapping », *Int. Conf. on Information Fusion*, p. 1-8, July 2006.

[WID p] WIDAYATI A., VERBIST B., MEIJERINK A., « Application of combined pixel-based and spatial-based approaches for improved mixed vegetation classification using IKONOS », *Proc. of the 23th Asian conference on remote sensing*, 2002, 8 p.

[YAN 99] YANG H., MEER F. V. D., BAKKER W., TAN Z. J., « A back-propagation neural network for mineralogical mapping from AVIRIS data », *International Journal of Remote Sensing*, vol. 20, n°1, p. 97–110, 1999.

[ZHO 05] ZHOU H., MAO Z., WANG D., « Classification of coastal areas by airborne hyperspectral image », *Optical Technologies for Atmospheric, Ocean, and Environmental Studies. Proc. of SPIE*, vol. 5832, p. 471–476, May 2005.

Chapitre 13

Parallelizing Image Analysis Applications for Spectral Microscopy

13.1. Opportunities in Spectral Microscopy

Spectral imaging combines two technologies that have independently evolved for the last several decades, namely spectroscopy and imaging. Systems that leverage the two technologies simultaneously have been used in the fields of astronomy [KIM 03], remote sensing [KER 02], and chemical compound analysis [KUL 03] to identify the composition and characteristics of celestial, terrestrial, and atmospheric elements. In recent years, spectral imaging has also been used in biomedical applications such as spectral karyotyping [WEI 05], general cell visualization [MAI 03], cell trafficking of variously colored fluorescent proteins [HAR 02], and differentiation of pathologies [ROY 95]. Spectral imaging systems developed to facilitate various biomedical applications have typically been coupled with traditional microscopes to enable spectral microscopy. Spectral microscopes provide the combined benefits of spectroscopy and imaging microscopy, resulting in the ability to acquire spectral images of microscopic specimen.

Spectral images are three dimensional (3D) cubes of data $I_{x,y}(\lambda)$ comprised of a series of two dimensional (2D) images $I_{x,y}$, one for each wavelength, λ. Each point within the spatial extent of the image, (x, y), is comprised of a spectral pixel, corresponding to a vector of intensity values, where the dimension of the vector is equivalent to the size of the third dimension of the data cube. In other words, each pixel location (x, y) contains a spectral signature stored along the λ axis corresponding to the spatial

_____

Chapitre rédigé par E. GABRIEL, S. SHAH.

location in the image. This representation is as depicted in figure 13.1. There are primarily three different ways that spectral images may be stored for digital processing. The most common approach is Band-Sequential (BSQ) where each wavelength is stored as a separate image and the entire set of images is stored as a cube [BEA 03]. Other mechanisms include the Band-Interleaved-Pixel (BIP) and the Band-Interleaved-Line (BIL) formats. In BIP, all corresponding wavelength intensity values belonging to each pixel location are stored in a sequential fashion. BIP is advantageous over BSQ in that it allows access to the spectral values of an individual pixel without having to read all of the image. Finally, BIL is an interleaved format that stores all spectral values associated to each scanline of an image in a sequential order.

./MassiveMulticomponent/gabriel/figs/SpectralCube.eps

Figure 13.1. *Graphic representation of Spectral Image*

Spectral microscopes capable of image acquisition combine different mechanisms of specimen excitation and spatial scanning. As defined in [LEV 06], they can be classified as wavelength-scan, spatial-scan, or time-scan methods. Wavelength-scan approaches are by far the most common, allowing for a series of 2D image acquisition, one for each wavelength of interest. Hence, the spectral data cube is populated one planar image at a time and corresponds directly to the BSQ image storage format. This mode of imaging is enabled mostly through the use of optical filters placed in the path of specimen excitation or in the path of image sensing. However, the use of optical filters can be limiting such that only a limited number of wavelengths can be measured due to the restriction on the number of filter wheel slots available on most microscopes. A more flexible method is through the use of tunable filters such as Circular-Variable Filters (CVF), Liquid-Crystal Tunable Filters (LCTF), or the Acousto-Optical Tunable Filters (AOTF) [WYA 75, MIL 91, GAT 00]. Each of these filters offer a broad range of selectable wavelengths for specimen examination. The limiting factor in this case is a relatively low transmission efficiency and fixed spectral resolution for a given wavelength [LEV 00]. Another alternative for wavelength-scanning is the use of a

monochromator for specimen illumination. A monochromator uses a prism or grating to emit narrow band of light. It is a mechanical device with many moving parts, and thus limits the speed at which one can change the wavelength of emitted light. On the other hand, it offers the user the ability to achieve the desired trade-off between the bandwidth of emitted light and the amount of light intensity desired for specimen examination. This offers significant advantage in application that require imaging with low sensitivity [LEV 00].

Spatial scanning methods are akin to line-scan imaging wherein the spectrum of an entire line of pixels is acquired simultaneously. This mode of imaging uses prisms or gratings to spread the light onto a 2D imaging sensor (e.g. CCD) such that all columns record a single row of pixels from the specimen, but each element of the column corresponds to a different wavelength. Hence, images are captured as $I(x, \lambda)$ and scanned across y. The scanning across y is typically performed by physically moving the specimen to be imaged. The spectral image format most suited for this mode of imaging is BIL.

The final alternative to the above two scanning methods is time-scan techniques that rely on data transformations to acquire spectral images. One such approach is Fourier spectroscopy [BAY 02, MAL 96], which does not require the use of any filters or spectral light sources, instead exploits principles of light interference to deconvolve the measured spectrum. This approach uses an interferometer that splits a beam of light generating an optical path difference (OPD). The result is a time delay in the travel between the two beams. As the beams travel through the sample, they are recombined to interfere at the detector. Multiple OPDs are generated and intensity corresponding to each is recorded, resulting in a pattern known as the interferogram. Fourier transform of the interferogram gives the spectral image. While the method is elegant and theoretically well-founded, it suffers from high computational complexity due to the required data transformations as well as an inability to select specific wavelengths of interest since the entire spectrum must be captured to acquire any portion of it.

Independent of the mode of acquisition, a spectral image contains thousands of spectral pixels. The data files generated are multidimensional and large in size, necessitating innovative mechanisms for analysis and visualization. Many modern data and image processing algorithms are relevant, and in some cases, adequate, for spectral image processing [CAR 03]. Some applications may require the design of new algorithms capable of leveraging both the spectral and image information [SCH 05, SHA 05]. In the following section, we introduce an application of spectral microscopy in the context of clinical cytology and cancer screening. The potential benefit of spectral imaging is presented along with challenges associated with design and development of computer systems capable of processing acquired images.

13.1.1. *Example Application : Thyroid Fine Needle Aspiration Cytology*

Cancer continues to remain a major health problem in the United States, with one of two men and one of three women developing cancer in their lifetime. Among various cancers, thyroid nodule is a common cancer of the thyroid gland. It has been estimated that up to 20% of the world population and approximately 50% of 60-year-old persons have palpable thyroid nodule or nodules [HEG 04]. The clinical spectrum ranges from the incidental, asymptotic, small, solitary nodule, in which the exclusion of cancer is a major concern, to the large, partly intrathoracic nodule that causes pressure symptoms, for which treatment is warranted regardless of cause. In spite of the growing incidences of thyroid lesions, the rate of thyroidectomies is on the decline. Early detection of the disease has been partly responsible for improved outcomes. Among screening and detection procedures, fine needle aspiration (FNA) is believed to be a safe, inexpensive, and minimally invasive procedure to diagnose tumors [GHA 93]. For cytological evaluation of FNA samples, smears are appropriately prepared and stained. Typically, the stain changes the color of the cells and tissue so that examination of the smear under standard microscopes with moderate magnification (20–40x) is sufficient for clinical evaluation.

With the advances in imaging technology, there is considerable interest in automated analysis of FNA cell smears that could help to reduce the time required for manual screening and increase the detection rate of abnormalities [ASS 03]. Several commercial products, such as ScanScope from Aperio Technologies, DX-40 from DMetrix, Inc., and iScan from BioImagene, Inc., have been developed to automate the process of digitizing microscope slides. They provide high throughput capabilities to digitize cell smears, resulting in a stitched image of scan areas of the order of 1.5cm x 1.5cm in less than 5 minutes. This provides a single image per smear that can be as large as 60,000 x 60,000 pixels under 40x magnification (resolution of 0.25μm/pixel). More recently, multispectral microscopes capable of acquiring spectral images under transmitted illumination have also been used to digitize and analyze cell smears [SHA 08, FEN 07]. Spectral imaging allows for the simultaneous measurement of spectral and spatial information of a sample such that the measurement of the spectral response at any pixel of a two-dimensional image is possible. Figure 13.2 shows a single field-of-view of a thyroid FNA cell smear. The image on the left is acquired using a standard brightfield optics while the one of the right shows four spectral bands acquired using an upright microscope attached to a monochromator for narrowband specimen illumination.

Studies have shown that biological tissue exhibits unique spectra in transmission. By exploring the spectral differences in tissue pathology, many chemical and physical characteristics not revealed under traditional imaging systems can be realized and used to improve the analysis efforts. Several efforts have already resulted in algorithms for cell segmentation, morphometric and karyometric feature analysis, as well as computer assisted diagnosis (CAD), with cell segmentation being the most challenging step for automated systems. However, to our knowledge, most of these efforts have been

./MassiveMulticomponent/gabriel/figs/SpectralFNAC.eps

Figure 13.2. *A single field-of-view of a thyroid fine needle aspirated cell smear showing both the brightfield image (left) and four spectral bands (right).*

relatively limited in size due to the large data size and the computational bottlenecks. It is not uncommon to acquire anywhere from 5 to 31 spectral channels for each sample. Considering an average size of the smear to be 1.0cm x 1.0cm, the image cube to be analyzed would be approximately 8GB to 50GB in size, depending on the number of spectral bands. This creates difficulties in analyzing the entire data set on a standard desktop. On the other hand, for any CAD system to be useful in clinical pathology, the time taken for analysis has to be minimized. This necessitates that an acceptable analytical solution maximize its ability to leverage modern computing systems.

13.2. Opportunities in Computing

Since more than two decades micro-processor developments are often characterized by Moore's Law [MOO 65]. Gordon E. Moore, one of the co-founders of Intel, observed that the number of transistors that can be placed on a integrated circuit doubles

every two years. From the perspective of the average end-user, one of the main implication of this law was/is, that each generation of micro-processors had a significant number of new features and thus better performance compared to the previous generation. In the nineties, most of the features introduced with a new processor generation focused on improving the performance of a single application process or thread. The most prominent representatives of this era are the introduction of dynamically scheduled processors which allow for the out-of-order execution of instructions, predicting the outcome of branches in the application and speculative execution of subsequent instructions, the ability to issue multiple instructions per cycle (e.g. superscalar architectures, very long instruction word architectures), and adding functional units capable of vector instructions (e.g. MMX, SSE, Altivec). Each of these features required a significant amount of logic and additional transistors on the processor. Combined with the ever increasing frequency of microprocessors, end-users got accustomed to the notion, that whenever their application faces performance problems, purchasing the most recent computer will increase the performance of their (sequential) code without requiring any modification to the application and therefore without investing any efforts.

Most processors available today run at a lower frequency than the systems available a couple of years ago. One of the main reasons for this is the power consumption of micro-processors, which increases with the frequency of the processor. At the same time, all major processor manufacturers acknowledged, that the performance of a single application cannot (easily) be further improved. The main reason for this is that the amount of parallelism available in a single instruction stream is fundamentally limited, and thus most of the functional units of a processor idle most of the time. As a result, architects focus on finding instructions from multiple instruction streams to execute concurrently, e.g. through the utilization of simultaneous multi-threading techniques (e.g. Intel's hyperthreading [MAR 02]), and through the usage of multiple compute cores on a single die.

As of today, all processors available in commodity systems are multi-core processors, which pose a major challenge to the average end-user : purchasing a new computer today will in fact give the user more compute cores and therefore more computer power compared to its predecessors. His (sequential) application will however not run any faster on such a system, since it is not able to utilize more than one compute core at a time. In order to fully exploit such a multi-core processor, the end-user will have to modify his code and explicitly denote the parallelism available in the application.

In the following, we would like to discuss some of the current developments in the processor market which might have potentially significant implications for the development of high performance applications for multi-spectral image analysis : homogeneous multi-core processors, heterogeneous multi-core systems potentially using compute accelerators, and clustered systems. We will conclude this section with a brief discussion of programming models for these systems.

13.2.1. *Homogeneous Multi-Core Processors*

Today's high end computing systems rely mostly on multi-core and multi-threaded processor technologies. In the TOP500 list of June 2008, a web page summarizing the 500 most powerful computer systems in the world, 56% of all machines listed use quad-core processors [TOP 08] and the vast majority of the reamining machines utilize dual-core processor technology. However, compilers and end-users perfoming processor-specific optimizations have to take into account the (significant) differences between the Intel, AMD, IBM, or SUN multi-core architectures in areas such as cache organization, connectivity between the cores, and organization of the I/O channels on the chip.

Consider as an example two of the most popular quad-core processor available today, the Intel Xeon processor (Tigerton) and the AMD Opteron processor (Barcelona). Figure 13.3 shows the basic architecture of a compute node consisting of four Intel Xeon processors, resulting in 16 compute cores in that system, and its AMD Opteron counterpart. Comparing a single chip with four cores each, the most striking difference between the two architectures is that the Intel processor offers a non-uniform connectivity between its four compute cores. The processor contains two pairs of two compute cores which are connected through the Level 2 cache (L2). This architecture has a consequence for the application developers in that it is sensitive to the location on where two application processes/threads are running. Consider a multi-threaded application which executes two concurrent threads. In case both threads run within the same pair of cores, the access to shared data items between the threads will be through the Level 2 cache. In case the threads are running however on cores which are not connected by the L2 cache, the synchronization of the shared data items will occur through the main memory, which will lead to a significantly lower performance compared to the first case. This problem is known as the thread affinity or process affinity problem. The AMD architecture does not have this hierarchy in the interconnect of the cores within a single die : all four cores of the Barcelona processor are connected through the L3 cache, making them insensitive to the thread-affinity problem described above. All modern operating systems offer mechanism to influence the location of a thread or process by overriding the default selection of the operating system.
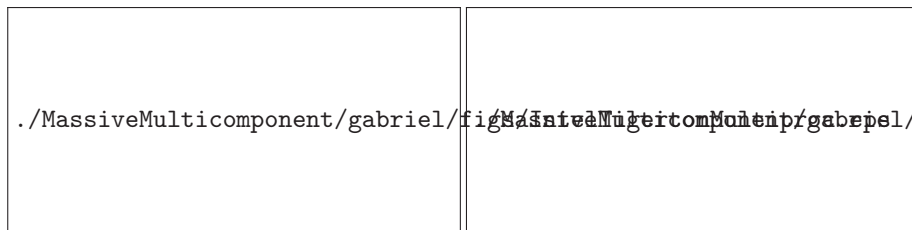
Figure 13.3. *Block Diagram of a multi-processor quad-core Intel Xeon (left) and a multi-processor quad-core AMD Opteron node (right).*

A multi-processor node containing more than one quad-core processor exposes however an entirely different picture. While the Intel Tigerton architecture offers a symmetric shared memory view, a system consisting of multiple AMD Opteron processors is organized as a Non-Uniform Memory Architecture (NUMA). In the latter case, each processor has its own main memory controller and manages a portion of the overall main memory. Although the address space is divided across all processors, every process still has access to any memory location in the system. However, depending on whether the data item is locally available in the main memory of the processor where the according process is executed, or whether it has to be accessed from the main memory of another processor, the costs for accessing the data item will vary significantly. This problem is often referred to as the memory affinity problem. Note, that both multi-processor configurations expose the process and thread affinity problem.

Comparing the performance of systems utilizing these two processors for various applications, such as reported in [BAR 08], reveals that the AMD Opteron system typically outperforms the Intel system for memory intensive applications. The NUMA architecture allows to utilize multiple memory controllers and memory banks in parallel, and thus has a higher overall data throughput between the main memory and the compute cores. On the other hand, applications that are not limited by the memory bandwidth often perform better on the Intel system due to the lower communication costs between the compute cores and the higher frequency of the Intel processors.

13.2.2. *Heterogeneous Multi-Core Systems*

In recent years, heterogeneous multi-core systems have emerged, many of them focusing on graphical and image processing applications. Most heterogeneous systems use compute accelerators in a commodity of-the-shelf PC, such as Clearspeed cards [Cle 08], Nvidia GPGPUs [WAS 07] or the Intel Larrabee [SEI 08] chipset. These accelerators offer tremendous performance benefits compared to regular CPUs for a small but relevant subset of compute operations.

The most widespread accelerator in image processing is the Nvidia G80 chip available on many NVIDIA graphics cards. The chip contains 128 *stream processors* (SPs). The 128 SPs are divided into groups of 16 SPs each, each group of SPs is connected by a separate L1 cache. The SPs have access to their own dedicated main memory, which guarantees very high data transfer rates. The G80 shows very good performance for operations which can utilize the large number of SPs efficiently, such as fast-fourier transforms or certain vector instructions. The main problem with using the G80 processor efficiently is its programing. Since the SPs have a different instruction set architecture than the regular CPU of the PC, the application developer has to create executables for each code section that shall be executed on the GPU and

launch the executable. Although these characteristics are abstracted into the programming language supported by the GPU (CUDA [THO 08]) and is a common approach in embedded systems, it is counterintuitive for many C/C++ programmers.

The IBM Cell processor [JOH 07] takes the heterogeneous computing aspects to the next level by providing a single processor that has different types of compute cores already built in. Originally targeting the Playstation console of Sony, the processor has gained popularity due to its overwhelming performance characteristics. *Roadrunner*, the first computer in the world which has a sustained performance of more than a PetaFLOP (10^{15} floating point operations per second), uses a combination of the Cell and AMD Opteron processors.

The Cell processor consists of a general purpose processor element (PPE) which is derived from the IBM PowerPC processor, and eight synergistic processor elements (SPE). The main function of the PPE is to coordinate the work of the SPEs. The SPEs, similar to the SPs in the Nividia GPU, are simplified processors optimized for a subset of the regular compute operations. Each SPE acts as a vector unit, which leads to optimal performance for instructions operating on multiple data items. The overall performance of a single Cell processor can reach up to 204.8 GFLOPS for single precision operations, or 102.4 GFLOPS for double precision operations. This outperforms a 2.0 GHz quad-core AMD Opteron processor by up to a factor of 6, which can achieve a theoretical peak performance of 32 GFLOPS. Similar to the G80 however, the main drawback with the Cell processor is that it requires programmers to significantly modify their source code in order to utilize the SPEs.

13.2.3. *Clustered System*

A complementary approach to adding more processors into a single PC is to combine multiple PCs to a compute cluster. There are two benefits of clustering PCs compared to further increasing the number of processors in a single system. First, certain internal resources such as the memory bandwidth scale with the number of PCs. Although scaling the memory bandwidth can also be achieved by the NUMA approach described above, a NUMA architecture is still sensitive to performance degradations due to cache coherence problems. Since each PC has its own independent main memory, a PC cluster provides a distributed memory architecture, which is not vulnerable to cache coherence problems. Second, a PC cluster is more robust than a single PC : the failure of a single compute node will not affect the other nodes in the cluster. This is not the case for 'fat' multi-processor PC with a large number of cores and CPUs. A shared memory system only runs a single instance of the operating system. Therefore, if a single compute core crashes, the entire PC/node will go down.

In order to efficiently utilize a PC cluster, the programmer has to modify once again his application. The main modification required is, that the application has to

be broken up into multiple pieces which can be executed concurrently. Typically, this happens either on the data level, i.e. each process works on a different portion of the overall data set, or on the functional level, i.e. each process performs a different operation on the data items. There are two system aspects which have to be taken into consideration in order to utilize a clustered system efficiently. The first aspect deals with the network interconnect used in the cluster, the second with accessing large files in a distributed memory environment.

The most widely utilized network interconnect available today is Gigabit Ethernet(GE). Most systems have one or more GE cards already on the motherboard, and GE switches are available in all sizes and price ranges. From the performance perspective, the network latency offered by a good GE switch using TCP/IP is in the range of $50\mu s$. The theoretical bandwidth of a GE link is 1 GBit/s. Most applications observe however an end-to-end bandwidth in the range of 80-100MBytes/sec due to TCP and Ethernet protocol overheads. More sophisticated network interconnects available today include *InfiniBand* [I.T 04], *Myrinet* [Myr 08], or *Quadrics* [Qua 08]. These networks, although significantly more expensive than GE, offer a much lower network latency and much higher bandwidth. As an example, 4xInfiniBand gives the user an end-to-end latency in the range of $3.5\mu s$ and a sustained bandwidth of up to 1 GByte/s. For many applications, the quality of the network interconnect will be the decisive factor in whether they can use a PC cluster efficiently.

Of similar importance for applications working with large files is the storage and the file system utilized in the cluster. Assuming that the storage provides sufficient bandwidth for the application, the main problem is how to give processes running on different nodes on the PC cluster access to the storage. The most popular solution today is to utilize the Network File System (NFS) [SAN 85]. NFS uses a client-server approach. File I/O requests such as read or write operations are forwarded from the compute nodes of the cluster to the main file server. The NFS server executes the required read/write operations and forwards the result of the operation to the client. The main problem with utilizing NFS in a clustered environment is however, that the NFS daemon on the server side can become a bottleneck, since a single NFS daemon serves all I/O requests occurring in the cluster. Furthermore, NFS is not very efficient when handling I/O requests from multiple processes to the same file, since this scenario was considered an unlikely event in the design of the NFS protocol. It is however highly relevant for the targeted applications discussed in this chapter, since a single image stored in a single file is being processed by multiple processes simultaneously. In order to overcome these limitations, one should use one of the parallel file systems available, such as *PVFS2* [CAR 00], *Lustre* [Lus] or *GPFS* [PRO 01], which offer much higher performance in this scenario than NFS.

