

# Statistical-Based Approaches for Noise Removal

State Luminița<sup>1</sup>, Cătălina-Lucia Cocianu<sup>1</sup> and Vlamos Panayiotis<sup>2</sup>

<sup>1</sup>University of Pitești

<sup>2</sup>Academy of Economic Studies

<sup>3</sup>Ionian University

<sup>1,2</sup>Romania

<sup>3</sup>Greece

## 1. Introduction

Image restoration