13.2.4. *Parallel Programming Models*

As mentioned previously, the main modification required to utilize a cluster or a multi-core processor is, that the application has to be broken up into multiple pieces. Most typically this happens on the data level, e.g. such that one process/thread works on the first part of the image, while the second process/thread works on the second part, etc. An important question arising for the parallelization of an application is whether to utilize a multi-threaded or a multi-process approach. The answer depends on the target architecture and some characteristics of the algorithms used. For the rest of this chapter, we will focus on a multi-process approach for two reasons : first, this approach can be utilized on a single PC having multiple compute cores as well as on a clustered environment. A multi-threaded approach on the other hand is strictly limited to a single PC/shared memory environment. Second, various analysis have shown, that even on a single PC, the multi-process approach often outperforms multi-threaded applications. The main reasons for this are the absence of cache coherence problems due to the separate address space of different processes.

The downside of using a multi-process approach is twofold. First, multi-process applications tend have a non-negligible memory overhead compared to multi-threaded applications, since many data items have to be replicated on all processes. Second, a multi-process approach requires rewriting the entire application for this model. In contrary to that, some multi-threaded approaches, e.g. OpenMP [CHA 08], allow for incremental parallelization of the sequential code. The initial investment in the multi-process approach is therefore significantly larger than that for the multi-threaded parallelization.

The dominant model for parallel multi-process applications as of today utilizes the Message Passing Interface (MPI) [Mes 97]. MPI specifies an Application Programming Interface (API) for data exchange between different processes. Compared to other communication interfaces such as sockets, it offers various advantages including

– Process startup and management capabilities ;

– Notion of data types for the data transfer routines, which allow an MPI application to be deployed in heterogeneous environments without requiring the application to perform the data conversion ;

– Functions supporting group communication operations, which abstract and optimize often occurring communication patterns in parallel applications.

There are a number of commercial and public domain MPI libraries available for virtually all operating systems used on commodity PCs and PC clusters. An application using MPI for data exchange between the processes works on all available MPI implementations, although the code has to be recompiled whenever a new MPI library is going to be used. Thus, MPI allows for source code portability, and not binary

compatibility between different implementations. For further information on programming using MPI please refer to the numerous books discussing this topic, such as [GRA 03, QUI 04].

In the remainder of this chapter, we focus on the problem of image segmentation as related to the application of spectral microscopy discussed in section 13.1.1. We present a pattern theoretic solution to the problem and give details on how to parallelize the most important components of the code in order to utilize a PC cluster and a multi-core processor efficiently.

13.3. Cell Segmentation in Spectral Images of Thyroid FNAC Smears

Diagnostic decisions in the evaluation of cytological specimen are guided by the measurement of cell properties and their spatial distribution. Specifically, cell area, radius, convexity, etc. and specimen topology provide critical information in the diagnostic process [WU 98]. Hence, the ability to segment and delineate cell boundaries forms an important step for any CAD system. A multitude of methods for automatic cell segmentation in a variety of image modalities have been proposed over the last three decades [CAS 79, WU 00, WU 98], including methods for immuno-stained cytological specimen images [MEA 02]. Nonetheless, cell segmentation remains a problem of interest due to the complexity of cell structures observed in images. Cytological smears, particularly ones obtained from fine-needle aspiration biopsy are often difficult to segment as they are contaminated with non-cellular debris. Moreover, the fine textured nature of the cell chromatin gives it a non-uniform intensity appearance in both color and gray images. The existence of dense intracellular as well as intercellular matter in the target cells makes it harder to draw a fine contour delineating the cell boundary without overshooting it.

Among the various solutions developed, the most common approach to cell segmentation is based on thresholding [SAH 88]. Such an approach is semi-automated at best and requires manual adjustment of threshold values due to variations in the staining process that result in different intensity distributions for different images. Methods that utilize local image information ranging from regions, edges, histograms, to clusters have also been developed [LIE 95]. Edge or gradient based approaches assume that discontinuities in the image intensity signify object boundaries. In general, edge and histogram based approaches have proved to be sensitive to noise and fail to provide good delineation of the cell boundary due to lack of contrast between cells and the background. To address this problem, edge based approaches have been coupled with parametric techniques such as Hough transforms [SER 95]. Their success has also been limited due to problems of noise and weak edges in typical smears. Morphological approaches that rely on region growing, splitting, and merging algorithms tend to be less sensitive to noise. As such, watershed based methods [VIN 91] have been extensively explored in this context. However, due to its iterative nature,

such approaches tend to be computationally expensive. In addition, watershed methods are known to suffer from the problem of over-segmentation. Heuristic methods to address this limitation have been proposed for specific domains [ADI 01, CHE 06]. Other methods including active contours, Fourier and Hough transforms, and neural networks [BAM 98, KUR 98] have also been developed. More generalized approaches relying on pattern theory coupled with clustering techniques have been exploited where pixels are assigned to either "cell" or "non-cell" class [HAR 79, FAU 80]. More recently, textural features have been used for cell segmentation, especially for cytological and histological samples [FER 98, YOG 96]. Texture in an image can be obtained using a multitude of methods ranging from gray-level co-occurrence matrices (GLCM), fractal measures, Law's texture measures, gradient structure tensors (GST), and Gabor filters [LAW 80, TOU 00, SHA 97]. A rigorous review of various segmentation techniques, specifically for cell segmentation can be found in [ABL 01].

In this work, we discuss the segmentation of cells in spectral images. We leverage a pattern theoretic approach of feature extraction for each pixel followed by clustering to achieve separation of cell pixels from those belonging to background. Finally, the pixels clustered as belonging to cells are refined using morphological operations.

13.3.1. *Imaging*

The core element of any spectral imaging system is the spectral dispersion component that separates light into its spectral components, and is coupled to a two-dimensional (2D) optical detector such as a CCD camera, or to an array of photo-multiplier tubes (PMT). In our system, we use a quarter-meter class, Czerny-Turner type monochromator that provides a tunable light emission spectrum at 10nm resolution. We utilize a wavelength range from 400-700nm for this study. The monochromator is connected to an Olympus BX51 upright optical microscope such that the light output from the monochromator feeds into the transmitted light path of the microscope. This allows for the use of conventional optical microscopy to acquire bright-field images at desired wavelengths (transmitted light). An Olympus UPlanApo 40X NA 0.9 was used for imaging. The Photometrics SenSys™CCD camera having 768 x 512 pixels (9x9μm) at 8-bit digitization is used which provides for high resolution low light image acquisition. To image each sample, the illumination from the monochromator was adjusted by achieving Köhler illumination for uniform excitation of the specimen. The condenser, aperture diaphragm, and the field stop were kept constant during measurements. Focusing was performed at the central wavelength of 550nm to minimize the chromatic aberration at all wavelengths. The system was calibrated as per the method proposed in [SHA 06]. Using a stepper controlled microscope stage, multiple images were acquired to cover the extent of the smear on the microscope slide. Resulting images were stitched to generate a composite mosaic.

A multispectral image allows the possibility to locate, discriminate, measure, and enumerate many entities within a specimen by detecting subtle differences among their individual spectral signatures[FAR 96]. Clearly, different stained cells will be spectrally distinct. However, spectral information in any cell can come from such optical processes as reflection and scattering. As long as the phenomenology is based on reproducible physical reality, classification of spectrally distinct species can be of great utility. Figure 13.2 (right) shows a subset of the spectral image (400nm, 500nm, 600nm, and 700nm) of a Papanicolaou stained cytological specimen. As seen the light absorption across various cellular constituents vary as a function of wavelength. This forms the spectral signature for each cellular entity.

To understand the spectral characteristics of biological samples, gray level image intensities may be used to determine the proportion of light transmitted by each cell across the exciting spectra. The transmission factor, T, is defined as :

$$T = I_t/I_i \tag{13.1}$$

where I_t is the intensity of the transmitted radiation at a point and I_i is the intensity of the incident light. Using the calibrated system, the incident light is fixed and known *a priori*. As such, one can then compute the absorption parameter for each pixel using the Beer-Lambert law[ORN 99] as :

$$A = \log(1/T) \tag{13.2}$$

For each pixel in the multispectral image, we measure the absorption parameter for all wavelengths to generate a feature vector representing the signature for that pixel.

13.3.2. *Cell Segmentation*

In this work, we use a bank of Gabor filters to extract a measure of texture at each pixel followed by clustering to group pixels belonging to the same class. Specifically, a two-dimensional Gabor function and its Fourier transform can be written as :

$$
\begin{aligned}
g(x,y) &= (\frac{1}{2\pi\sigma_x\sigma_y})\exp[-\frac{1}{2}(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}) + 2\pi jWx], \\
G(u,v) &= \exp-\frac{1}{2}[\frac{(u-W)^2}{\sigma_u^2} + \frac{v^2}{\sigma_v^2}],
\end{aligned}
\tag{13.3}
$$

where W is the frequency, σ_x and σ_y define the spread of the Gaussian in the x and y directions, respectively, and $\sigma_u = 1/2\pi\sigma_x$ and $\sigma_v = 1/2\pi\sigma_y$. These filters at various orientations and scales act as edge and line detectors and thus provide micro-features in a given region. A bank of such filters can be obtained by rotating and scaling of
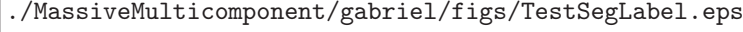
the base filter. By ensuring that the half-peak magnitude of the filter response in the frequency spectrum touch each other, we can obtain a set of self-similar functions known as Gabor Wavelets. Each new filter can be obtained as :

$$
\begin{aligned}
g_{mn}(x,y) &= a^{-m}g(x',y'), & (13.4) \\
x' &= a^{-m}(x\cos\theta + y\sin\theta), \\
y' &= a^{-m}(-x\sin\theta + y\cos\theta),
\end{aligned}
$$

where $a > 1$, m, n, is an integer such that $\theta = n\pi/K$ where K is the number of orientations and m is the number of scale out of S. We generate a bank of self-similar filters through appropriate dilations and translations of the basis Gabor function, $g(x,y)$, as defined by Manjunath et al. [MAN 96]. We use 3 scales and 4 orientations resulting in a total of 12 filters in the bank. To efficiently compute a measure of texture for each pixel in the multispectral image, we generate an average image from the multispectral stack. For each pixel in the average image, a feature vector is constructed comprising of the magnitude of each of the 12 filter responses, and the mean and standard deviation of the 12 magnitude values. In addition, the absorption is measured according to equation 13.2 for each pixel in each channel of the multispectral image. Hence, we generate a 45-dimensional feature vector for each pixel belonging to a spectral image with 31 channels. The extracted features are clustered using the standard k-means algorithm [HAR 79] which results in effective grouping of pixels belonging to the thyroid cells and partitioning of the image. Finally, morphological filtering [WU 08] is performed on the resultant labeled set of pixels to ensure homogeneity of labels and obtain smooth cell regions. Figure 13.4 (left) shows an example of a thyroid cell smear image at 520nm (1 of 31 channels) and the result of clustering (right) with the cluster of pixels detected as thyroid cells overlayed in black. Overall, the process flow for the algorithm is as shown in figure 13.5.

13.3.3. *Parallel Implementation*

In order to enable the processing of multi-gigabyte images, we detail the modifications to the code described above for a multi-process parallelization approach using MPI. The main source for parallelism in this applications comes from the size of the image. Thus, the code relies on data decomposition using a 1-D distribution of the image across the process. In a scenario utilizing np processes, the image is split into np parts, each part consisting of a fixed number of rows of the image. Process 0 will therefore own and operate on the first part of the image, process 1 the second one, etc.. Figure 13.6 highlights this approach for a two processes scenario. From the algorithmic point of view, the most simple approach would treat each portion of the image as an entire image and execute the sequential code accordingly. The main advantage

Figure 13.4. *One channel of a multispectral cell smear image (left) and result of clustering overlayed to highlight the cluster representing the cells of interest (right).*

Input :
    $I(\lambda) = \{I_{\lambda 1}, ..., I_{\lambda k}\}$: Spectral Image with k spectral channels
    n : number of expected image regions (cell and non-cell)
Output :
    $L = \{L_1, ..., L_r\}$: labeled image regions associated with cells of interest
Algorithm :
    <u>Initialization :</u> Generate Gabor filter bank with m scales and n orientations
    <u>Filtering :</u> Perform padding of image and generated Gabor filter bank
    Perform forward FFT of the image $\rightarrow I_{FFT}$
    for each (filter in the filter bank)
        perform forward FFT of each filter $\rightarrow G_{mnFFT}$
        perform complex multiplication of I_{FFT} and G_{mnFFT}
        perform inverse FFT of the resultant
    end for
    Compute average and standard deviation of magnitude response
    Compute spectral absorption coefficient for each image channel λ_i
    <u>Clustering :</u> Initialize cluster centers and perform k-means clustering
    Discard all pixels belonging to non-cell cluster
    <u>Smoothing :</u> Perform morphological smoothing for all pixels in cell cluster
    Remap pixels from cell cluster to visualize segmented image

Figure 13.5. *Sequential algorithm flow for cell segmentation in spectral image.*

of this approach is that processes are entirely decoupled, i.e. there is no communication required between the processes during execution. Furthermore, the modifications on the source to accomodate for this scenario are minimal. However, this approach has the disadvantage that the information gained from one portion of the image is not

propagated to other parts. This can lead to non-contiguous results and significant differences between the result of sequential and parallel code. Therefore, we follow for the rest of this section an approach, where the result of the parallel execution is as close to the sequential as possible. We detail the steps required to parallelize the convolution operation, the k-means clustering, the morphological smoothing, and finally for the occurring file I/O operations.
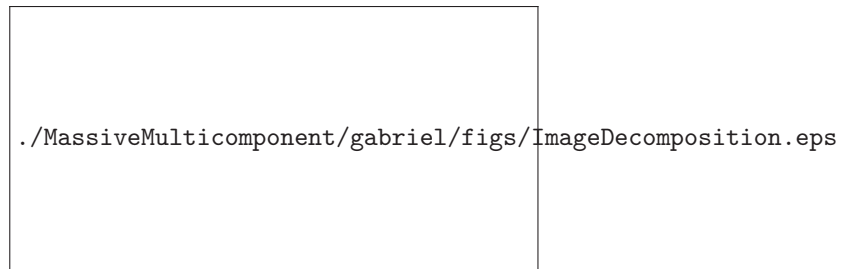
Figure 13.6. *1-D decomposition of an image on two processes.*

13.3.3.1. *Convolution*

As outlined in the previous section, the convolution operation is applied on a set of Gabor filters. From the computational perspective, the operation requires a Fast Fourier Transform (FFT) of the image, an FFT operation of each filter, a complex multiplication of the Fourier Transforms of image and filter, and finally an inverse FFT on the result of the complex multiplication. Since the image is distributed across multiple processes, the FFT operation requires an all-to-all communication operation internally. In the most straight forward implementation, the all-to-all operation requires np^2 messages with an overall data volume of n.

Several numerical libraries provide parallel FFTs using MPI. In the following, we give details on how to utilize one of the most widespread FFT libraries in scientific computing, namely FFTW [FRI 05]. FFTW version 2.1.5 supports MPI parallel applications. To compute an FFT using this library, the user has to create first a *plan* for the according FFT operation. The goal of creating a plan is to give the library the opportunity to evaluate the performance of multiple algorithms internally and choose the best performing one for the given problem size and number of processes. All subsequent invocations of the actual FFT operation will use the plan determined to deliver the best performance.

For our application, FFTW requires the image to be padded on the right and the lower side by the same number of rows/columns as the dimensions of the filter. This increases the overall image size, and has to be taken into account when determining the 1-D distribution of the image. The padded regions are shown in Figure 13.6 as a

gray border. Furthermore, in order for the convolution operation to make mathematical sense, the filters have to be padded to the same dimension as the padded image. Since the size of padded image is very large for the scenarios of interest in the chapter, one has to avoid to keep the filters in the padded format in order to minimize the memory consumption of the application. Note, that the padding borders are flipped compared to the original image after the inverse FFT operation, and appear on the upper and left part of the image. This requires a redistribution of the filter response for the subsequent operations such as clustering, since the first and the last processes have now a different number of rows of the image compared to the original distribution.

13.3.3.2. *k-means Clustering*

The input to the k-means clustering algorithm is a set of instances to be clustered (e.g. pixels) and the number of desired clusters. The result of the clustering step is a set of cluster centroids and a mapping of instances to the cluster. The sequential implementation starts by determining initial values for the cluster centroids. In the main compute loop, each instance/pixels determines its distance to each of the current centroids, and chooses the cluster with the minimum distance. For each cluster the code maintains the number of instances assigned to it as well as a weight. After all instances have made the according calculation, the cluster centroids can be adjusted based on the weight of the clusters and its members. The procedure is repeated until the difference between the mappings of two subsequent iterations is below a given threshold, i.e. the mapping of instances to clusters does not change anymore.

Figure 13.7 displays the individual steps of this algorithm along with the modifications required for the parallel version. These are shown as underlined statements. The overall goal of the parallel version is for each process to apply the clustering algorithm on its part of the image. Thus, all cluster related information such as the cluster centroids have to be replicated on all processes and kept synchronized. The first modification compared to the sequential version is, that a single process has to determine the initial values for the centroids. This process broadcasts using `MPI_Bcast` these values to all processes working on the same image.

The second modification required to the code is to introduce an additional global summation in order to determine for each cluster the overall number of pixels assigned to it and the overall weight of each cluster. This can be accomplished in MPI using the `MPI_Allreduce` function with the operator `MPI_SUM`. As a result of this function, all processes will have the overall number of pixels per cluster and the overall weight of each cluster. Each process can now compute the new cluster centroids independently. Due to the fact, that the global summations are only executed once per iteration, the algorithm has a favorable compute to communication ratio for large images.

13.3.3.3. *Morphological Smoothing*

The main idea behind this step is to remove isolated points in a cluster. As an example, if a pixels is assigned to a cluster a, but all neighboring pixels are assigned

Input :

    $I = \{i_1, ..., i_k\}$: Instances to be clustered

    n : number of clusters

Output :

    $C = \{c_1, ..., c_n\}$: cluster centroids

    $m : I \rightarrow C$: cluster membership

Algorithm :

    One process : determine initial values for C

    Broadcast initial values to all processes

    while (m has changed)

        $c_k^{temp} = 0, k = 1, ..., n$

        $count_k = 0, k = 1, ..., n$

        for each $i \in I$

            $m(i) = \min distance(i, c_k), k = 1, ..., n$

            $c_k^{temp} + = i$

            $count_k + +$

        end for

        Determine global sum of c^{temp} and $count$ across all processes

        Recompute C based on c^{temp} and $count$

    end while

Figure 13.7. *Parallel k-means clustering*

to cluster b, it makes under many circumstances sense to assign the particular pixel also to the cluster b. In order to implement this algorithm, the code has to compare the cluster assignments for all pixels with its neighboring pixels. In the parallel code, where each process holds a contiguous portion of the image, this algorithm can be applied without any major modification to all interior pixels. However, the pixels at the boundary of the domain need access to the clustering information for the previous respectively next rows of pixels, which are located on other processes.

This problem is solved by using a common technique in parallel computing, namely introducing the notion of ghost-cells. A ghost-cell is a copy of an item which is owned by another process. A process is allowed to read a ghost-cell, but is not allowed to modify it, since he is not the owner of the according element. On the other hand, the owner has the responsibility to update ghost-cells located on other processes every time the content of the original cell changes. Figure 13.8 shows an example for a three process configuration. In this example, each process has access to one additional row of pixels above and below their portion of the image. The ghost-cell distribution algorithm has to take into account, that the process owning on the first portion of the image does not have an upper row of ghost-cell pixels, and the last process does not have a lower row of ghost-cell pixels.

./MassiveMulticomponent/gabriel/figs/GhostCells.eps
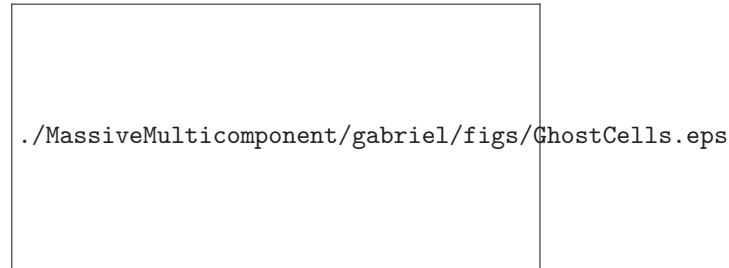
Figure 13.8. *Ghostcells in a 1-D decomposition of an image on three processes.*

In our code, we use two pixels in each direction for the morphological smoothing. Therefore, we create two rows of ghost-cells into each direction. The parallel algorithm can then perform the smoothing for all pixels assigned to his domain without introducing any additional communication overhead.

13.3.3.4. *File I/O*

There are two significant I/O operations occurring in the application. At the beginning of the analysis, all processes read portions of same image from the same file, although they access the file at different locations. Similarly, different processes write different portions of the same output file at the end of the analysis. For the subsequent discussion, we assume that the image is available in an uncompressed format, using either of the storage formats discussed in section 13.1. Following the initial calculations of the domain boundaries for each process, the user can easily modify his read/write routines to work in the parallel environment.

Consider as an example an image of $n_1 \times n_2$ pixels and n_{ch} channels. Assume furthermore for the sake of simplicity, that each pixel takes four bytes per channel, and the image is stored in a Band Sequential Format, i.e. all $n_1 \times n_2$ pixels of channel 1 are stored first in the file, followed by the entire data of channel 2 etc.. If the image analysis code is executed using n_p processes, each process holds a portion of the image of size $n_1 \times (n_2/n_p)$. For a very large image, it is unrealistic to assume, that each process can read the entire data into its main memory and extract afterwards the portions of the image which it owns. In the following we show the pseudo-code to read the according portions of the image on the ith process, which relies on the `fseek` function to move the file pointer in the file.

```
for ( j=0; j<n_ch; j++ ) {
    // Determine where to read the next portion of the file
  // for process i
```

```
filepos = j * n_1 * n_2 + i * n_1 * n_2/n_p;

  // move the file pointer to that position
fseek (fd, filepos, SEEK_SET);

  //Determine where to store the data in memory
mempos = j * n_1 * n_2/n_p;

//Execute the actual read operation
fread ( &(img[mempos]), n_1 \times (n_2/n_p)), 4, fd );
}
```

The code sequence above can be executed by all process simultaneously, and can easily be adapted also for write operations. It exposes however performance problems for two reasons. First and foremost, the number of overall I/O operations, i.e. the number of function calls to `fread`, is $n_p * n_{ch}$. Assuming n_p is large, this can cause significant problems, since I/O requests are very expensive. Note, that the I/O latency of a hard drive, i.e. the time the hard drive requires to move its head to the required position on the disk, start reading the data and transfer it to the main CPU, is in the range of 7-10ms, which is multiple orders of magnitudes higher than any other operation performed by the CPU. Reducing the number of I/O requests will therefore improve the performance of read/write operations, even if the overall amount of data transfered to and from the hard drive remains constant.

Second, since all processes execute the code sequence simultaneously, the I/O requests will arrive at the hard drive in an arbitrary order. Assuming that the image is too large to fit into the cache of the hard drive or operating system, which will typically be the case for the multi-gigabyte images discussed in this chapter, the hard drive will have to reposition the head constantly, instead of reading larger portions of the image in a single operation. Coordinating the `fread` requests of multiple processes could lead to reading the image for an entire channel in a single operation, and forward the according portions of the file to each process.

MPI offers collective read and write operations through the functions `MPI_File_ read_all` and `MPI_File_ write_all`. These operations require all processes to participate in the function call, although each process might read or write different portions of the file. The collective nature of these functions allows the MPI library to coordinate the access to the file among multiple process, reduce the number of I/O requests presented to the file system and the storage and thus improve the performance of read and write operations in parallel applications. For more details to MPI I/O and the according concepts, please refer to [MAY 01].

13.4. Performance Evaluation

In the following, we present the performance of the code described in the previous section. The measurements have been executed on the *shark* cluster at the University of Houston. The cluster consists of 24 single processor, dual core AMD Opteron nodes and 5 dual processor quad core AMD Opteron nodes, providing a total of 88 compute cores in the cluster. Each processor is running at 2.2 GHz, and has 1 GB of main memory per core. The nodes are connected by a 4xInfiniBand and a Gigabit Ethernet network. Each compute node has access to an NFS mounted home file system as well as to a parallel PVFS2 file system.

For our tests, we used gcc v4.2 and Open MPI v1.2.5 [GAB 04]. Tests have been executed for three image sizes, namely a

– an image consisting of 512×512 pixels with 31 spectral channels, with a total images size of 7.75 MB,

– an image consisting of 2048×2048 pixels with 21 spectral channels, with a total image size of 84 MB, and

– an image consisting of 8192×8192 pixels with 21 spectral channels and a total size of 1344 MB.

Each test presented here has been repeated three times, and the minimum time over the three runs has been used.

./MassiveMulticomponent/gabriel/figs/overall-time-8cores.eps

Figure 13.9. *Overall scaling of the application on a single 8-core node.*

./MassiveMulticomponent/gabriel/figs/overall-time.eps

Figure 13.10. *Overall scaling of the application on the cluster.*

Figures 13.9 and 13.10 shows the execution time for each image on various number of processes for the MPI parallel version. The first figure shows the results achieved on a single 8-core node of the cluster, while the results shown in the second figure utilize multiple dual-core nodes. The number of processes used was restricted by the main memory required for a given image and a minimum number of rows per process. As an example, the results obtained on a single compute node do not contain the largest image size, since the code could not be executed within the 8GB of main memory available on the node.

The code shows very good scaling for both scenarios, i.e. increasing the number of processes used decreases the execution time of the application. For the single-node scenario, the medium image scales up to eight cores. Analyzing this image takes for the sequential version more than 250 seconds, while using all eight available cores within the PC allows to decrease the execution for the same image to just 68 seconds. The smaller image scales similarly well when utilizing two and four cores on the node. However, further increasing the number of cores for the small image does not further decrease the execution time, since the communication costs exceed in this case the computational benefit of further splitting up the data.

The results obtained utilizing multiple nodes follow generally the same pattern. In most cases, doubling the number of processors used on the cluster decreases the execution time of the code by a factor of 1.6 - 1.9. The main exception from that behavior is observed for the largest image when utilizing 32 and 44 processes. The reason for this deviation is discussed in later subsections.

./MassiveMulticomponent/gabriel/figs/details.eps

Figure 13.11. *Scaling of the convolution, k-means clustering and morphological smoothing.*

An in-depth analysis of the application reveals, that the most time consuming parts are the k-means clustering and the convolution operation for both test scenarios. Figure 13.11 shows the scaling behavior of all three components discussed previously, namely the convolution, k-means clustering and smoothing for the multi-node experiments. All three algorithms scale very well for all problem sizes. Most notably, the deviation from the very good scaling behavior of the MPI code for the 32 and 44 processor test cases mentioned above is clearly not due to the performance of these routines. Note, that we skipped the performance of the morphological smoothing for the 512×512 image, since its execution time was less than one second, and it would have distorted the graph significantly.

13.4.1. *Influence of the Network Interconnect*

The results presented so far use the InfiniBand network of the cluster. In this section, we evaluate the influence of the network interconnect on the two most time consuming routines in the code, namely the convolution and the k-means clustering. Figure 13.12 compares the execution time of these two routines for the 2048×2048 image over 4xInfiniBand (latency : $3.5\mu s$, bandwidth : $1GB/s$) and Gigabit Ethernet (latency : $55\mu s$, bandwidth : $80MB/s$). The k-means clustering is nearly insensitive to the difference in the quality of the networks due to the low communication overhead introduced by the global reductions. However, the convolution, which consists of a large number of FFTs, does not scale beyond four processes on the Gigabit Ethernet network, while it still shows a performance improvement for 16 processes for the

very same test case when using InfiniBand. This can be explained with the communication pattern of a parallel FFT, which involves a large number of (small) messages, and therefore shows a sensitivity to the latency of a network.

Figure 13.12. *Performance of the convolution and k-means clustering of InfiniBand and Gigabit Ethernet.*

13.4.2. *Influence of the File System*

Beyond the convolution and the k-means clustering, the next most time consuming code sections are the routines dealing with reading input files and writing output files. The MPI version of the code has the option to write the texture data into output files. This is mostly to facilitate future processing steps in realizing a complete CAD solution.

Tableau 13.1. *Comparing MPI I/O with POSIX I/O over PVFS2 and NFS.*

Image	No. of processes	MPI I/O on PVFS2	POSIX I/O on PVFS2	POSIX I/O on NFS
8192×8192	16	21.9	49.0	115.9
8192×8192	32	57.0	51.0	114.4
8192×8192	44	66.7	50.4	115.9

As described in the previous section, I/O operations are performed using collective MPI I/O routines. For the 8192×8192 image, the size of the texture information stored

in different files sums up to 6GB. Table 13.1 shows, that using PVFS2 over the NFS file system on the cluster reduce the overall execution time of I/O operations by more than 50%. In some instances, the MPI I/O version of the code outperforms the version of the MPI parallel code, which uses regular POSIX/C style read and write operations. However, the time spent in I/O operations increases dramatically for the 32 and 44 processes test cases when using MPI I/O. The reason is probably, that other test case has only a single MPI process running per node, while the 32 and 44 processes test cases run two MPI processes on each node, one per CPU core. The implementation of the collective I/O routines within Open MPI seem to generate for these instances a non-negligble overhead, and the POSIX I/O based parallel version outperforms in this case the MPI I/O version. Thus, the conclusion of this subsection is, that utilizing a parallel file system such as PVFS2 is mandatory for large images. Using MPI I/O over regular sequential I/O operation might further increase the performance for certain scenarios, and lead to a minor performance degradation in others. Improving the performance of collective I/O operations is however an active research topic, and we would expect to see the scale shift significantly in favor of MPI I/O operations over the next years.

13.5. Conclusions

Application requirements and architectural developments force programmers to rethink and reformulate their applications. In order to fully exploit the potential of modern computer systems, the application developer has to explicitly expose the parallelism available within his code. We demonstrated successfully, how the individual sections of an image analysis application can be parallelized for multi-spectral images such as occurring in spectral microscopy. The performance benefits of this approach on both, a single PC/node utilizing eight compute cores and a cluster of PCs have been documented. We also highlighted the necessity to utilize state-of-the-art components for the cluster, specifically for the network interconnect and the file system, in order for the application to scale to a large number of processes.

Although encouraging, the current results are just the starting point for further work. We plan to parallelize further components typical for image analysis applications, such as feature extraction and classification. Furthermore, we explore also the potential for parallelism for other segmentation algorithms, such as graph based approaches [FEL 04]. A final piece of the puzzle to fully utilize multiple compute processes efficiently lies in the development of parallel read and write routines for compressed images. Since utilizing compression techniques to store large scale images is the norm as of today, this component is critical for the successful deployment of the parallel code version into every-day production environments.

[ABL 01] Ablameyko S., Nedzved A., Lagunovsky D., Patsko O., Kirillov V., « Cell image segmentation : review of approaches », *Proc. Int. Conf. Pattern Recognition*, vol. 2, p. 26–34, 2001.

[ADI 01] ADIGA P. U., CHAUDHURI B., « An efficient method based on watershed and rule-based merging for segmentation of 3-D histo-pathological images », *Pattern Recognition*, vol. 34, p. 1449-1459, 2001.

[ASS 03] ASSOCIATION T. A. T., « *Consensus guidelines for thyroid testing in the new millennium* », vol. 6(13), Chapitre Thyroid fine needle aspiration (FNA) and cytology, p. 1–80, 2003.

[BAM 98] BAMFORD P., LOVELL B., « Unsupervised cell nucleus segmentation with active contours », *Signal Processing*, vol. 71, n°2, p. 203–213, décembre 1998.

[BAR 08] BARKER K., DAVIS K., HOISIE A., KERBYSON D., LANG M., PAKIN S., SAN-CHO J., « Experiences in Scaling Scientific Applications on Current-generation Quad-Core Processors », *IEEE International Parallel and Distributed Processing Symposiom (IPDPS) 2008*, April 14-18, p. 1 – 8, 2008.

[BAY 02] BAYANI J., SQUIRE J. A., « Spectral karyotyping », *Methods Mol Biol*, vol. 204, p. 85-104, 2002, Journal Article United States.

[BEA 03] BEARMAN R. L. G., « Biological Imaging Spectroscopy », VO-DINH T., Ed., *Biomedical Photonics Handbook*, CRC Press, Boca Raton, 2003.

[CAR 00] CARNS P. H., III R. B. W. B. L., THAKUR R., « PVFS : A Parallel File System for Linux Clusters », *Proceedings of the 4th Annual Linux Showcase and Conference*, 2000.

[CAR 03] CARDULLO R. A., « Fundamentals of image processing in light microscopy », *Methods Cell Biol*, vol. 72, p. 217-42, 2003, Journal Article Review United States.

[CAS 79] CASTLEMAN K. R., *Digital Image Processing*, Prentice Hall, 1979.

[CHA 08] CHAPMAN B., JOST G., VAN DER PAS R., *Using OpenMP : Portable Shared Memory Parallel Programming*, The MIT Press, Cambridge, Massachusetts and London, England, 2008.

[CHE 06] CHENG X., ZHOU X., WONG T., « Automated segmentation, classification and tracking of cancer cell nuclei in time-lapse microscopy », *IEEE Transactions on Biomedical Engineering*, vol. 53, p. 762-766, 2006.

[Cle 08] CLEARSPEED HOMEPAGE, « http ://www.clearspeed.com », 2008.

[FAR 96] FARKAS D., BALLOU B., FISHER F., FISHMAN D., « Microscopic and mesoscopic spectral bio-imaging », *Proceedings of SPIE*, vol. 2678, p. 200-209, 1996.

[FAU 80] FAUGERAS O. D., PRATT W. K., « Decorrelation Methods of Texture Feature Extraction », *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 2, n°4, p. 323–332, juillet 1980.

[FEL 04] FELZENSZWALB P. F., HUTTENLOCHER D. P., « Efficient Graph-Based Image Segmentation », *Int. J. Comput. Vision*, vol. 59, n°2, p. 167–181, Kluwer Academic Publishers, 2004.

[FEN 07] FENG C., SHUZHEN C., LIBO Z., « New abnormal cervical cell detection method of multi-spectral pap smears », *Wuhan University Journal of Natural Sciences*, vol. 12, n°3, p. 476-480, 2007.

[FER 98] FERRER-ROCA O., GOMEZ J. A. P., ESTEVEZ M., « Chromatin texture from hematoxylin stained thyroid lesions », *Anal Cell Pathol.*, vol. 17, n°4, p. 209–217, 1998.

[FRI 05] FRIGO M., JOHNSON S. G., « The Design and Implementation of FFTW3 », *Proceedings of IEEE*, vol. 93, n°2, p. 216–231, 2005.

[GAB 04] GABRIEL E., FAGG G. E., BOSILCA G., ANGSKUN T., DONGARRA J. J., SQUYRES J. M., SAHAY V., KAMBADUR P., BARRETT B., LUMSDAINE A., CASTAIN R. H., DANIEL D. J., GRAHAM R. L., WOODALL T. S., « Open MPI : Goals, Concept, and Design of a Next Generation MPI Implementation », *Proceedings, 11th European PVM/MPI Users' Group Meeting*, Budapest, Hungary, p. 97–104, September 2004.

[GAT 00] GAT N., « Imaging spectroscopy using tunable flters : A review. », *Proc. of SPIE*, vol. 4056, p. 50-64, 2000.

[GHA 93] GHARIB H., GOELLNER J. R., « Fine-needle aspiration of the thyroid : an appraisal », *Ann. Intern. Med.*, vol. 118, p. 282–289, 1993.

[GRA 03] GRAMA A., KARYPIS G., KUMAR V., GUPTA A., *Introduction to Parallel Computing (2nd Edition)*, Pearson Education Limited, 2003.

[HAR 79] HARTIGAN J. A., WONG M. A., « A K-means clustering algorithm », *Applied Statistics*, vol. 28, p. 100–108, 1979.

[HAR 02] HARAGUCHI T., ET. AL., « Spectral imaging fluorescence spectroscopy », *Genes to Cells*, vol. 7(9), p. 881-887, 2002.

[HEG 04] HEGEDUS L., « The thyroid nodule », *New Eng J Med.*, vol. 351, p. 1764–1771, 2004.

[I.T 04] I.T. ASSOCIATION,   « InfiniBand architecture specification. Release 1.2 », http ://www.infinibandta.org/specs, 2004.

[JOH 07] JOHNS C. R., BROKENSHIRE D. A., « Introduction to the Cell Broadband Engine Architecture », *IBM Journal of Research and Development*, vol. 51, n°5, p. 503–520, 2007.

[KER 02] KEREKES J., BAUM J., « Spectral imaging system analytical model for subpixel object detection », *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40(5), p. 1088-1101, 2002.

[KIM 03] KIM E., CHOI Y., YANG H. S., KANG M., JEONG S., YANG S., KIM E. D., LEE J. H., RASHEED A. A., NASIR H. M., ROSDI M. R., « A high-resolution multi-spectral imaging system for small satellites », *Acta Astronautica*, vol. 52(9), p. 813-818, 2003.

[KUL 03] KULCKE A., GURSCHLER C., SPOCK G., LEITNER R., KRAFT M., « On-line classification of synthetic polymers using near infrared spectral imaging », *Journal of near infrared spectroscopy*, vol. 11(1), p. 71-81, 2003.

[KUR 98] KURUGOLLU F., SANKUR B., « Color Cell Image Segmentation Using Pyramidal Constraint Satisfaction Neural Network », *IAPR Workshop on Machine Vision Applications*, p. 85–88, 1998.

[LAW 80] LAWS K. I., Textured Image Segmentation, PhD thesis, 1980.

[LEV 00] LEVENSON R.M ; HOYT C., « Spectral Imaging and Microscopy », *American Laboratory*, vol. 32, n°22, p. 26-33, 2000.

[LEV 06] Levenson R. M., Mansfield J. R., « Multispectral imaging in biology and medicine : slices of life », *Cytometry A*, vol. 69, n°8, p. 748-58, 2006, R01 ca108468/ca/nci R44 ca88684/ca/nci Journal Article Research Support, N.I.H., Extramural Review United States the journal of the International Society for Analytical Cytology.

[LIE 95] Lie W., « Automatic target segmentation by locally adaptive image thresholding », *IEEE Trans. Image Process.*, vol. 4, p. 1036–1041, July 1995.

[Lus] Lustre Home Page, « http ://wiki.lustre.org ».

[MAI 03] Maiti D., Sennoune S., Martinez-Zaguilan R., « Proton gradients in human breast cancer cells determined by confocal, multiphoton, and spectral imaging microscopy », *FASEB Journal*, vol. 17(4), pageA467, 2003.

[MAL 96] Malik Z., Cabib D., Buckwald R., Talmi A., Garini Y., Lipson S., « Fourier transform multipixel spectroscopy for quantitative cytology », *Journal of microscopy*, vol. 182, p. 133-140, 1996.

[MAN 96] Manjunath B. S., Ma W., « Texture features for browsing and retrieval of image data », *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI - Special issue on Digital Libraries)*, vol. 18, n°8, p. 837–42, Aug 1996.

[MAR 02] Marr D. T., Binns F., Hill D. L., Hinton G., Koufaty D. A., Miller J. A., Upton M., « Hyper-Threading Technology Architecture and Microarchitecture », *Intel Technology Journal*, vol. 6, n°1, p. 4–15, 2002.

[MAY 01] May J., *Parallel I/O for High Performance Computing*, Morgan Kaufmann, 2001.

[MEA 02] Meas-Yedid V., Tilie S., Olivo-Marin J.-C., « Color Image Segmentation Based on Markov Random Field Clustering for Histological Image Analysis », *ICPR '02 : Proceedings of the 16 th International Conference on Pattern Recognition (ICPR'02) Volume 1*, Washington, DC, USA, IEEE Computer Society, page10796, 2002.

[Mes 97] Message Passing Interface Forum, MPI-2 : Extensions to the Message Passing Interface, July 1997, http ://www.mpi-forum.org/.

[MIL 91] Miller P. J., « Use of tunable liquid crystal filters to link radiometric and photometric standards. », *Metrologia*, vol. 28, p. 145-149, 1991.

[MOO 65] Moore G. E., « Cramming more components onto integrated circuits », *Electronics*, vol. 38, n°8, April 1965.

[Myr 08] Myricom homepage, « http ://www.myricom.com », 2008.

[ORN 99] Ornberg R. L., Woerner M., Edwards D. A., « Analysis of stained objects in histopathological sections by spectral imaging and differential absorption », *J. Histochem. Cytochem.*, vol. 47, n°10, p. 1307–1313, 1999.

[PRO 01] Prost J.-P., Treumann R., Hedges R., Jia B., Koniges A., « MPI-IO/GPFS, an Optimized Implementation of MPI-IO on top of GPFS », *Proceedings of Supercomputing*, November 2001.

[Qua 08] Quadrics homepage, « http ://www.quadrics.com », 2008.

[QUI 04] Quinn M. J., *Parallel Programming in C with MPI and OpenMP*, McGraw Hill Higher Education, 2004.

[ROY 95] ROY K., BOTTRILL I., INGRAMS D., PANKRATOV M., REBEIZ E., WOO P., SHAPSHAY S., KABANI S., MANOHARAN R., « Diagnostic fluorescence spectroscopy of oral mucosa », *Proceedings of SPIE*, vol. 2395, p. 135-142, 1995.

[SAH 88] SAHOO P. K., SOLTANI S., WONG A. K. C., « A Survey of Thresholding Techniques », *Computer Vision, Graphics and Image Processing*, vol. 41, p. 233–260, 1988.

[SAN 85] SANDBERG R., GOLDBERG D., KLEIMAN S., WALSH D., LYON B., « Design and Implementation or the Sun Network Filesystem », 1985,.

[SCH 05] SCHWARTZKOPF W. C., BOVIK A. C., EVANS B. L., « Maximum-likelihood techniques for joint segmentation-classification of multispectral chromosome images », *IEEE Trans Med Imaging*, vol. 24, n°12, p. 1593-610, 2005, Journal Article United States.

[SEI 08] SEILER L., CARMEAN D., SPRANGLE E., FORSYTH T., ABRASH M., DUBEY P., JUNKINS S., LAKE A., SUGERMAN J., CAVIN R., ESPASA R., GROCHOWSKI E., JUAN T., HANRAHAN P., « Larrabee : a many-core x86 architecture for visual computing », *ACM Trans. Graph.*, vol. 27, n°3, p. 1–15, 2008.

[SER 95] SER P., SIU W., « Novel detection of conics using 2-D Hough planes », *IEEE Proc. Vision, Image Signal Processing*, vol. 142, p. 262–270, 1995.

[SHA 97] SHAH S., AGGARWAL J. K., « A Bayesian Segmentation Framework for Textured Visual Images. », *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, p. 1014–1020, 1997.

[SHA 05] SHAH S., « Multispectral Integration for Segmentation of Chromosome Images », *Proceedings 11th International Conference on Computer Analysis of Images and Patterns*, p. 506–513, 2005.

[SHA 06] SHAH S., THIGPEN J., MERCHANT F., CASTLEMAN K., « Photometric Calibration for Automated Multispectral Imaging of Biological Samples », METAXAS D., WHITAKER R., RITTSCHER J., SEBASTIAN T., Eds., *Proceedings of 1st Workshop on Microscopic Image Analysis with Applications in Biology (in conjunction with MICCAI, Copenhagen)*, p. 27–33, 2006.

[SHA 08] SHAH S., SCHWARTZ M. R., MODY D. R., SCHEIBER-PACHT M., AMRIKACHI M., « The Role of Multispectral Microscopy in Differentiating Benign and Malignant Thyroid Nodules : A Pilot Study of 24 Cases », 2008.

[THO 08] THORMAN P., *Multigrid Methods on GPUs*, VDM Verlag Dr. Mueller e.K., 2008.

[TOP 08] TOP 500 WEBPAGE, « http ://www.top500.org/ », 2008.

[TOU 00] TOURASSI G. D., FREDERICK E. D., VITTITOE N. F., COLEMAN R. E., « Fractal Texture Analysis of Perfusion Lung Scans », *Comput Boimed Res*, vol. 33, n°3, p. 161–171, June 2000.

[VIN 91] VINCENT L., SOILLE P., « Watersheds in Digital Spaces : An Efficient Algorithm Based on Immersion Simulations », *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 13, n°6, p. 583–598, 1991.

[WAS 07] WASSON S., « Nvidia GeForce 8800 graphics processor », 2007.

[WEI 05] WEIER H.-U. G., WEIER J. F., RENOM M. O., ZHENG X., COLLS P., NUREDDIN A., PHAM C. D., CHU L. W., RACOWSKY C., MUNN S., « Fluorescence In Situ Hybridization and Spectral Imaging Analysis of Human Oocytes and First Polar Bodies », *Journal of Histochemistry and Cytochemistry*, vol. 53, n°3, p. 269-272, 2005.

[WU 98] WU H., BARBA J., GIL J., « A parametric fitting algorithm for segmentation of cell images », *IEEE Trans Biomed Eng*, vol. 45, p. 400–407, 1998.

[WU 00] WU H. S., BERBA J., GIL J., « Iterative thresholding for segmentation of cells from noisy images », *Journal of Microscopy*, vol. 179, n°3, p. 296–304, 2000.

[WU 08] WU Q., MERCHANT F. A., CASTLEMAN K. R., *Microscope Image Processing*, Academic Press, 2008.

[WYA 75] WYATT C. L., « Infrared Spectrometer. Liquid-helium-cooled rocketborne circular-variable filter. », *Applied Optics*, vol. 14, p. 3086-3091, 1975.

[YOG 96] YOGESAN K., JORGENSEN T., ALBREGTSEN F., TVETER K. J., DANIELSEN H. E., « Entropy-based texture analysis of chromatin structure in advanced prostate cancer », *Cytometry*, vol. 24, p. 268–276, 1996.

Chapitre 14

Hypercomplex models and processing for vector images

The problem of processing colour images has been of considerable interest since the 1990s, when advances in technology made processing of these images practical using readily-available desktop computers and workstations. Researchers that studied the subject soon realised that there were fundamental theoretical advances needed. For example, although it was possible to process colour images using separate luminance and chrominance, there was no method available to handle the whole image at once (so-called 'holistic' processing). Sangwine and Thornton in the UK, and McCabe *et al* in Australia, tried the idea of using complex numbers to represent the chrominance information [SAN 98b, MCC 00]. This led to the idea of using hypercomplex numbers to handle the pixels as vectors, which is the subject of this chapter. Independently, Pei, in Taiwan, studied the use of hypercomplex numbers in colour image processing [PEI 97, PEI 99] and there were others who used hypercomplex numbers for one-dimensional signal processing, mainly as a tool for designing high-order filters.

14.1. Hypercomplex numbers

Hypercomplex numbers are a generalisation of the idea of complex numbers to dimensions higher than 2. Like the complex numbers, they depend on the existence of square roots of -1. Unlike the complex numbers, which are unique, there are many possible definitions of hypercomplex numbers, and their algebras vary considerably in properties. However, among the many possibilities, one particular algebra stands out :

Chapitre rédigé par Stephen J. SANGWINE, Todd A. ELL, Nicolas LE BIHAN.

the *quaternions*, discovered by Hamilton in 1843 [HAM 53]. The quaternions stand out from other hypercomplex algebras because the algebra is a *division algebra* (every non-zero quaternion has a multiplicative inverse), and apart from non-commutative multiplication, the algebra has no difficult properties. In contrast, all other hypercomplex algebras suffer from additional limitations such as non-associativity, divisors of zero, lack of a norm etc. These problems are not insurmountable, but they are a considerable drawback in numerical processing and in theoretical development, since algebraic manipulations can be very difficult, and it is necessary to consider carefully what happens if a value in a numerical algorithm is a divisor of zero – can it cause a numerical error, for example ? The Clifford algebras have been proposed as an alternative to the quaternions (in fact one of the Clifford algebras is isomorphic to the quaternions), but no other algebra has been developed to the point where such a wide variety of numerical and algebraic tools has been implemented. The unique position of the quaternions among hypercomplex algebras was established in 1877 by the Theorem of Frobenius, which places the reals, complex numbers and quaternions together as the only finite-dimensional division algebras over the reals [KAN 89, Chapter 19][1].

The fundamental idea used in applying hypercomplex numbers to image processing is that each image pixel can be represented by a single hypercomplex number, with several components (three for colour images). Then, hypercomplex operations (arithmetic, algebraic) can be applied to processing of the image, or to the theoretical development of processing techniques that can then be implemented in some other way (for example, using matrix or standard vector arithmetic).

In the remainder of this chapter we concentrate on the quaternion algebra. Our work makes use of an open source Matlab® toolbox, developed in 2005 [SAN 05] which implements all of the fundamental quaternion operations needed to reproduce the results presented here. We also use a newer toolbox, which builds on the first to implement linear quaternion systems as described in § 14.4 [ELL 07d].

14.1.1. *Definitions*

Table 14.1 summarises the notation used in this chapter.

Quaternions can be represented in a variety of ways, but the simplest is the Cartesian form $q = w + x\boldsymbol{i} + y\boldsymbol{j} + z\boldsymbol{k}$, where $w, x, y, z \in \mathbb{R}$. $\boldsymbol{i}, \boldsymbol{j}$ and $\boldsymbol{k}$ are roots of -1 and obey the rules discovered by Hamilton in 1843 : $\boldsymbol{i}^2 = \boldsymbol{j}^2 = \boldsymbol{k}^2 = \boldsymbol{ijk} = -1$. From these rules it can be seen that $\boldsymbol{ij} = \boldsymbol{k}$ and $\boldsymbol{ji} = -\boldsymbol{k}$, for example, exhibiting the lack of commutative multiplication for which the quaternions are famous (as the

---

1. It is possible to define quaternions over other fields, especially the complex numbers (the complexified quaternions or biquaternions).

Notation	Meaning
$\mathbb{R}$	Reals
$\mathbb{H}$	Quaternion numbers
$\mathbb{C}_x$	Complex subfield of $\mathbb{H}$
$GL_4(\mathbb{R})$	General linear group of 4×4 invertible real matrices
S [ ]	Scalar operator, extracts the scalar part of a quaternion
U [ ]	Unit operator, yields a quaternion of unit modulus
V [ ]	Vector operator, extracts the vector part of a quaternion
UV [ ]	Unit-vector operator, equivalent to U [V [ ]]
$\mathbb{H}^{n \times m}$	Set of n-by-m quaternion matrices
V $[\mathbb{H}]^{N \times M}$	Set of n-by-m pure quaternion matrices

Tableau 14.1. *Notation used in this chapter.*

first example of an algebra that lacked commutativity), and which caused Hamilton so much difficulty in discovering the multiplication rules [WAE 76].

As with complex numbers, a quaternion has a norm : $\|q\| = w^2 + x^2 + y^2 + z^2$, and a modulus which is the square root of the norm : $|q| = \sqrt{\|q\|}$. The quaternions are a normed algebra, meaning that the 'rule of the norms' is obeyed and every non-zero quaternion has a multiplicative inverse. Thus, for any two quaternions q_1 and q_2, including $q = 0$, we have that : $\|q_1 q_2\| = \|q_1\| \, \|q_2\|$.

A quaternion with $w = 0$ is called a *pure* quaternion, and if it also has unit modulus, it is a square root of -1. In what follows we use the symbol $\boldsymbol{\mu}$ to denote such a unit pure quaternion (there are an infinite number of these, corresponding to the set of points on the surface of a unit radius sphere). The three quaternion 'operators' $\boldsymbol{i}, \boldsymbol{j}$ and $\boldsymbol{k}$ are special cases of unit pure quaternions.

It is often useful to regard a quaternion as composed of two parts, called the *scalar* and *vector* parts, defined thus : $S[q] = w$ and $V[q] = x\boldsymbol{i} + y\boldsymbol{j} + z\boldsymbol{k}$. In this form, the quaternion is analogous to a complex number, an analogy that can be made closer by separating the modulus of the vector part thus : $q = S[q] + \boldsymbol{\mu} \, |V[q]|$. In this form, $\boldsymbol{\mu}$ is a square root of -1, and the scalar part and the modulus of the vector part are analogous to the real and imaginary parts of a complex number.

A very significant alternative representation for a quaternion is the polar form, which is very similar to the polar form of a complex number : $q = |q| \exp \boldsymbol{\mu}\theta$. $\boldsymbol{\mu}$ is known as the *axis* of the quaternion, and it defines a direction in 3-space which is the imaginary axis of a plane containing the value of the quaternion. This plane is exactly analogous to the Argand plane of a complex number, except that here the plane is a plane section of 4-space. All quaternions with the same axis lie in the same plane, and

they commute. Such a subset of quaternions is isomorphic to the complex numbers, a fact that we use later. The argument of the quaternion, θ, normally has a value between 0 and π, that is the quaternion is limited to half of the plane in which it lies. The other half of the plane is occupied by quaternions with an oppositely directed axis. The axis of a quaternion can be determined easily from the Cartesian form by normalising the vector part : $\boldsymbol{\mu} = \mathrm{V}[q]/|\,\mathrm{V}[q]|$. The argument is obtained in the same way as for complex numbers[2] : $\theta = \tan^{-1}\frac{|V[q]|}{S[q]}$.

A biquaternion, or *complexified* quaternion, is a quaternion with complex components : $a + b\boldsymbol{i} + c\boldsymbol{j} + d\boldsymbol{k}$, $a, b, c, d \in \mathbb{C}$. The biquaternion algebra is *not* a normed division algebra (there exist divisors of zero, and hence the 'rule of the norms' does not always hold). This algebra has been used in vector image and signal processing, for example [MIR 05, § 4.4] and [SAI 08], but much further work remains to be done before it is as well understood as the quaternion algebra $\mathbb{H}$. In particular, the impact of the existence of divisors of zero has not yet been fully studied [ALF 07].

14.1.2. *Basic properties*

The multiplication of two quaternions in Cartesian form is completely determined by Hamilton's famous multiplication rules for the three quaternion 'operators' $\boldsymbol{i}$, $\boldsymbol{j}$ and $\boldsymbol{k}$. However, it is also useful to understand multiplication in terms of the scalar and vector parts. We state the result here without proof, since it is given in many sources, for example [WAR 97, §2.2] :

$$q_1 q_2 = \mathrm{S}[q_1]\,\mathrm{S}[q_2] - \mathrm{V}[q_1] \cdot \mathrm{V}[q_2] + \mathrm{S}[q_1]\,\mathrm{V}[q_2] + \mathrm{S}[q_2]\,\mathrm{V}[q_1] + \mathrm{V}[q_1] \times \mathrm{V}[q_2]$$

where the symbols $\cdot$ and $\times$ represent the scalar and vector (dot and cross) products of classical vector analysis. This formula shows clearly where the non-commutativity arises, since the vector or cross product is non-commutative. It also shows why multiplication of quaternions with co-planar vector parts (common axis) *is* commutative : the cross product in this case vanishes, because the two vector parts are parallel.

The conjugate of a quaternion is obtained by negating the vector part, or equivalently the three Cartesian components of the vector part, or by negating the exponent in the polar form. It is usually denoted by an overbar, $\bar{q}$. The norm, described in §14.1.1 can also be expressed as the product of a quaternion with its conjugate, just as for complex numbers : $\|q\| = q\,\bar{q}$.

The quaternions are justly famous for their utility in representing and calculating rotations in 3-space. For example, the operation of rotation about an axis in 3-space

---

2. In practice, numerically this is computed using an `atan2` function or similar. See for example, the method used in the Matlab® toolbox used by the authors [SAN 05].

is given by the mapping $x \rightarrow px\bar{p}$ where $p = e^{\boldsymbol{\mu}\phi/2}$ (a quaternion expressed in polar form). The axis of the rotation is $\boldsymbol{\mu}$ and the angle of rotation about this axis is ϕ [COX 46, Theorem 3.2]. Equally compact formulae exist for reflections about a plane, and rotations and reflections in 4-space. We omit the details here, but refer the reader to [COX 46] or [WAR 97]. It is a straightforward matter of algebraic manipulation to devise formulae for more elaborate geometrical operations which can be composed from elementary operations.

The theorem of de Moivre generalises to quaternions, and expresses the scalar and vector parts in terms of the polar form : $q = |q|e^{\boldsymbol{\mu}\theta} = |q|(\cos\theta + \boldsymbol{\mu}\sin\theta)$, where the scalar part is $S[q] = |q|\cos\theta$ and the vector part is $V[q] = |q|\boldsymbol{\mu}\sin\theta$. This generalisation works because all unit pure quaternions are roots of -1.

14.1.3. *Symplectic form*

Quaternions may be decomposed into pairs of complex numbers, and this is of considerable utility in the numerical implementation of fast quaternion Fourier transform algorithms, presented in § 14.3. The existence of such decompositions was known in Hamilton's time, in the form of the famous 'Cayley-Dickson' decomposition : $q = (w + x\boldsymbol{i}) + (y + z\boldsymbol{i})\boldsymbol{j}$ which multiplies out to give the standard quaternion Cartesian form. The idea may be generalised to an arbitrary basis of three mutually perpendicular unit pure quaternions $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2$ and $\boldsymbol{\mu}_3$ such that $\boldsymbol{\mu}_1\boldsymbol{\mu}_2 = \boldsymbol{\mu}_3$. Then we can express any quaternion in the form $q = w + x'\boldsymbol{\mu}_1 + y'\boldsymbol{\mu}_2 + z'\boldsymbol{\mu}_3$ by a change of basis (w remains invariant), and then we have the so-called 'symplectic'[3] decomposition [ELL 00a] and [ELL 07c, §IV.A] : $q = (w + x'\boldsymbol{\mu}_1) + (y' + z'\boldsymbol{\mu}_1)\boldsymbol{\mu}_2$ in which the two terms in parentheses are isomorphic to complex numbers. We could write this form of the quaternion more compactly as $q = C_1 + C_2\boldsymbol{\mu}_2$, where $C_1, C_2 \in \mathbb{C}_{\boldsymbol{\mu}_1}$, the set of quaternions with axis $\boldsymbol{\mu}_1$ which are isomorphic to elements of $\mathbb{C}$. It is straightforward to verify that a change of order of the second complex value with the 'outer' operator requires a complex conjugation of the complex value : $q = C_1 + C_2\boldsymbol{\mu}_2 = C_1 + \boldsymbol{\mu}_2 C_2^*$, a fact that we use later, in § 14.3.2, Theorem 14.3.1.

An alternative is to consider the decomposition in terms of a geometric decomposition of the vector part of the original quaternion into components parallel and perpendicular to the direction $\boldsymbol{\mu}_1$. This can be expressed in terms of quaternions by the formulae : $q_{\parallel} = \frac{1}{2}(q - \boldsymbol{\mu}_1 q \boldsymbol{\mu}_1)$ and $q_{\perp} = \frac{1}{2}(q + \boldsymbol{\mu}_1 q \boldsymbol{\mu}_1)$, which resolve the vector part of the quaternion into components parallel and perpendicular to $\boldsymbol{\mu}_1$ (the formulae are based on reflections about a plane normal to $\boldsymbol{\mu}_1$, which contains $\boldsymbol{\mu}_2$ and $\boldsymbol{\mu}_3$).

---

3. The term 'symplectic' was used because the two components of the decomposition can be called 'simplex' and 'perplex' components.

14.1.4. *Linear quaternion equations*

Processing of vector images requires a vector generalisation of the concept of *scaling* a pixel. Quaternions provide a powerful means to perform geometric operations on vectors, as we have seen above with rotations. What is not so obvious is that these geometric operations are in some cases *linear* and in fact can be made linear in all cases by using homogeneous coordinates.

Linear equations in quaternions are not trivial, because of the non-commutative multiplication. In a commutative algebra, a general linear equation is of the form $y = mx + c$, whereas in the quaternion algebra a general linear form would be $y = a_1 x b_1 + a_2 x b_2 + \ldots + a_N x b_N$. Until recently it was not known whether it was possible to reduce such an equation to a fixed number of terms (a *canonic* form), but in 2007 it was shown by one of the authors [ELL 07b] that there is indeed a canonic form : $y = Ax + Bx\boldsymbol{i} + Cx\boldsymbol{j} + Dx\boldsymbol{k}$ (or an alternative in which the fixed coefficients are on the right, and $\boldsymbol{i}$, $\boldsymbol{j}$ and $\boldsymbol{k}$ are on the left). Methods for solving for the four coefficients are now known both algebraically [ELL 07b] and numerically [SAN 08]. The significance of this result is that, no matter how complex a quaternion filter becomes when composed of more elementary quaternion filters, it is always possible to reduce the implementation of the filter to at most four quaternion convolutions. Furthermore, the fact that the canonic form has 16 degrees of freedom delimits the possible set of operations that can be implemented. There is an equivalence between the canonic form of the linear quaternion equation, and the operations that can be represented using 4×4 matrices (the general linear group of order 4, $GL_4(\mathbb{R})$). Thus an equivalent implementation would be a matrix equation operating on pixels as 4-vectors with 4×4 matrices as the coefficients. This is a useful equivalence, but the matrix form is not at all easy to use for devising and composing geometric operations, whereas the quaternion form offers simple ways to compose operations in terms of geometric concepts. Work is ongoing to study the set of operations in terms of the geometric transformations that are possible in 3-space and 4-space using the canonic form of the linear quaternion equation.

Using the hypercomplex operator product rules, the quaternion product of two arbitrary quaternions, $q, p \in \mathbb{H}$, written in Cartesian form as

$$p = p_0 + p_1\boldsymbol{i} + p_2\boldsymbol{j} + p_3\boldsymbol{k} \quad \text{and} \quad q = q_0 + q_1\boldsymbol{i} + q_2\boldsymbol{j} + q_3\boldsymbol{k}$$

is

$$r_0 = p_0 q_0 - p_1 q_1 - p_2 q_2 - p_3 q_3$$
$$r_1 = p_1 q_0 + p_0 q_1 - p_3 q_2 + p_2 q_3$$
$$r_2 = p_2 q_0 + p_3 q_1 + p_0 q_2 - p_1 q_3$$
$$r_3 = p_3 q_0 - p_2 q_1 + p_1 q_2 + p_0 q_3$$

where $r = pq = r_0 + r_1\boldsymbol{i} + r_2\boldsymbol{j} + r_3\boldsymbol{k}$. Gathering terms into matrix-vector notation we obtain two distinct matrix-vector products : the *standard* form

$$\begin{bmatrix} r_0 \\ r_1 \\ r_2 \\ r_3 \end{bmatrix} = \begin{bmatrix} p_0 & -p_1 & -p_2 & -p_3 \\ p_1 & p_0 & -p_3 & p_2 \\ p_2 & p_3 & p_0 & -p_1 \\ p_3 & -p_2 & p_1 & p_0 \end{bmatrix} \begin{bmatrix} q_0 \\ q_1 \\ q_2 \\ q_3 \end{bmatrix} \tag{14.1}$$

which preserves the order of the product, and the *transmuted* form

$$\begin{bmatrix} r_0 \\ r_1 \\ r_2 \\ r_3 \end{bmatrix} = \begin{bmatrix} q_0 & -q_1 & -q_2 & -q_3 \\ q_1 & q_0 & q_3 & -q_2 \\ q_2 & -q_3 & q_0 & q_1 \\ q_3 & q_2 & -q_1 & q_0 \end{bmatrix} \begin{bmatrix} p_0 \\ p_1 \\ p_2 \\ p_3 \end{bmatrix} \tag{14.2}$$

which reverses the order. Let the standard matrix form be denoted $[\![p]\!]$ and the transmuted matrix form as $[\![q]\!]^{\dagger}$, where the '$\dagger$' signifies transposing the lower-right $3{\times}3$ sub-matrix. Further, let the quaternion 4-vector be denoted $[q]$. Then the equivalent matrix-vector form for the quaternion product becomes

$$pq \leftrightarrow [\![p]\!]\,[q] = [\![q]\!]^{\dagger}\,[p]\,.$$

The transmuted form allows one to commute the order of the product as needed.

14.1.5. *Quaternion vectors and matrices*

Quaternions themselves may be represented as matrices, with either real or complex elements. The quaternion algebra is isomorphic to the algebra of complex 2×2 matrices with the following layout of elements (where w, x, y and z are the Cartesian components of the quaternion) :

$$\begin{pmatrix} w + ix & y + iz \\ -y + iz & w - ix \end{pmatrix} \tag{14.3}$$

In this form, the norm of the complex matrix is the modulus of the quaternion and the inverse of the matrix is equivalent to the inverse of the quaternion. Of course, multiplication of complex matrices is equivalent to multiplication of the quaternions represented by the matrices. The complex representation given in equation 14.3 can be extended to the case of matrices with quaternion elements, using a *block matrix* in which each quaternion element is represented by a 2×2 block of complex elements. This is described in detail by Zhang in [ZHA 97, §4] and is referred to as the *complex adjoint* matrix. Using this representation, many matrix operations may be generalised to quaternion matrices (but the numerical implementation is neither efficient nor accurate, due to the duplication of entries). Nevertheless, the existence of the complex

adjoint makes it possible to calculate some decompositions of quaternion matrices, even before a direct quaternion algorithm has been developed. This has been applied to the singular value decomposition and the eigenvalue decomposition, both of which are discussed below in § 14.5.

14.2. Hypercomplex representation of vector images

We now present two means by which a vector image may be represented using quaternions (an array of quaternions, with one element corresponding to each pixel of the image). The simplest approach, used in all the early work by the authors and others is to represent a colour image (or in general any vector image with three components per pixel) as a pure quaternion array, or as a quaternion array with zero scalar part. For example, let $f(n, m) = (r(n, m), g(n, m), b(n, m))$ be a $n \times m$ RGB colour image with components $r(n, m)$, $g(n, m)$ and $b(n, m)$, the pixels being represented by RGB triples $(r_{n,m}, g_{n,m}, b_{n,m})$. We can represent this image in quaternion form as : $q(n, m) = r(n, m)\boldsymbol{i} + g(n, m)\boldsymbol{j} + b(n, m)\boldsymbol{k}$ and manipulate it using quaternion operations as described below. After processing, for example, using Fourier transforms, it may be that we obtain a result with a small scalar part $w(n, m)$ which we discard in order to recover an RGB image result. Of course, a correctly designed operation will not usually result in a significantly non-zero scalar part – any non-zero values are usually due to rounding errors.

A more recent development, motivated by the discovery of the canonic form of a linear quaternion equation [ELL 07b], is to represent the colour image using a quaternion array in which the four components hold the components of a *homogeneous coordinate* representation of the image pixels. In homogeneous coordinates, a point in 3-space is represented by four coordinates (which we denote w, x, y, z). w is a *weight* which is arbitrarily assigned a value of 1 when the homogeneous coordinate is created (x, y and z take the values of the three components of the coordinate in 3-space). Any subset of homogeneous coordinates with the same ratio of $\frac{x}{w} : \frac{y}{w} : \frac{z}{w}$ represent the same point (provided $w > 0$). Conversion of a homogeneous coordinate to a normal 3-space coordinate requires the normalised values $\frac{x}{w}, \frac{y}{w}, \frac{z}{w}$ to be computed to yield the three components of the coordinate in 3-space. (The value of w may become different to 1 as a result of multiplications and additions of homogeneous coordinate values.)

The utility of homogeneous coordinates lies in the representation of geometric operations (and for this reason they are widely used in computer graphics). Geometric operations in homogeneous coordinates require multiplication by 4×4 real matrices, and all such transformations are *linear in the homogeneous coordinates*, even where the operation is non-linear in conventional Cartesian coordinates – a trivial example is translation (addition of a constant 3-space offset to every point). Linear operations can be combined (in this case by matrix multiplication) so that a sequence of transformations may be merged into a single 4×4 real matrix (or its equivalent linear quaternion

equation). Homogeneous coordinates also permit a useful distinction between *points* and *vectors* – a vector can be represented by a homogeneous coordinate with a weight of zero – representing a point in 3-space 'at infinity' in the direction indicated by the x, y, z components of the coordinate.

Many colour spaces used in image processing are linear transformations of the standard RGB colour space[4] [PAL 98]. Using homogeneous coordinates, transformation in and out of such colour spaces may be combined with geometrical transformations while in the colour space to yield a single transformation matrix or set of quaternion coefficients.

14.3. Hypercomplex spectral analysis

Scalar image processing has a well-developed theory covering image manipulation in the frequency domain. The most common frequency domain technique makes use of the discrete Fourier transform, where image pixels are transformed from the spatial domain into the frequency domain. Once the spectrum of the image has been obtained, computationally costly processes, such as linear filtering and image registration, become more efficient.

This section discusses several discrete Fourier transforms based on quaternions providing a holistic vector spectrum. It also presents generalized convolution and correlation definitions both of which are also well suited to vector images. But more importantly, both can be manipulated in the frequency domain using equivalent quaternion spectral operational formulae.

14.3.1. *Quaternion Fourier Transforms*

Recall the definition of the standard 2D complex Fourier transform

$$H\left[u, v\right] = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} h\left(x, y\right) e^{-i(ux+vy)} dx\,dy.$$

Here i denotes the standard complex operator $\sqrt{-1}$. Extending this definition from complex to quaternion numbers involves modifying the exponential $e^{-i(ux+vy)}$. Immediately a number of choices arise due to the non-commutative nature of quaternions. First, one could replace the standard complex operator with a single hypercomplex operator from the set $\{\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}\}$. Second, one could factor the exponential as

---

4. There are several RGB colour spaces, and transformation between them is usually a linear transformation.

$e^{-i(ux+vy)} = e^{-ivy}e^{-iux}$, then replace each complex operator with different hyper-complex operators to obtain, for example, $e^{-\boldsymbol{j}vy}e^{-\boldsymbol{k}ux}$. Notice, however, that $e^{-\boldsymbol{j}vy-\boldsymbol{k}ux} \neq e^{-\boldsymbol{j}vy}e^{-\boldsymbol{k}ux}$; so there are twelve possible combinations. Finally, since the function $h(x,y)$ and the exponential are quaternion valued, their order alters the transform result. This allows for placing the exponent to the left, right, or splitting portions to either side. Ell was the first to study this topic extensively [ELL 92, ELL 93]. His original quaternion Fourier transform was defined as :

$$H\left[\boldsymbol{j}\omega, \boldsymbol{k}\nu\right] = \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} e^{-\boldsymbol{j}\omega t} h\left(t,\tau\right) e^{-\boldsymbol{k}\nu\tau} dt d\tau.$$

The motivation for this definition was to simplify the analysis of systems of two-dimensional partial differential equations. Sangwine [SAN 96] demonstrated the applicability of this quaternion Fourier transform to colour images (using a discrete version of Ell's transform). However, subsequent attempts to implement convolution and correlation using a simple generalization of the standard complex operational formula failed, and this motivated the search for alternate formulations. The result of this search is the following quaternion Fourier transform [SAN 00] where the hypercomplex operator was generalized to $\boldsymbol{\mu} \in \mathrm{UV}[\mathbb{H}]$.

$$\mathcal{F}\left[h\left(n,m\right)\right] = \frac{1}{k} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} e^{-\boldsymbol{\mu}2\pi\left(\frac{mv}{M} + \frac{nu}{N}\right)} h\left(n,m\right) = H\left[u,v\right] \qquad (14.4)$$

with inverse

$$\mathcal{F}^{-1}\left[H\left[u,v\right]\right] = \frac{1}{k} \sum_{v=0}^{M-1} \sum_{u=0}^{N-1} e^{+\boldsymbol{\mu}2\pi\left(\frac{mv}{M} + \frac{nu}{N}\right)} H\left[u,v\right] = h\left(n,m\right)$$

where $k = \sqrt{MN}$. The operators $\boldsymbol{i}$, $\boldsymbol{j}$, $\boldsymbol{k}$ are now special cases of $\boldsymbol{\mu}$; and the complex Fourier transform is a special case of this transform in which $\boldsymbol{\mu} = \boldsymbol{i}$, and the function to be transformed is complex, the result $H\left[u,v\right]$ is also complex. This can be seen by noting that $\mathbb{C}_{\boldsymbol{i}} \subset \mathbb{H}$ can be associated with the standard complex numbers. The choice of $\boldsymbol{\mu}$, called the *axis* of the transform is arbitrary, but not without consequence.

A closely related transform is defined by reversing the order of $h(n,m)$ and the exponential [SAN 00]. This is called the 'transpose' transform. In the special complex case mentioned above, where $\boldsymbol{\mu} = \boldsymbol{i}$, the transform and its transpose yield the same result. Although the transforms defined in equation (14.4) and its transpose form are sufficient for defining operational formula, there is another definition which reduces the complexity of the operational formula. This alternative to the transform definition is obtained by noting that the choice of which transform to regard as the 'forward' transform is arbitrary. The transform defined in equation (14.4) can be utilized as a transform from the spatial frequency domain to the spatial domain, thus swapping

the forward and inverse definitions. The transform with the negative exponential is called the 'forward' transform and the one with the positive exponential is called the 'reverse' transform. The term 'inverse' transform is used to denote the transform from spatial frequency domain to the spatial domain. This gives four distinct transforms, all of which are interrelated and used.

Definition 14.3.1 *Quaternion Fourier Transform. Let $f(n,m) \in \mathbb{H}^{N \times M}$, then the quaternion Fourier transforms of $f(n,m)$ are defined as*

$$\mathcal{F}^{\pm L}[f(n,m)] = \frac{1}{k} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} e^{\mp \boldsymbol{\mu} 2\pi \left(\frac{mv}{M} + \frac{nu}{N} \right)} f(n,m) = F^{\pm L}[u,v]$$

$$\mathcal{F}^{\pm R}[f(n,m)] = \frac{1}{k} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f(m,n) e^{\mp \boldsymbol{\mu} 2\pi \left(\frac{mv}{M} + \frac{nu}{N} \right)} = F^{\pm R}[u,v]$$

where $k = \sqrt{MN}$ and the inverses are obtained by changing the sign of the exponential and summing over u and v, instead of n and m.

The quaternion fast Fourier transform (QFFT) may be implemented by decomposition into two independent complex Fourier transforms and thus be implemented by building upon existing complex code. This is significant because it makes available to researchers using hypercomplex Fourier transforms all the investment made by others in efficient complex Fast Fourier Transform (FFT) implementations and requires substantially less effort than coding hypercomplex versions of existing code. Consider definition 14.3.1 and write the image function to be transformed in symplectic form thus :

$$f(n,m) = c_1(n,m) + c_2(n,m)\,\boldsymbol{\mu}_2$$

where $c_1(n,m), c_2(n,m) \in \mathbb{C}_{\boldsymbol{\mu}_1}$, and $\boldsymbol{\mu}_1 \perp \boldsymbol{\mu}_2$. Rewriting equation (14.4) using this symplectic form for $f(n,m)$ and using the same unit pure quaternion, $\boldsymbol{\mu}_1$ for the transform axis and the 'inner' operator of the symplectic form :

$$\mathcal{F}[f(n,m)] = \frac{1}{\sqrt{MN}} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} e^{-\boldsymbol{\mu}_1 2\pi \left(\frac{mv}{M} + \frac{nu}{N} \right)} \left(c_1(n,m) + c_2(n,m)\,\boldsymbol{\mu}_2 \right)$$

This may be separated as follows :

$$\mathcal{F}[f(n,m)] = C_1[u,v] + C_2[u,v]\,\boldsymbol{\mu}_2 = F[u,v]$$

where

$$C_p[u,v] = \frac{1}{\sqrt{MN}} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} e^{-\boldsymbol{\mu}_1 2\pi \left(\frac{mv}{M} + \frac{nu}{N} \right)} c_p(n,m)$$

for $p \in \{1, 2\}$, each of which is isomorphic to the standard complex 2D Fourier transform and may be computed using standard complex code [ELL 00a].

In order to understand what the spectral coefficients represent, consider the 'inverse' transform (that which reconstructs the image from its spectral coefficients and the exponential basis function). Any given pair of horizontal and vertical frequencies is represented in the spatial frequency domain by four quaternion values, one in each of the four quadrants. (Exceptions are the Nyquist term, and any pair of frequencies including a DC or Nyquist term.) The coefficients in the four quadrants correspond to the four combinations of positive and negative horizontal and vertical spatial frequencies. Extract four terms from the spectrum at a given pair of horizontal and vertical spatial frequencies and the resulting inverse transform image will contain colours which follow an orbit in colour space.

A demonstration of this interpretation of the spectral components is done through the use of a colour-cube scatter plot. If each pixel value of an image is plotted in colour-space, the resulting scatter plot shows the distribution of the image's colour contents. The first row of figure 14.1 shows an example of the Lena image and its colour-cube scatter plot. If the original image is quaternion Fourier transformed, a single harmonic (in one direction) extracted and inverse transformed, the resulting image will show a 'rainbow grating' pattern due to harmonic oscillations in luminance and chrominance. The scatter plot of this rainbow grating image should draw an elliptic orbit about the center of the cube. The figure also shows the results of this process against the Lena image for the first five harmonics in both the horizontal and vertical directions. As expected, the resulting scatter plots for each harmonic show that the colour content is distributed in an orbital path about the mid-gray point. (Notice the magnification scale factor used to zoom into the center of the cube since the higher harmonics are typically smaller in absolute magnitude.)

14.3.2. *Convolution*

Owing to the non-commutative multiplication of quaternions the most general quaternion convolution is a two-sided convolution.

Definition 14.3.2 *Two-sided Quaternion Convolution*

$$g \prec f \succ h = \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} g_{i,j} f_{n-i,m-j} h_{i,j}$$

where f is the input image and $g, h \in \mathbb{H}^{N \times M}$ are the convolution masks.

When either of the masks is the identity, a one-sided convolution results, which gives the handed convolution definitions.

Figure 14.1. *Color-cube scatter plots of Lena image.*

Definition 14.3.3 *One-sided Quaternion Convolution. The left-handed and right-handed convolutions, respectively, are*

$$g \circ f = \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} g_{i,j} f_{n-i,m-j}, \quad and \quad f \circ h = \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} f_{n-i,m-j} h_{i,j}.$$

Using the quaternion Fourier transform, the spectral-domain formula for the one-sided convolution is given as [ELL 07c] :

$$\mathcal{F}^{\pm L} [f \circ h] = k \left\{ F_1^{\pm L} [u, v] H^{\pm L} [u, v] + F_2^{\pm L} [u, v] \, \boldsymbol{\mu}_2 H^{\mp L} [u, v] \right\} \qquad (14.5)$$

and

$$\mathcal{F}^{\pm R} [f \circ h] = k \left\{ F^{\pm R} [u, v] H_1^{\pm R} [u, v] + F^{\mp R} [u, v] H_2^{\pm R} [u, v] \, \boldsymbol{\mu}_2 \right\} \qquad (14.6)$$

where $k = \sqrt{NM}$ and $F [u, v]$ and $H [u, v]$ are symplectically decomposed as :

$$F^{\pm L} [u, v] = F_1^{\pm L} [u, v] + F_2^{\pm L} [u, v] \, \boldsymbol{\mu}_2,$$

and

$$H^{\pm R} [u, v] = H_1^{\pm R} [u, v] + H_2^{\pm R} [u, v] \, \boldsymbol{\mu}_2,$$

respectively, depending on whether the left or right Fourier transform is used.

A direct operational formula for the bi-convolution equation is unnecessary since any bi-convolution can instead be rewritten as a sum of (at most) four one-sided convolutions.

Theorem 14.3.1 *Bi-Convolution as handed-convolution. Let* $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2 \in \mathrm{UV}\,[\mathbb{H}]$ *and* $\boldsymbol{\mu}_1 \perp \boldsymbol{\mu}_2$. *Decompose the left mask and image function into symplectic form as* $g = g_1 + g_2\,\boldsymbol{\mu}_2$, *and* $f = f_1 + f_2\,\boldsymbol{\mu}_2$ *where* $f_1, f_2, g_1, g_2 \in \mathbb{C}_{\boldsymbol{\mu}_1}$, *then*

$$g \prec f \succ h = f_1 \circ (g_1 h) - \bar{f}_2 \circ (g_2 h) + \boldsymbol{\mu}_2 \left\{ \bar{f}_2 \circ (\bar{g}_1 h) + f_1 \circ (\bar{g}_2 h) \right\}$$

The proof is a straightforward application of the symplectic swap rule discussed in §14.1.3. See [ELL 07c] for details.

Now it is a straightforward matter of applying the one-sided spectral convolution equations (14.5) and (14.6) to each of the four convolutions. Since the quaternion Fourier transform is a linear operation, the image function can be transformed first, then split into symplectic form.

14.3.3. *Correlation*

The discrete version of the cross-correlation of two *real*-valued functions is defined as

$$f * g = \sum_{p=-\infty}^{\infty} \sum_{q=-\infty}^{\infty} f_{p,q} g_{p-n,q-m} = \sum_{p=-\infty}^{\infty} \sum_{q=-\infty}^{\infty} f_{p+n,q+m} g_{p,q}$$

Here substituting $n' = p - n$ and $m' = q - m$ yields the second equality. So, unlike the convolution of two real-valued functions, the correlation of two real-valued functions does not commute, i.e., $f * g \neq g * f$. For real-valued functions this definition is sufficient. However, for complex valued functions, a conjugate operation is required to ensure that the autocorrelation surface $R_f(n, m)$ and the power-spectral density $S_f[u, v]$ constitute a Fourier-transform pair as required by the *Wiener-Khintchine Theorem* [MAR 65, p.169]. So for complex-valued functions the discrete correlation formula becomes

$$f * g = \sum_{p=-\infty}^{\infty} \sum_{q=-\infty}^{\infty} f_{p,q} \bar{g}_{p-n,q-m} = \sum_{p=-\infty}^{\infty} \sum_{q=-\infty}^{\infty} \bar{f}_{p+n,q+m} g_{p,q}.$$

The last equality uses the fact that shift reversal cancels conjugation and that conjugation of a product does not reorder the product, i.e., $\overline{z_1 z_2} = \bar{z}_1 \bar{z}_2$ for any two complex numbers. Further, it requires that the direction of the shift $(p{\pm}n, q{\pm}m)$ be immaterial.

However, in the case of image processing shift direction is important, as for example in image registration [MOX 03].

Taking all of this into consideration, there are two equivalent definitions for the cross-correlation of two images f and g when extended to hypercomplex images [SAN 99].

Definition 14.3.4 *Quaternion Cross-Correlation. Let $f, g \in \mathrm{V}\left[\mathbb{H}\right]^{N \times M}$, then their cross-correlation is given by*

$$f * g = \sum_{q=0}^{M-1} \sum_{p=0}^{N-1} f_{p,q} \bar{g}_{p-n,q-m} = \sum_{q=0}^{M-1} \sum_{p=0}^{N-1} f_{p+m,q+n} \bar{g}_{p,q} = R_{fg} \qquad (14.7)$$

where the shift operator on f and g is implemented cyclically using modulo arithmetic.

R_{fg} is the correlation surface of the images as an $N \times M$ image. If the correlated functions are the same, *i.e.*, $f = g$, the auto-correlation surface R_f is computed. These equations effectively treat the images as one tile sliding over an infinite plane of tiles. The direct evaluation of the cross-correlation is impractical for all but the smallest images due to the high computational costs ($O\left(N^4\right)$ for a $N \times N$ image). This necessitates the use of the fast quaternion Fourier transform given earlier.

Using the Fourier transforms (§ 14.3.1) and symplectic decomposition the spectral form of the cross-correlation is given as [ELL 00b] :

$$R_{fg} = \mathcal{F}^{\mp R}\left\{F^{+R}\overline{G_{\parallel}^{\pm L}} + F^{-R}\overline{G_{\perp}^{\pm L}}\right\}$$

or

$$R_{fg} = \mathcal{F}^{\mp L}\left\{F_{\parallel}^{\pm L}\overline{G^{\pm R}} + F_{\perp}^{\pm L}\overline{G^{\mp R}}\right\}$$

where $G = \mathcal{F}\left\{g\right\}$, $F = \mathcal{F}\left\{f\right\}$, $\left(G_{\parallel}, F_{\parallel}\right) \parallel \boldsymbol{\mu}$, and $\left(G_{\perp}, F_{\perp}\right) \perp \boldsymbol{\mu}$.

14.4. Linear filtering

Fundamental to linear image filtering is the convolution operator. Recall the definition of the real 2D discrete convolution :

$$y_{n,m} = \sum_{k} \sum_{l} m_{k,l}\, x_{n-k,m-l}$$

where $m, x \in \mathbb{R}$. Each summand, $m_{k,l}\, x_{n-k,m-l}$, is a linear function that scales the image pixel value x by the mask value m. At its basic arithmetic level convolutions

consist of shifted and scaled accumulations based on the relative location of scalar values in the convolution filter mask. Scalar image convolution is therefore restricted to the scaling of pixel-values. Whereas an individual summand of the real convolution forms the classical linear function, $y = mx$, a summand of the quaternion bi-convolution is one part of a multi-part quaternion linear function given by :

$$y = \sum_{p=1}^{P} m_p \, x \, n_p$$

where $m_p, x, n_p \in \mathbb{H}$. Without the loss of generality, functions of this type can be reduced to at most four terms as discussed in § 14.1.4. This upper limit is not as obvious as it first appears ; non-commutative quaternion multiplication causes difficulty in reducing functions with more terms.

Theorem 14.4.1 *Quaternary canonical form. Let $m_p, n_p, q \in \mathbb{H}$. Without loss of generality the multi-nomial linear equation*

$$f(q) = \sum_{p=1}^{P} m_p \, q \, n_p$$

can be reduced to

$$f(q) = A \, q + B \, q \, \boldsymbol{i} + C \, q \, \boldsymbol{j} + D \, q \, \boldsymbol{k} \tag{14.8}$$

where

$$A = \tfrac{1}{2} \sum_{p=1}^{P} m_p \, (n_p + \bar{n}_p), \quad B = \tfrac{1}{2} \sum_{p=1}^{P} m_p \, (\boldsymbol{i}\bar{n}_p\boldsymbol{i} + \bar{n}_p) \, \boldsymbol{i},$$

$$C = \tfrac{1}{2} \sum_{p=1}^{P} m_p \, (\boldsymbol{j}\bar{n}_p\boldsymbol{j} + \bar{n}_p) \, \boldsymbol{j}, \quad D = \tfrac{1}{2} \sum_{p=1}^{P} m_p \, (\boldsymbol{k}\bar{n}_p\boldsymbol{k} + \bar{n}_p) \, \boldsymbol{k}.$$

One can expand the right-hand terms as $n_p = n_p^w + n_p^x\boldsymbol{i} + n_p^y\boldsymbol{j} + n_p^z\boldsymbol{k}$, then each summand can be factored as $m_pqn_p = m_pn_p^wq + m_pn_p^xq\boldsymbol{i} + m_pn_p^yq\boldsymbol{j} + m_pn_p^zq\boldsymbol{k}$ which results in the required number of four terms. However the formulas above for A, B, C and D require no knowledge of the Cartesian components of n_p. For a detailed proof see [ELL 07b].

Once reduced to canonical form, linear combinations of such functions can be maintained in this form. Let two such functions be given as : $f_1(q) = A_1q + B_1q\,\boldsymbol{i} + C_1q\,\boldsymbol{j} + D_1q\,\boldsymbol{k}$ and $f_2(q) = A_2q + B_2q\,\boldsymbol{i} + C_2q\,\boldsymbol{j} + D_2q\,\boldsymbol{k}$ then their sum $f_1 + f_2$ is given by sum of the individual components, i.e.,

$$f_1(q) + f_2(q) = A_3q + B_3q\,\boldsymbol{i} + C_3q\,\boldsymbol{j} + D_3q\,\boldsymbol{k}$$

where $A_3 = A_1 + A_2$, $B_3 = B_1 + B_2$, $C_3 = C_1 + C_2$, and $D_3 = D_1 + D_2$. Likewise, their composition, $f_2\left(f_1(q)\right)$, is given by

$$f_2\left(f_1(q)\right) = A_3 q + B_3 q \boldsymbol{i} + C_3 q \boldsymbol{j} + D_3 q \boldsymbol{k}$$

where

$$A_3 = A_2 A_1 - B_2 B_1 - C_2 C_1 - D_2 D_1,$$

$$B_3 = A_2 B_1 + B_2 A_1 - C_2 D_1 + D_2 C_1,$$

$$C_3 = A_2 C_1 + B_2 D_1 + C_2 A_1 - D_2 B_1,$$

$$D_3 = A_2 D_1 - B_2 C_1 + C_2 B_1 + D_2 A_1.$$

Given this canonical form, the general quaternion convolution (given in definition 14.3.2) can be rewritten as

$$y_{n,m} = \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} \left\{ \begin{array}{l} A_{i,j} f_{n-i,m-j} + B_{i,j} f_{n-i,m-j} \boldsymbol{i} \\ + C_{i,j} f_{n-i,m-j} \boldsymbol{j} + D_{i,j} f_{n-i,m-j} \boldsymbol{k} \end{array} \right\}.$$

Or in left-convolution operator notation

$$y = (A \circ f) + (B \circ f)\boldsymbol{i} + (C \circ f)\boldsymbol{j} + (D \circ f)\boldsymbol{k}.$$

Now the canonical linear quaternion function can be written in matrix form as

$$A q + B q \boldsymbol{i} + C q \boldsymbol{j} + D q \boldsymbol{k} \leftrightarrow \mathbf{M}\left[q\right]$$

where $\mathbf{M} \in \mathbb{R}^{4 \times 4}$ is given by

$$\mathbf{M} = [\![A]\!] + [\![B]\!]\,[\![i]\!]^{\dagger} + [\![C]\!]\,[\![j]\!]^{\dagger} + [\![D]\!]\,[\![k]\!]^{\dagger}.$$

What should be noted is that $\mathbf{M}$ does not necessarily represent a *single* quaternion, which has only four degrees of freedom. But it is always true that any *arbitrary* matrix $\mathbf{M}$ can be written as an equivalent canonical linear quaternion function. Further, if $\mathbf{M} \in GL_4(\mathbb{R})$ (the set of invertible 4×4 matrices), then $f(q) \leftrightarrow \mathbf{M}\left[q\right]$ is invertible also.

14.4.1. *Geometric filtering*

It will now be shown that the linear functions in canonical form can be used to perform a number of 3-space geometric transformations on pixel vectors beyond scaling such as : rotations, shears, translations, reflections and even perspective projections of the vectors. These image pixel 'point'-operators are then used to build convolution masks. The transformations discussed are applied to the image pixels in the space of

the samples; these are not transformations applied to an image's index. For example, when we talk of a rotation we mean rotation of pixel values, not a rotation of the image about its index system; hence colour image pixel rotation about the grey line $r = g = b$ changes the hue of a pixel, not its saturation or luminance.

Given the equivalence between canonical linear functions and members of $GL_4(\mathbb{R})$, we can start with the geometric transforms possible using matrices in $GL_4(\mathbb{R})$. The following equation diagrams the possibilities for a unit-weight point $[q]^T = [1, x, y, z]$:

$$
\begin{bmatrix} w' \\ x' \\ y' \\ z' \end{bmatrix} = \left[\begin{array}{c|ccc} k & p_x & p_y & p_z \\ \hline t_x & u_x & s_{xy} & s_{xz} \\ t_y & s_{yx} & u_y & s_{yz} \\ t_z & s_{zx} & s_{zy} & u_z \end{array} \right] \begin{bmatrix} 1 \\ x \\ y \\ z \end{bmatrix}
$$

where the unit-weight output point is given by $[q']^T = [1, x'/w', y'/w', z'/w']$. In this equation t_i correspond to translations, u_i to non-uniform scaling, s_{ij} to shears, p_i to projections, and k to uniform scaling, where $i, j \in \{x, y, z\}$ and $i \neq j$. The 3×3 sub-matrix consisting of the u_i's and s_{ij}'s can be fashioned as a direction-cosine matrix to yield rotations, and with appropriate sign changes to yield reflections. In short, all 3-space geometric transformations are available with this single matrix equation. Individual operations can be cascaded to create more complex transformations by multiplying their individual matrices together.

One could brute-force build the canonical linear functions by first constructing the individual operations in $GL_4(\mathbb{R})$, then convert them to quaternion functions, however, a more elegant way is to construct them directly. We give just two examples to show their expressive power. Each of these is overly complicated when represented in $GL_4(\mathbb{R})$. More examples can be found in [ELL 07a].

The first example is a radial dilation function. Radial dilations are a set of transformations that expand 3-dimensional space outwards from an invariant line defined by a unit quaternion $\boldsymbol{\mu}$ using a scale factor α. The transformation is represented by

$$
f(q) = \frac{2 + \alpha}{2} q + \frac{\alpha}{2} \boldsymbol{\mu} q \boldsymbol{\mu}
$$

where $\alpha > 0$ expands space, and $\alpha < 0$ compresses space. As a simple example of what this transformation achieves, if applied to an RGB encoded image, with the grey-line $r = g = b$ as the invariant line, it allows saturation to be increased or decreased without affecting hue or luminance.

The second example is an axial neutral shear function. They are axial neutral since the axis line has no shear effect. The axial shear equation is

$$
f(q) = q \left(1 - \tfrac{\alpha}{2} \boldsymbol{\mu}_3 \right) - \tfrac{\alpha}{2} \boldsymbol{\mu}_2 q \boldsymbol{\mu}_1,
$$

where $\boldsymbol{\mu}_3$ is the axis of the shear and α is the shear factor. It is required that $\boldsymbol{\mu}_1 \perp \boldsymbol{\mu}_2 \perp \boldsymbol{\mu}_3$ and $\boldsymbol{\mu}_1\boldsymbol{\mu}_2 = \boldsymbol{\mu}_3$ (*i.e.*, they form an orthogonal right-handed triad). This requirement ensures that vectors are mapped to vectors.

14.4.2. *Examples*

We illustrate the use of hypercomplex geometric filtering through a chrominance-edge detection example. This filter will have the maximum response to colour transitions that are opponent colours in the Hering opponent colour $(R_g Y_b W_k)$ space. It will not respond to luminance changes, but will respond to saturation changes. Further, the filtered image will represent the colour transitions in the output that are coloured in accordance to the direction of the transition in the opponent colour space. There are four canonical functions needed to build such a convolution filter.

The first function maps the input image from RGB to $R_g Y_b W_k$ colour space. Since this mapping is linear, it can be represented as

$$f_1\left(q\right) = \tfrac{1}{4}\left(2 - \boldsymbol{i} - \boldsymbol{j}\right)q + \tfrac{1}{4}\left(1 - 2\boldsymbol{i} + 2\boldsymbol{j} - \boldsymbol{k}\right)q\boldsymbol{i}$$
$$+ \tfrac{1}{4}\left(1 + 2\boldsymbol{i} + 2\boldsymbol{j} - 3\boldsymbol{k}\right)q\boldsymbol{j} + \tfrac{1}{4}\left(-\boldsymbol{i} - 3\boldsymbol{j} - 2\boldsymbol{k}\right)q\boldsymbol{k}.$$

To remove the luminance information we need a second function which projects pixel values onto the mid-grey plane in the $R_g Y_b W_k$ colour space. To enhance the output response (for better viewing) we scale the chrominance information uniformly by an arbitrary factor of 10. The function which does this is

$$f_2\left(q\right) = \tfrac{1}{4}\left(21\right)q - \tfrac{1}{4}\left(\boldsymbol{i}\right)q\boldsymbol{i} - \tfrac{1}{4}\left(\boldsymbol{j}\right)q\boldsymbol{j} + \tfrac{1}{4}\left(19\boldsymbol{k}\right)q\boldsymbol{k}.$$

A scalar-image edge detector usually consists of neighboring mask elements with opposing signs such as

$$\begin{bmatrix} +1 & +1 & +1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix}$$

for a horizontal edge detector. A direct generalization of this concept (*i.e.*, the ±1's) for *vector* pixels is to rotate the pixels in the upper row clockwise $90°$ and the pixels in the lower row counter-clockwise by the same amount. In a uniformly coloured region of the image the results of the mask operation exactly cancel (since the vectors will point in opposite directions). For an image region which contains a transition from one colour to its opponent colour the rotations align the pixel vectors in the same direction giving the maximum result. (This of course assumes that the mid-grey point is the origin of the colour space, which is how we encode the RGB values when converting them to a quaternion weighted-points.) Therefore, the third function needed is a $90°$ rotation about the grey-line, in $R_g Y_b W_k$ coordinates, and is given by

$$f_3\left(q\right) = \tfrac{1}{2}\left(1 + \boldsymbol{k}\right)q - \tfrac{1}{2}\left(1 + \boldsymbol{k}\right)q\boldsymbol{k}$$

The fourth function is used to 'tune' the output response of the filter. We will rotate output colours, after conversion back to RGB space, by a $45°$ angle about the grey-line using

$$
f_4\left(q\right) = \begin{array}{llll}
(& 0.8536 & + & 0.20410\,\boldsymbol{i} & + & 0.20410\,\boldsymbol{j} & + & 0.20410\,\boldsymbol{k})\,q \\
+ & (-0.2041 & - & 0.04882\,\boldsymbol{i} & - & 0.04882\,\boldsymbol{j} & - & 0.04882\,\boldsymbol{k})\,q\,\boldsymbol{i} \\
+ & (-0.2041 & - & 0.04882\,\boldsymbol{i} & - & 0.04882\,\boldsymbol{j} & - & 0.04882\,\boldsymbol{k})\,q\,\boldsymbol{j} \\
+ & (-0.2041 & - & 0.04882\,\boldsymbol{i} & - & 0.04882\,\boldsymbol{j} & - & 0.04882\,\boldsymbol{k})\,q\,\boldsymbol{k}
\end{array}
$$

(This is a purely arbitrary choice used to illustrate the power of this technique, as there is no additional computational burden in doing this step.) We are now ready to construct the final horizontal edge-detection mask using these quaternion functions as

$$
\begin{bmatrix}
U & U & U \\
0 & 0 & 0 \\
L & L & L
\end{bmatrix}
$$

where $U\left(q\right) = f_4 \circ f_1^{-1} \circ f_3 \circ f_2 \circ f_1$, and $L\left(q\right) = f_4 \circ f_1^{-1} \circ f_3^{-1} \circ f_2 \circ f_1$.

Figures 14.2 and 14.3 show the input and output images, respectively, for this colour-edge detector.

Figure 14.2. *Original Image*

Figure 14.3. *Edge-Detection Results*

14.5. Quaternion Principal Component Analysis

Principal Component Analysis (PCA) is a well established tool in image and signal processing. Its applications are countless and can be found in standard textbooks. Here, we consider the extension of PCA to the case where the data are quaternion valued. The idea of Hypercomplex PCA originated in polarized seismic signal processing [Le 04] and was then developed for colour images [Le 03]. Here, we present

some colour image denoising results based on hypercomplex PCA and a rank reduction technique.

14.5.1. *Hypercomplex matrices and vectors*

Matrices and vectors of quaternions are a straightforward generalization of real or complex valued matrices and vectors. A quaternion valued matrix composed of N rows and M columns, denoted $\mathbf{A}$ is an element of $\mathbb{H}^{N \times M}$. A quaternion valued vector of dimension N is an element of $\mathbb{H}^N$. Obviously, a quaternion matrix $\mathbf{A} \in \mathbb{H}^{N \times M}$ can be written in its Cartesian form as : $\mathbf{A} = \mathbf{A}_w + \mathbf{A}_x \boldsymbol{i} + \mathbf{A}_y \boldsymbol{j} + \mathbf{A}_z \boldsymbol{k}$, where $\mathbf{A}_{w,x,y,z} \in \mathbb{R}^{N \times M}$ are its real matrix components. The same Cartesian form exists for vectors. Here we consider right vector spaces over $\mathbb{H}$ [Le 04] which means that scalars multiply vectors from the right. This choice is motivated by the fact that this convention leads to 'standard' notation for matrix manipulation and also because the theory of right eigenvalues is well established for quaternion matrices, while the use of left eigenvalues is not obvious from their theory [ZHA 97].

An N-dimensional vector space is called a *right* vector space over the set of quaternions $\mathbb{H}$ if, $\forall \boldsymbol{v} \in \mathbb{H}^N$ and $\xi, \mu \in \mathbb{H}$, the following stands : $(\boldsymbol{v}\xi)\mu = \boldsymbol{v}(\xi\mu)$. Given a quaternion valued vector $\boldsymbol{v} \in \mathbb{H}^N$, then its transpose is $\boldsymbol{v}^T$ while its conjugate-transpose is $\boldsymbol{v}^\dagger$. The scalar product of two quaternion vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ is given by : $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \boldsymbol{u}\boldsymbol{v}^\dagger = \sum_{\alpha=1}^N \overline{u}_\alpha v_\alpha$. The norm of a quaternion vector is defined as : $\|\boldsymbol{u}\| = \langle \boldsymbol{u}, \boldsymbol{u} \rangle^{\frac{1}{2}}$. Two quaternion vectors $\boldsymbol{u}$ and $\boldsymbol{v} \in \mathbb{H}$ are said to be *orthogonal* if $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = 0$. It is possible to demonstrate [Le 04] that a right vector space over $\mathbb{H}$ can be decomposed into a sum of orthogonal and linearly independent quaternion vectors.

Most of the quaternion matrix properties can be found in [ZHA 97]. Here we list some properties without proofs. The Frobenius norm of a quaternion matrix $\mathbf{A} \in \mathbb{H}^{N \times M}$ is $\|\mathbf{A}\| = \sum_{i=1}^N \sum_{j=1}^M |a_{ij}|^2$. A quaternion matrix is called unitary if $\mathbf{A}\mathbf{A}^\dagger = \mathbf{Id}$ where $\mathbf{Id}$ is the real identity matrix. Any quaternion matrix $\mathbf{A} \in \mathbb{H}^{N \times M}$ can be written as [ZHA 97, Le 04] :

$$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^\dagger$$

where $\mathbf{U} \in \mathbb{H}^{N \times N}$ and $\mathbf{V} \in \mathbb{H}^{N \times N}$ are unitary matrices and where $\mathbf{D} \in \mathbb{R}^{N \times N}$ is a diagonal matrix with the singular values of $\mathbf{A}$ on its diagonal. The rank of a quaternion matrix is given by the number of its non-null singular values[5]. A quaternion matrix is called *Hermitian* if $\mathbf{A} = \mathbf{A}^\dagger$. More properties of quaternion matrices can be found in [ZHA 97].

---

5. Note that the rank definition is not obvious for quaternion matrices due to the non-commutativity of the multiplication over $\mathbb{H}$, but can be defined in a non-ambigious way using complex representation of quaternion matrices (see [ZHA 97] for example).

14.5.2. *Eigenvalue decomposition of Hermitian quaternion matrices*

As we will be interested in the eigenvalue decomposition (EVD) of the covariance matrix, which is Hermitian by construction, we only present the EVD for such matrices here.

A Hermitian quaternion matrix $\mathbf{A} \in \mathbb{H}^{N \times N}$ can be decomposed like :

$$\mathbf{A} = \mathbf{U}\Delta\mathbf{U}^{\dagger} \tag{14.9}$$

where $\mathbf{U} \in \mathbb{H}^{N \times N}$ is a unitary quaternion matrix and $\Delta \in \mathbb{R}^{N \times N}$ is a real diagonal matrix with the eigenvalues $\delta_{\beta}, \beta = 1, \ldots, r$ of $\mathbf{A}$ on its diagonal (r is the rank of $\mathbf{A}$). Since $\mathbf{U}$ is unitary, its columns are orthogonal, and thus it is possible to write $\mathbf{A}$ as a sum of rank 1 subspaces :

$$\mathbf{A} = \sum_{\beta=1}^{r} \delta_{\beta}\boldsymbol{u}_{\beta}\boldsymbol{u}_{\beta}^{\dagger} = \sum_{\beta=1}^{l} \delta_{\beta}\boldsymbol{u}_{\beta}\boldsymbol{u}_{\beta}^{\dagger} + \sum_{\beta=l+1}^{r} \delta_{\beta}\boldsymbol{u}_{\beta}\boldsymbol{u}_{\beta}^{\dagger} = \mathbf{A}_l + \mathbf{A}_{r-l} \tag{14.10}$$

where the subspaces spanned by $\mathbf{A}_l$ and $\mathbf{A}_{r-l}$ are orthogonal. Now, assuming that the eigenvalues are ranked in decreasing order of magnitude, then a truncation of rank l of $\mathbf{A}$ consists in considering $\mathbf{A}_l$. As proved in [Le 04], this rank-l approximation of the matrix $\mathbf{A}$ is the best approximation (in the mean square sense) as the error ε :

$$\varepsilon = \text{Trace}\left[(\mathbf{A} - \mathbf{A}_l)^{\dagger}(\mathbf{A} - \mathbf{A}_l)\right] = \sum_{\beta=l+1}^{r} \delta_{\beta}^2 \tag{14.11}$$

is minimum (see [Le 04]). Note also that thanks to (14.10) the following is true :

$$(\mathbf{A} - \mathbf{A}_l)\mathbf{A}_l^{\dagger} = 0$$

The subspace defined by $\mathbf{A}_l$ is made up of the sum of l rank-1 subspaces corresponding to the highest eigenvalues of $\mathbf{A}$.

In order to compute the eigenvalues of a quaternion Hermitian matrix, there exist two main possibilities :

   1) To use the isomorphism between quaternions and 2×2 complex matrices

   2) To compute the decomposition in quaternion arithmetic.

The first possibility uses the fact that any quaternion matrix $\mathbf{A} \in \mathbb{H}^{N \times M}$ can be written in its Cayley-Dickson form as : $\mathbf{A} = \mathbf{A}_{(1)} + \mathbf{A}_{(2)}\boldsymbol{j}$, with $\mathbf{A}_{(1)}, \mathbf{A}_{(2)} \in \mathbb{C}^{N \times M}$. Then, it is possible to define for $\mathbf{A}$ its *complex adjoint* matrix $\chi_{\mathbf{A}} \in \mathbb{C}^{2N \times 2M}$ as :

$$\chi_{\mathbf{A}} = \begin{pmatrix} \mathbf{A}_{(1)} & \mathbf{A}_{(2)} \\ -\overline{\mathbf{A}}_{(2)} & \overline{\mathbf{A}}_{(1)} \end{pmatrix}$$

The eigenvalues of $\chi_\mathbf{A}$ are directly linked with those of $\mathbf{A}$ (see [ZHA 97] for details). The use of $\chi_\mathbf{A}$ to estimate the eigenvalues of $\mathbf{A}$ has the advantage that it can be done using algorithms working in complex arithmetic. The main drawback is the computation time, as it is necessary to compute the EVD of a matrix whose size is twice that of the original matrix. This approach was used in [Le 04]. For the second possibility recent results in quaternion matrix algebra [Le 07, SAN 06] may be used to compute the eigenvalues directly in quaternion arithmetic. The main advantage of this approach is computation time as well as accuracy of the eigenvalues. Such algorithms are part of the QTFM (Quaternion Toolbox For Matlab) [SAN 05].

14.5.3. *Quaternion PCA for colour image denoising and dimensionality reduction*

The use of quaternion PCA could be transposed to any PCA-like problem where quaternion models arise for colour images. Here, we mention the denoising and dimensionality reduction problem. In fact, this former problem is nothing more than a noise-free PCA of the original image and a lower-dimension representation of the image. We present the noisy case before considering the dimensionality reduction special case.

Consider the case of a colour image corrupted by some white Gaussian noise. We use the quaternion representation of colour images, *i.e.* the image of interest $\mathbf{X} \in V[\mathbb{H}]^{N \times N}$ has N rows and N columns and each pixel is a pure quaternion. The noise we consider here is a matrix $\mathbf{N} \in V[\mathbb{H}]^{N \times N}$ where each pure quaternion is a 3D Gaussian random variable with covariance $\Sigma_\mathbf{N} = \sigma^2 \mathbf{Id}_3$. The noise is supposed decorrelated from the image of interest. The noisy colour image $\mathbf{Y} \in V[\mathbb{H}]^{N \times N}$ can thus be expressed as :

$$\mathbf{Y} = \mathbf{X} + \mathbf{N} \tag{14.12}$$

PCA consists in finding an estimate of $\mathbf{X}$ given $\mathbf{Y}$ with the assumptions previously stated. Before discussing the limitations of quaternion PCA, we briefly explain its principle. First, assuming decorrelation between the noise and the signal, then the covariance of $\mathbf{Y}$ is given by :

$$\boldsymbol{\Gamma}_\mathbf{Y} = \mathbb{E}\left[\mathbf{X}\mathbf{X}^\dagger\right] + \mathbb{E}\left[\mathbf{N}\mathbf{N}^\dagger\right] = \boldsymbol{\Gamma}_\mathbf{X} + \sigma^2 \mathbf{Id} \tag{14.13}$$

Practically, the mathematical expectation $\mathbb{E}[.]$ has to be replaced by some sub-image averaging, assuming ergodicity and stationarity of the noise $\mathbf{N}$, as we only have one realization of the noisy image. Then, computing the EVD of $\boldsymbol{\Gamma}_\mathbf{Y}$, one gets :

$$\boldsymbol{\Gamma}_\mathbf{Y} = \sum_{\alpha=1}^{N_s} \delta_\alpha \boldsymbol{u}_\alpha \boldsymbol{u}_\alpha + \sum_{\alpha=N_s+1}^{r} \delta_\alpha \boldsymbol{u}_\alpha \boldsymbol{u}_\alpha \tag{14.14}$$

where N_S is the dimension of the *signal subspace* (here the image of interest $\mathbf{X}$) and r is the rank of $\mathbf{Y}$. Natural images are in fact full rank images, which means that a N_s truncation will be an approximation of the original image. However, if the approximation is good for small values of N_s, then the denoising will be efficient. The noise is spread over all the eigenvalues of $\mathbf{\Gamma_Y}$ and thus the noise reduction is of the order of $r/(r - N_s)$. Using the eigenvectors of $\mathbf{\Gamma_Y}$ corresponding to the highest eigenvalues (and so the *signal subspace*), one can build the projector on this subspace of dimension N_s :

$$\mathcal{P}_{Sig} = \sum_{\alpha=1}^{N_s} \boldsymbol{u}_\alpha \boldsymbol{u}_\alpha^\dagger \tag{14.15}$$

This projector must be applied to the data $\mathbf{Y}$ to obtain an estimate of the original image $\mathbf{X}$ in the following way :

$$\widetilde{\mathbf{X}} = \mathcal{P}_{Sig} \mathbf{Y} \tag{14.16}$$

where the symbol $\widetilde{\ }$ stands for the estimated value of the quantity of interest. The decomposition of $\mathbf{\Gamma_Y}$ into orthogonal subspaces thanks to the quaternion EVD ensures orthogonality between $\widetilde{\mathbf{X}}$ and the remaining part $\mathbf{Y} - \widetilde{\mathbf{X}}$ which is composed of the high frequencies of the original image $\mathbf{X}$ and the noise $\mathbf{N}$. Note that quaternion PCA does not allow a perfect reconstruction and only minimises the mean square error between the estimated and original image under the above mentioned assumptions. This limitation of PCA-like techniques is well known and has motivated the large amount of work on Independent Component Analysis (ICA) which tries to overcome such problems in searching for statistically independent components rather than orthogonal ones (but still with an assumption of Gaussian noise). We mention that ICA has been extended to quaternion valued signals [Le 06], but has not been studied in the framework of colour image denoising up to now. Considering now the case of dimensionality reduction, it is straightforward from equation (14.16) to see that if $\mathcal{P}_{Sig}$ is estimated using a noise free image (*i.e.* with $\mathbf{N} = \mathbf{0}$ in (14.12)), then one gets for $\widetilde{\mathbf{X}}$ a lower rank (specified by N_s) representation of $\mathbf{X}$. Thanks to the orthogonality constraints imposed during quaternion PCA, the error between $\mathbf{X}$ and $\widetilde{\mathbf{X}}$ is minimal in the mean square sense. In the following section, we present some results of this rank reduction technique.

14.5.4. *Examples*

We illustrate the use of quaternion PCA on colour images with an example of dimensionality reduction. In figure 14.4, the original image of size 256×256 is presented, left. Then, in figures 14.4, right and 14.5, the 'reduced versions' of lower ranks (respectively 1, 20 and 100) are displayed. In order to have an idea of the reconstruction error, we give in figure 14.6 this error as a function of the rank truncation r.

From figure 14.5, it is obvious that a rank-100 reconstruction is already a good approximation of the original image, even if the original image was of full rank (256).

MassiveMulticomponent/LeBihanSangwine/Fig1_nlb.eps   MassiveMulticomponent/LeBihanSangwine/Fig2_nlb.eps

Figure 14.4. *Original (left) and rank-1 truncation (right) images*

MassiveMulticomponent/LeBihanSangwine/Fig3_nlb.eps   MassiveMulticomponent/LeBihanSangwine/Fig4_nlb.eps

Figure 14.5. *Rank-20 (left) and rank-100 truncation (right) images*

MassiveMulticomponent/LeBihanSangwine/Fig5_nlb.eps
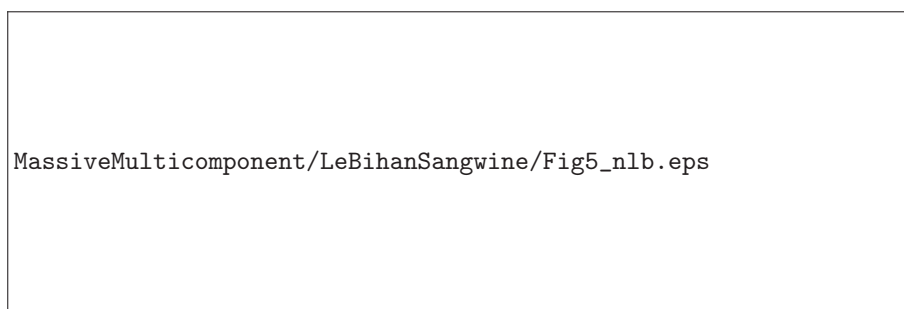
Figure 14.6. *Reconstruction error (mean square)*

This is confirmed in figure 14.6 where it can be seen how the image is well reconstructed with a low rank approximation. Most of the colour and spatial information is contained in the low ranks. Then the reconstruction error tends to zero as the rank tends to 256. Similar results were obtained on other images in [Le 03]. This confirms that as a dimensionality reduction technique, the quaternion PCA offers a practical tool that handles colour in a very natural way.

14.6. Conclusions and future directions

In this chapter we have given an overview of the utility of hypercomplex, particularly quaternion, methods for representing and processing vector images, with particular examples drawn from the field of colour image processing. Quaternion Fourier transforms are now well understood, and coupled with the developing understanding of linear quaternion systems and their links to the algebra of the general linear group, $GL_4(\mathbb{R})$, there is much promise of new linear filter developments. This is particularly promising where linear implementations are possible using homogeneous coordinates, but the operation performed in the space of the image pixels themselves is non-linear, because it is clear that linear operations in the space of the pixel values are very limited in effect. At the same time, new developments in hypercomplex generalisations of linear algebra can be applied to vector signal and image processing. The principal component analysis technique can be extended to the case of quaternion valued images, thanks to recent advances in quaternion matrix decomposition. Possible applications of quaternion PCA remain to be investigated. Further work should include the study of quaternion independent component analysis (ICA) for colour images and its potential use for denoising of colour images. The applications of linear algebra in signal and image processing tend to be at a deep level of algorithmic design, and this field is not yet so far advanced because it was started about 5 years later than the study of Fourier transforms and linear filtering. However, the tools now available for research make it inevitable that new developments will occur quite rapidly.

[ALF 07] ALFSMANN D., GÖCKLER H. G., « On Hyperbolic Complex LTI Digital Systems », *Proceedings of EUSIPCO 2007, 15th European Signal Processing Conference*, Poznan, Poland, European Association for Signal Processing, p. 1332–1336, 3–7 septembre 2007.

[COX 46] COXETER H. S. M., « Quaternions and Reflections », *American Mathematical Monthly*, vol. 53, n°3, p. 136-146, mars 1946.

[ELL 92] ELL T. A., Hypercomplex Spectral Transformations, PhD thesis, University of Minnesota, Minneapolis, MN, USA, 1992.

[ELL 93] ELL T. A., « Quaternion-Fourier transforms for analysis of two-dimensional linear time-invariant partial differential systems », *Decision and Control, 1993., Proceedings of the 32nd IEEE Conference on*, vol. 2, p. 1830–1841, 15-17 Dec. 1993.

[ELL 00a] ELL T. A., SANGWINE S. J., « Decomposition of 2D Hypercomplex Fourier Transforms into Pairs of Complex Fourier Transforms », GABBOUJ M., KUOSMANEN P., Eds.,

Proceedings of EUSIPCO 2000, Tenth European Signal Processing Conference, vol. II, Tampere, Finland, European Association for Signal Processing, p. 1061-1064, 5–8 septembre 2000.

[ELL 00b] ELL T. A., SANGWINE S. J., « Hypercomplex Wiener-Khintchine Theorem with Application to Color Image Correlation », *IEEE International Conference on Image Processing (ICIP 2000)*, vol. II, Vancouver, Canada, Institute of Electrical and Electronics Engineers, p. 792-795, 11–14 septembre 2000.

[ELL 07a] ELL T. A., « Hypercomplex Color Affine Filters », *IEEE International Conference on Image Processing (ICIP 2007)*, vol. V, San Antonio, TX, Institute of Electrical and Electronics Engineers, p. 249-252, 16–19 septembre 2007.

[ELL 07b] ELL T. A., « On Systems of Linear Quaternion Functions », Preprint : arXiv :math/0702084v1, février 2007.

[ELL 07c] ELL T. A., SANGWINE S. J., « Hypercomplex Fourier Transforms of Color Images », *IEEE Trans. Image Process.*, vol. 16, n°1, p. 22–35, janvier 2007.

[ELL 07d] ELL T. A., SANGWINE S. J., « Linear Quaternion Systems Toolbox for Matlab® », [Online], 2007, Software library, available at : http ://lqstfm.sourceforge.net/.

[HAM 53] HAMILTON W. R., *Lectures on Quaternions*, Hodges and Smith, Dublin, 1853, Available online at Cornell University Library : http ://historical.library.cornell.edu/math/.

[KAN 89] KANTOR I. L., SOLODNIKOV A. S., *Hypercomplex numbers, an elementary introduction to algebras*, Springer-Verlag, New York, 1989.

[Le 03] LE BIHAN N., SANGWINE S. J., « Quaternion principal component analysis of color images », *Image Processing, 2003. ICIP 2003. Proceedings 2003 International Conference on*, vol. 1, p. I-809-12, septembre 2003.

[Le 04] LE BIHAN N., MARS J., « Singular value decomposition of quaternion matrices : a new tool for vector-sensor signal processing », *Signal Processing*, vol. 84, n°7, p. 1177–1199, 2004.

[Le 06] LE BIHAN N., BUCHHOLZ S., « Quaternionic Independent Component Analysis using hypercomplex nonlinearities », 7th *IMA Conference on Mathemathics in Signal Processing*, Cirencester, England, 2006.

[Le 07] LE BIHAN N., SANGWINE S. J., « Jacobi Method for Quaternion Matrix Singular Value Decomposition », *Applied Mathematics and Computation*, vol. 187, n°2, p. 1265–1271, 15 April 2007.

[MAR 65] MARSHALL J. L., *Introduction to Signal Theory*, International Textbook Company, Scranton, Pennsylvania, 1965.

[MCC 00] MCCABE A., CAELLI T., WEST G., REEVES A., « Theory of Spatiochromatic Image Encoding and Feature Extraction », *Journal of the Optical Society of America A*, vol. 17, n°10, p. 1744-1754, octobre 2000.

[MIR 05] MIRON S., Méthodes Multilinéaires et Hypercomplexes en Traitement d'Antenne Multicomposante Haute Résolution, PhD thesis, Institut National Polytechnique de Grenoble, France, octobre 2005, Available : http ://tel.archives-ouvertes.fr/.

[MOX 03] MOXEY C. E., SANGWINE S. J., ELL T. A., « Hypercomplex Correlation Techniques for Vector Images », *IEEE Trans. Signal Process.*, vol. 51, n°7, p. 1941–1953, juillet 2003.

[PAL 98] PALUS H., « Representations of Colour Images in Different Colour Spaces », Sangwine, Horne [SAN 98a], Chapitre 4, p. 67-90.

[PEI 97] PEI S.-C., CHENG C.-M., « A Novel Block Truncation Coding of Color Images using a Quaternion-Moment-Preserving Principle », *IEEE Trans. Commun.*, vol. 45, n°5, p. 583-595, 1997.

[PEI 99] PEI S. C., CHENG C. M., « Color Image Processing by Using Binary Quaternion Moment-Preserving Thresholding Technique », *IEEE Trans. Image Process.*, vol. 8, n°5, p. 614-628, 1999.

[SAI 08] SAID S., LE BIHAN N., SANGWINE S. J., « Fast complexified quaternion Fourier transform », *IEEE Trans. Signal Process.*, vol. 56, n°4, p. 1522–1531, avril 2008.

[SAN 96] SANGWINE S. J., « Fourier Transforms of Colour Images using Quaternion, or Hypercomplex, Numbers », *Electronics Letters*, vol. 32, n°21, p. 1979-1980, 10 octobre 1996.

[SAN 98a] SANGWINE S. J., HORNE R. E. N., Eds., *The Colour Image Processing Handbook*, Chapman and Hall, London, 1998.

[SAN 98b] SANGWINE S. J., THORNTON A. L., « Frequency Domain Methods », Sangwine, Horne [SAN 98a], Chapitre 12, p. 228-241.

[SAN 99] SANGWINE S. J., ELL T. A., « Hypercomplex auto- and cross-correlation of color images », *Image Processing, 1999. ICIP 99. Proceedings. 1999 International Conference on*, vol. 4, Kobe, Japan, p. 319–322, 24-28 Oct. 1999.

[SAN 00] SANGWINE S. J., ELL T. A., « The Discrete Fourier Transform of a Colour Image », BLACKLEDGE J. M., TURNER M. J., Eds., *Image Processing II Mathematical Methods, Algorithms and Applications*, Chichester, Horwood Publishing for Institute of Mathematics and its Applications, p. 430-441, 2000, Proceedings Second IMA Conference on Image Processing, De Montfort University, Leicester, UK, September 1998.

[SAN 05] SANGWINE S. J., LE BIHAN N., « Quaternion Toolbox for Matlab® », [Online], 2005, Software library available at : http ://qtfm.sourceforge.net/.

[SAN 06] SANGWINE S. J., LE BIHAN N., « Quaternion Singular Value Decomposition based on Bidiagonalization to a Real or Complex Matrix using Quaternion Householder Transformations », *Applied Mathematics and Computation*, vol. 182, n°1, p. 727–738, 1 November 2006.

[SAN 08] SANGWINE S. J., « Canonic form of linear quaternion functions », Preprint : arXiv :0801.2887v1, janvier 2008.

[WAE 76] VAN DER WAERDEN B. L., « Hamilton's Discovery of Quaternions », *Mathematics Magazine*, vol. 49, n°5, p. 227–234, novembre 1976.

[WAR 97] WARD J. P., *Quaternions and Cayley Numbers : Algebra and Applications*, vol. 403 de *Mathematics and Its Applications*, Kluwer, Dordrecht, 1997.

[ZHA 97] ZHANG F., « Quaternions and matrices of quaternions », *Linear algebra and its applications*, vol. 251, p. 21-57, 1997.

Chapitre 15

Panoramic integral-field spectrograph : huge ultraspectral data to understand Universe history

In this chapter, we present a new method to classify specific objects in astronomy, the galaxies, with new multiband observation that are now available within Virtual Observatory framework (see next paragraph). To do such a task, we think that multichannel observation should be used to classify such objects according to their global behavior, and not only the shape of the pattern within the single optical band. We propose to decompose each object into structures and to reconstruct them (inverse problem) by using multichannel observations with a Bayesian model and Monte Carlo sampling. The next step, will be to extend this approach of galaxy decomposition to hyperspectral data given by MUSE instrument which is under construction. We conclude this chapter by some perspectives, drawing how we would like to try to analyse huge amount of data, that are today only simulated but that are scheduled to arrive at the begining of 2012.

15.1. Virtual Observatory (VO) and new objectives

With the development of Virtual Observatory (VO) since 2001, astronomers have now access, through internet network, to huge amount of image data, taken with different telescope, with different time exposure, with different point spread function

---

(psf), etc. Data access has been solved in the past years by the development of inter-operability tools which are now efficient, even if some improvements remains possible. The main goal, now, consists in managing such huge amount of data, to link them together, to evaluate their redundancy and exploit their uncertainty. To achieve this goal, we develop data fusion methods to obtain a target image to be inferred from a number of blurred noisy images from a single or different sensors under various conditions of acquisition (various lattice grid, varying shift, random orientation, blur, noise, etc). We propose to use Bayesian framework to invert the forward model which describes accurately the image formation process. By this way, we try to take into account some a priori knowledge for automatic registration and spatial resampling, in order to be able to address the ill-posed problem in a rigourous Bayesian scheme. Our goal consists in the modelling of the whole instrument acquisition process, to outperform existing methods such as resample-and-add or drizzling methods since we can handle different instrument behavior and compute uncertainty estimate as well. Moreover, we develop this approach in a recursive way to fuse ultraspectral cubes from MUSE intrument, so that the model can be updated when new data become available. This chapter is organized as follows : in section (2.2) galaxy classification is presented, whereas in section (2.3) we present a new pattern decomposition to classify them. Section (2.4) present Bayesian model and Monte Carlo sampling for parameter estimation. Section (2.5) show some results we obtained on simulated and raw data. Finally, we draw the perspective of ultraspectral fusion on huge data set via a Bayesian perspective.

15.2. Galaxy classification goals

Galaxy classification is a necessary step in analysing and then understanding the evolution of these objects in relation to their environment at different spatial scales. Current classifications rely mostly on the De Vaucouleurs scheme [1] which is an evolution of the original idea by Hubble. These classifications are based only on the visible aspect of galaxies and identifies five major classes : ellipticals, lenticulars, spirals with or without bar, and irregulars. Each class is characterized by the presence, with different strengths, of physical structures such as a central bright bulge, an extended fainter disc, spiral arms, . . . and each class and the intermediate cases are themselves divided into finer stages.

Nowadays wide astronomical image surveys provide huge amount of multiwave-length data. For example, the Sloan Digital Sky Survey (SDSS[1]) has already produced more than 15 Tb of 5-band images. Nevertheless, most classifications still do not take advantage of colour information, although this information gives important clues on galaxy evolution allowing astronomers to estimate the star formation history, the current amount of dust, etc. This observation motivates the research of a more efficient

---

1. http ://www.sdss.org/

classification including spectral information over all available bands. Moreover due to the quantity of available data (more than 930,000 galaxies for the SDSS), it appears relevant to use an automatic and unsupervised method.

Two kinds of methods have been proposed to automatically classify galaxies following the Hubble scheme. The first one measures galaxy features directly on the image (*e.g.* symmetry index [2], Pétrosian radius [3], concentration index [4], *clumpiness* [5], . . .). The second one is based on decomposition techniques (shapelets [6], the basis extracted with principal component analysis [7], and the pseudo basis modelling of the physical structures : bulge and disc [8]). Parameters extracted from these methods are then used as the input to a traditional classifier such as a support vector machine [9], a multi layer perceptron [10] or a Gaussian mixture model [6].

These methods are now able to reach a good classification efficiency (equal to the experts' agreement rate) for major classes [7]. Some attempts have been made to use decomposition into shapelets [11] or feature measurement methods [12] on multispectral data by processing images band by band. Fusion of spectral information is then performed by the classifier. But the lack of physical meaning of data used as inputs for the classifiers makes results hard to interpret. To avoid this problem we propose to extend the decomposition method using physical structures to multiwavelength data. This way we expect that the interpretation of new classes will be straightforward.

In this context, three 2D galaxy decomposition methods are publicly available. Gim2D [13] performs bulge and disc decomposition of distant galaxies using MCMC methods, making it robust but slow. Budda [14] handles bulge, disc, and stellar bar, while Galfit [15] handles any composition of structures using various brightness profiles. Both of them are based on deterministic algorithms which are fast but sensitive to local minima. Because these methods cannot handle multispectral data, we propose a new decomposition algorithm. This works with multispectral data and any parametric structures. Moreover, the use of MCMC methods makes it robust and allows it to work in a fully automated way.

In Sec. 15.3, we extend current models to multispectral images. Then, we present in Sec. 15.4 the Bayesian approach and a suitable MCMC algorithm to estimate model parameters from observations. The first results on simulated and raw images are discussed in Sec. 15.5. Finally perspectives are drawn in Sec. 15.6.

15.3. Galaxy Model

Decomposition into Structures

It is widely accepted by astronomers that spiral galaxies for instance can be decomposed into physically significant structures such as bulge, disc, stellar bar and

spiral arms (Fig. 15.4, first column). Each structure has its own particular shape, populations of stars and dynamic. The bulge is a spheroidal population of mostly old red stars located in the centre of the galaxy. The disc is a planar structure with different scale heights which includes most of the gas and dust if any and populations of stars of various ages and colour from old red to younger and bluer ones. The stellar bar is an elongated structure composed of old red stars across the galaxy centre. Finally, spiral arms are over-bright regions in the disc that are the principal regions of star formation. The visible aspect of these structures are the fundamental criterion in the Hubble classification. It is noteworthy that this model only concerns regular galaxies and that no model for irregular or peculiar galaxies is available.

We only consider in this chapter bulge, disc, and stellar bar. Spiral arms are not included because no mathematical model including both shape and brightness informations is available ; we are working at finding such a suitable model.

Structure Model

We propose in this section a multispectral model for bulge, disc, and stellar bar. These structures rely on the following components : a generalized ellipse (also known as super ellipse) is used as a shape descriptor and a Sérsic law is used for the brightness profile [16]. These two descriptors are flexible enough to describe the three structures.

The major axis r of a generalized ellipse centred at the origin with axis parallel to coordinate axis and passing trough point $(x, y) \in \mathbb{R}^2$ is given by :

$$r(x, y) = \left(|x|^{c+2} + \left| \frac{y}{e} \right|^{c+2} \right)^{\frac{1}{c+2}} \tag{15.1}$$

where e is the ratio of the minor to the major axis and c controls the misshapenness : if $c = 0$ the generalized ellipse reduces to a simple ellipse, if $c < 0$ the ellipse is said to be *disky* and if $c > 0$ the ellipse is said to be *boxy* (Fig. 15.1). Three more parameters are needed to complete shape information : the centre (c_x, c_y) and the position angle α between abscissa axis and major axis.

The Sérsic law [16] is generally used to model the brightness profile. It is a generalization of the traditional exponential and De Vaucouleurs laws usually used to model disc and bulge brightness profiles. Its high flexibility allows it to vary continuously from a nearly flat curve to a very piked one (Fig. 15.2). The brightness at major axis r is given by :

$$I(r) = I \exp -k_n \left(\left(\frac{r}{R} \right)^{\frac{1}{n}} - 1 \right) \tag{15.2}$$

where R is the effective radius, n is the Sérsic index, and I the brightness at the effective radius. k_n is an auxiliary function such that $\Gamma(2n) = 2\gamma(2n, k_n)$ to ensure

Figure 15.1. *Left : a simple ellipse with position angle α, major axis r and minor axis r/e. Right : generalized ellipse with variations of parameter c (displayed near each ellipse).*

Figure 15.2. *The Sérsic law for different Sérsic index n. $n = 0.5$ yields a Gaussian, $n = 1$ yields an exponential profile and for $n = 4$ we obtain the De Vaucouleurs profile.*

that half of the total flux is contained in the effective radius (Γ and γ are respectively the complete and incomplete gamma function).

Then, the brightness at pixel (x, y) is given by :

$$F(x, y) = (F_1(x, y), \ldots, F_B(x, y)) \tag{15.3}$$

with B the number of bands and the brightness in band b is defined as :

$$F_b(x, y) = I_b \exp -k_{n_b} \left(\left(\frac{r(x, y)}{R_b} \right)^{\frac{1}{n_b}} - 1 \right) \tag{15.4}$$

As each structure is supposed to represent a particular population of stars and galactic environment, we also assume that shape parameters do not vary between bands. This strong assumption seems to be verified in observations suggesting that shape

variations between bands is negligible compared with deviation induced by noise. Moreover, this assumption reduces significantly the number of unknowns. The stellar bar has one more parameter which is the cut-off radius R_{max} ; its brightness is zero beyond this radius. For the bulge (respectively the stellar bar), all Sérsic parameters are free which leads to a total number of $5 + 3B$ (respectively $6 + 3B$) unknowns. For the disc, parameter c is set to zero and Sérsic index is set to one leading to $4 + 2B$ free parameters. Finally, we assume that the centre is identical for all structures yielding a total of $11 + 8B$ unknowns.

Observation Model

Atmospheric distortions can be approximated by a spatial convolution with a Point Spread Function (PSF) H given as a parametric function or an image. Other noises are a composition of several sources and will be approximated by a Gaussian noise $\mathcal{N}(0, \Sigma)$. Matrix Σ and PSF H are not estimated as they can be measured using a deterministic procedure. Let Y be the observations and e the noise, we then have :

$$Y = Hm + e \qquad \text{with} \qquad m = F_{\mathfrak{B}} + F_{\mathfrak{D}} + F_{\mathfrak{B}a} \qquad (15.5)$$

with $\mathfrak{B}$, $\mathfrak{D}$, and $\mathfrak{B}a$ denoting respectively the bulge, the disc, and the stellar bar.

15.4. Bayesian Model and Monte Carlo Sampling

The problem being clearly ill-posed, we adopt a Bayesian approach. Priors assigned to each parameter are summarized in Tab. 15.1 ; they were determined from literature when possible and empirically otherwise. Indeed experts are able to determine limits for parameters but no further information is available : that is why Probability Density Functions (pdf) of chosen priors are uniformly distributed. However we expect to be able to determine more informative priors in future work. The posterior reads then :

$$P\left(\phi|Y\right) = \frac{1}{\left(2\pi\right)^{\frac{N}{2}} \det\left(\Sigma\right)^{\frac{1}{2}}} \exp -\frac{1}{2}\left(Y - Hm\right)^{T}\Sigma^{-1}\left(Y - Hm\right)P\left(\phi\right) \quad (15.6)$$

where $P\left(\phi\right)$ denotes the priors and ϕ the unknowns. Due to its high dimensionality it is intractable to characterize the posterior pdf with sufficient accuracy. Instead, we aim at finding the Maximum A Posteriori (MAP). Because of the posterior complexity, the need for a robust algorithm leads us to choose MCMC methods [17]. MCMC algorithms are proven to converge in infinite time, and in practice the time needed to obtain a good estimation may be quite long. Thus several methods are used to improve

convergence speed : simulated annealing, adaptive scale [18] and direction [19] Hastings Metropolis (HM) algorithm. As well, highly correlated parameters like Sérsic index and radius are sampled jointly to improve performance. The main algorithm is a Gibbs sampler consisting in simulating variables separately according to their respective conditional posterior. One can note that the brightness factors posterior reduces to a truncated positive Gaussian $\mathcal{N}^+\left(\mu, \sigma^2\right)$ which can be efficiently sampled using an accept-reject algorithm [20]. Other variables are generated using the HM algorithm.

Some are generated with a Random Walk HM (RWHM) algorithm whose proposal is a Gaussian. At each iteration a random move from the current value is proposed. The proposed value is accepted or rejected with respect to the posterior ratio with the current value. The parameters of the proposal have been chosen by examining several empirical posterior distributions to find preferred directions and optimal scale. Sometimes the posterior is very sensitive to input data and no preferred directions can be found. In this case we decided to use the Adaptive Direction HM (ADHM). ADHM algorithm uses a sample of already simulated points to find preferred directions. As it needs a group of points to start with we choose to initialize the algorithm using simple RWHM. When enough points have been simulated by RWHM, the ADHM algorithm takes over. Algorithm and parameters of proposal distributions are summarized in Tab. 15.1.

Also, parameters I_b, R_b, and n_b are jointly simulated. R_b, n_b are first sampled according to $P\left(R_b, n_b \mid \phi_{\backslash\{R_b, n_b, I_b\}}\right)$ where I_b has been integrated and then I_b is sampled [21]. Indeed, the posterior can be decomposed in :

$$P\left(R_b, n_b, I_b \mid \phi_{\backslash\{R_b, n_b, I_b\}}, Y\right) = P\left(R_b, n_b \mid \phi_{\backslash\{R_b, n_b, I_b\}}, Y\right) P\left(I_b \mid \phi_{\backslash\{I_b\}}, Y\right)$$
(15.7)

15.5. Validation and Results

We measured two values for each parameter : the MAP and the variance of the chain in the last iterations. The latter gives an estimation of the uncertainty on the estimated value. A high variance can have different interpretations. In case of an observation with a low SNR, the variance naturally increases. But the variance can also be high when a parameter is not relevant. For example, the position angle is significant if the structure is not circular, the radius is also significant if the brightness is strong enough. We have also checked visually the residual image (the difference between the observation and the simulated image) which should contain only noise and non modelled structures.

Parameters are initialized by generating random variables according to their priors. This procedure ensures that the algorithm is robust so that it will not be fooled by a bad initialisation, even if the burn-in period of the Gibbs sampler is quite long (about 1,500 iterations corresponding to 1.5 hours).

Tableau 15.1. *Parameters and their priors. All proposal distributions are Gaussians whose covariance matrix (or deviation for scalars) are given in the last column.*

Structure	Parameter	Prior Support	Algorithm
$\mathfrak{B}, \mathfrak{B}a, \mathfrak{D}$	centre (c_x, c_y)	Image domain	RWHM with $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$
$\mathfrak{B}$	major to minor axis (e)	$[1; 10]$	RWHM with 1
	position angle (α)	$[0; 2\pi]$	RWHM with 0.5
	ellipse misshapenness (c)	$[-0.5; 1]$	RWHM with 0.1
	brightness factor (I)	$\mathbb{R}^+$	direct with $\mathcal{N}^+\left(\mu, \sigma^2\right)$
	radius (R)	$[0; 200]$	ADHM with $\begin{pmatrix} 0.16 & -0.02 \\ -0.02 & 0.01 \end{pmatrix}$
	Sérsic index (n)	$[1; 10]$	
$\mathfrak{D}$	major to minor axis (e)	$[1; 10]$	RWHM with 0.2
	position angle (α)	$[0; 2\pi]$	RWHM with 0.5
	brightness factor (I)	$\mathbb{R}^+$	direct with $\mathcal{N}^+\left(\mu, \sigma^2\right)$
	radius (R)	$[0; 200]$	RWHM with 1
$\mathfrak{B}a$	major to minor axis (e)	$[4; 10]$	RWHM with 1
	position angle (α)	$[0; 2\pi]$	RWHM with 0.5
	ellipse misshapenness (c)	$[0.6; 2]$	RWHM with 0.1
	brightness factor (I)	$\mathbb{R}^+$	direct with $\mathcal{N}^+\left(\mu, \sigma^2\right)$
	radius (R)	$[0; 200]$	ADHM with $\begin{pmatrix} 0.16 & -0.02 \\ -0.02 & 0.01 \end{pmatrix}$
	Sérsic index (n)	$[0.5; 10]$	
	cut-off radius (R_{max})	$[10; 100]$	RWHM with 1

Test on Simulated Images

We have validated the procedure on simulated images to test the ability of the algorithm to recover input parameters. The results showed that the algorithm is able to provide a solution leading to a residual image containing only noise (Fig. 15.3). Some parameters like elongation, position angle, or centre are retrieved with a very good precision (relative error less than 0.1%). On the other hand, Sérsic parameters are harder to estimate. Thanks to the extension of the disc, its radius and its brightness, are estimated with a relative error of less than 5%. For the bulge and the stellar bar, the situation is complex because information is held by only a few pixels and an error in the estimation of Sérsic parameters does not lead to a high variation in the likelihood. Although the relative error increases to 20%, the errors seem to compensate each other.
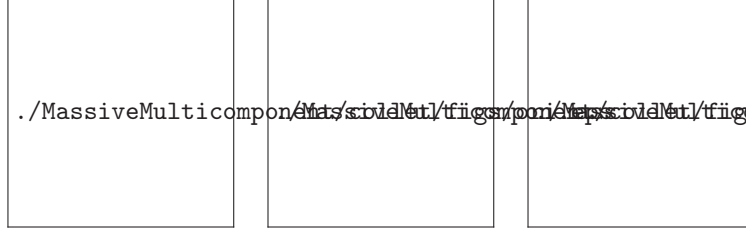
Figure 15.3. *Example of estimation on a simulated image (only one band on five is shown). Left : simulated galaxy with a bulge, a disc and a stellar bar. Centre : estimation. Right : residual. Images are given in inverse gray scale with enhanced contrast.*

Another problem is the evaluation of the presence of a given structure. Because the algorithm seeks at minimizing the residual, all the structures are always used. This can lead to solutions where structures have no physical significance. Therefore, we tried to introduce a Bernoulli variable coding the structure occurrence. Unfortunately, we were not able to determine a physically significant Bernoulli parameter. Instead we could use a pre- or post-processing method to determine the presence of each structure. These questions are highly linked to the astrophysical meaning of the structures we are modelling and we have to ask ourselves why some structures detected by the algorithm should in fact not be used. As claimed before, we need to define more informative joint priors.

Test on Real Images

We have performed tests on about 30 images extracted from the EFIGI database [7] which is composed of thousands of galaxy images extracted from the SDSS. Images are centred on the galaxy but may contain other objects (stars, galaxies, artefacts, ...). Experiments showed that the algorithm performs well as long as no other bright object is present in the image (see Fig. 15.4 for example). As there is no ground truth available on real data we compared the results of our algorithm on monospectral images with those provided by Galfit. This shows a very good agreement since Galfit estimations are within the confidence interval proposed by our method.

Computation Time

Most of the computation time is used to evaluate the likelihood. Each time a parameter is modified, this implies the recomputation of the brightness of each affected structure for all pixels. Processing 1,000 iterations on a 5-band image of 250×250 pixels takes about 1 hour with a Java code running on an Intel Core 2 processor (2,66 GHz). We are exploring several ways to improve performance such as providing
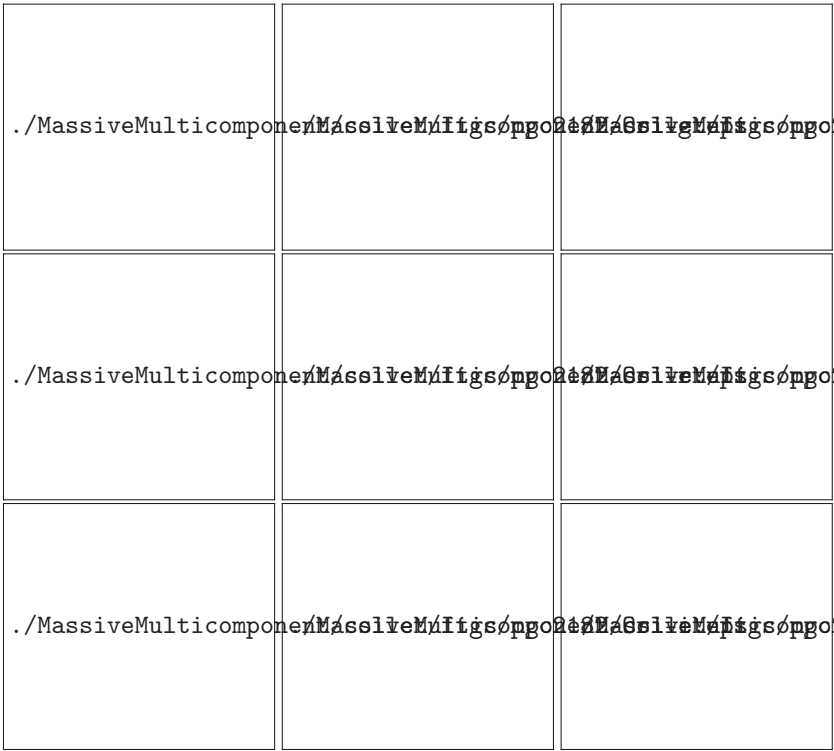
Figure 15.4. *Left column : galaxy PGC2182 (bands g, r, and i) is a barred spiral. Centre column : estimation. Right column : residual. Images are given in inverse gray scale with enhanced contrast.*

a good initialisation using fast algorithms or finely tuning the algorithm to simplify exploration of the posterior pdf.

15.6. Perspectives, future requirements for astronomical image analysis and fusion of huge data set

The Multi Unit Spectroscopic Explorer (MUSE) is a second generation VLT panoramic integral-field spectrograph operating in the visible wavelength range. MUSE has a field of 1x1 arcmin, sampled at 0.2x0.2 arcsec (Wide Field Mode, hereafter WFM), and of 7.5x7.5 arcsec, sampled at 25 milli-arcsec (Narrow Field Mode, hereafter NFM), both assisted by adaptive optics. The simultaneous spectral range is 0.465-0.93 μm, at a resolution of R 3000. This instrument will produce in 2012 the first observations of the far Universe, with high spatial and spectral resolution through cube of size 300 by 300 pels on 4000 spectral bands at each of the 90.000 locations.

A single observation (data cube) will weight 1.5 Gbyte. MUSE couples the disco-very potential of a large imaging device to the measuring capabilities of a high-quality spectrograph, while taking advantage of the increased spatial resolution provided by adaptive optics. This makes MUSE a unique and tremendously powerful instrument for discovering and characterizing objects that lie beyond the reach of even the dee-pest imaging surveys. The most challenging scientific and technical application, and the most important driver for the instrument design, is the study of the progenitors of normal nearby galaxies out to redshifts beyond 6. These systems are extremely faint and can only be found by their $Ly\alpha$ emission. MUSE will be able to detect these in large numbers (15.000 for a typical program) through a set of nested surveys of different area and depth. The deepest survey will require very long integrations (80 hrs each field), corresponding to 80 ultraspectral cubes of 1 hour exposure each / 1.5 Tbytes each, let be a total of 100 Gbytes. In this way, the instrument will be able to reach extreme depths by means of long total integrations.

The overall complexity of integral-field spectrographic data and the size of indivi-dual datacubes will somewhat prevent the user from having a direct and easy handle of the scientific content of the datasets. Indeed, the impact of atmosphere (including AO correction) ; PSF varying with field of view, wavelength and spatial position ; at-mospheric dispersion effects ; cirrus absorption ; night sky continuum and emission as function of time and moon ; sky absorption as function of zenital distance, including telluric absorption ; detector readout noise and dark current ; geometric distorsion ; flatfield and wavelength calibration residuals etc. require to modelize the wole MUSE instrument before any image processing.

Multisource data fusion for bandlimited signals : a Bayesian perspective

A new approach of fusion has already been theoretically validated with Space-Fusion program grant (lsiit-miv.u-strasbg.fr/paseo) funded by an ANR (JC05-41500) Jeune Chercheur in 2005, for multi-source data fusion and super-resolution from re-mote sensing and astronomical images [22, 23]. In this study, we focused on astrono-mical images, obtained through different telescopes. We are going to generalize this approach to astronomical images given by MUSE at different time with the same ins-trument : the main challenge will consist in developing an iterative fusion process to deal with spectral and spatial high resolution. These images suffer from aliasing (fi-nite spatial discretization depending on the location), from varying spectral psf, from atmospheric turbulence and cosmic rays, etc. We are going to reconstruct an image through probabilistic framework, wih Shannon sampling theory, geometry and noise modeling as basic ingredients. Indeed, we believe that Bayesian theory is most appro-priate for handling uncertainties (stochastic and systematic errors) in the images : as we combine observations from a single source at different time, we expect a reduction of uncertainties [23]. Thus, the image fusion problem can be stated as the computation of the posteriori probability density function of the unknown image given the set of

observations. Via Bayes' rule, joint and posterior pdfs are proportional given the data (unwanted variables are implicitely integrated out). It is crucial to understand image formation (forward model corresponding how MUSE image formation is physically done) in order to effectively invert the acquisition process and reconstruct the actual origin of the set of measurements. Whereas the functions involved in the likelihood are constrained by the physic of the used instrument, the priors should only enforce objective constraints, or generic properties easy to check experimentally (e.g., image smoothness).

De Vaucouleurs, G. : Classification and Morphology of External Galaxies. Handbuch der Physik 53, 275 (1959)

Yagi, M., Nakamura, Y., Doi, M., Shimasaku, K. and Okamura, S. : Morphological classification of nearby galaxies based on asymmetry and luminosity concentration. Monthly Notices of Roy. Astr. Soc. 368, 211–220 (2006)

Petrosian, V. : Surface brightness and evolution of galaxies. Astrophys. J. Letters 209, L1–L5 (1976)

Abraham, R. G., Valdes, F., Yee, H. K. C. and van den Bergh, S. : The morphologies of distant galaxies. 1 : an automated classification system. Astrophys. J. 432, 75–90 (1994)

Conselice, C. J. : The Relationship between Stellar Light Distributions of Galaxies and Their Formation Histories. Astrophys. J. Suppl. S. 147, 1–28 (2003)

Kelly, B. C. and McKa, T. A. : Morphological Classification of Galaxies by Shapelet Decomposition in the Sloan Digital Sky Survey. Astron. J. 127, 625–645 (2004)

Baillard, A., Bertin, E., Mellier, Y., McCracken, H. J., Géraud, T., Pelló, R., Leborgne, F. and Fouqué, P. : Project EFIGI : Automatic Classification of Galaxies. Astron. Soc. Pac. Conf. ADASS XV 351, 236 (2006)

Allen, P. D., Driver, S. P., Graham, A. W., Cameron, E., Liske, J. and de Propris, R. : The Millennium Galaxy Catalogue : bulge-disc decomposition of 10095 nearby galaxies. Monthly Notices of Roy. Astr. Soc. 371, 2–18 (2006)

Tsalmantza, P., Kontizas, M., Bailer-Jones, C. A. L., Rocca-Volmerange, B., Korakitis, R., Kontizas, E., Livanou, E., Dapergolas, A., Bellas-Velidis, I., Vallenari, A. and Fioc, M. :

Towards a library of synthetic galaxy spectra and preliminary results of classification and parametrization of unresolved galaxies for Gaia : Astron. Astrophys. 470, 761–770 (2007)

Bazell, D. : Feature relevance in morphological galaxy classification. Monthly Notices of Roy. Astr. Soc. 316, 519–528 (2000)

Kelly, B. C. and McKay, T. A. : Morphological Classification of Galaxies by Shapelet Decomposition in the Sloan Digital Sky Survey. II. Multiwavelength Classification. Astron. J. 129, 1287–1310 (2005)

Lauger, S., Burgarella, D. and Buat, V. : Spectro-morphology of galaxies : A multi-wavelength (UV-R) classification method. Astron. Astrophys. 434, 77-87 (2005)

Simard, L., Willmer, C. N. A., Vogt, N. P., Sarajedini, V. L., Phillips, A. C., Weiner, B. J., Koo, D. C., Im, M., Illingworth, G. D. and Faber, S. M. : The DEEP Groth Strip Survey. II. Hubble Space Telescope Structural Parameters of Galaxies in the Groth Strip. Astrophys. J. Suppl. S. 142, 1–33 (2002)

de Souza, R. E., Gadotti, D. A. and dos Anjos, S. : BUDDA : A New Two-dimensional Bulge/Disk Decomposition Code for Detailed Structural Analysis of Galaxies. Astrophys. J. Suppl. S. 153, 411–427 (2004)

Peng, C. Y., Ho, L. C., Impey, C. D. and Rix, H.-W. : Detailed Structural Decomposition of Galaxy Images. Astron. J. 124, 266–293 (2002)

Sérsic, J. L. : Atlas de galaxias australes. Cordoba, Argentina : Observatorio Astronomico (1968)

Gilks, W.R., Richardson, S. and Spiegelhalter, D.J. : Markov Chain Monte Carlo In Practice. Chapman & Hall/CRC, Washington, D.C. (1996)

Gilks, W. R., Roberts, G. O. and Sahu, S. K. : Adaptive Markov chain Monte Carlo through regeneration. J. Amer. Statistical Assoc. 93, 1045–1054 (1998)

Roberts, G. O. and Gilks, W. R. : Convergence of adaptive direction sampling. J. of Multivariate Ana. 49, 287–298 (1994)

Mazet, V., Brie, D. and Idier, J. : Simulation of positive normal variables using several proposal distributions. IEEE Workshop on Statistical Sig. Proc, 37–42 (2005)

Devroye, L. : Non-Uniforme Random Variate Generation. Springer-Verlag, New York (1986)

Jalobeanu, A. and Gutierrez, J.A. : Multisource data fusion for bandlimited signals : a Bayesian perspective. Proc. of 25th workshop on Bayesian Inference and Maximum Entropy methods (MaxEnt'06), Paris, France, Aug 2006

Jalobeanu, A., Gutierrez, J.A. and Slezak, E. : Multisource data fusion and super-resolution from astronomical images. Statistical Methodology (STAMET), Special issue on Astrostatistics, 5(4), Jul 2008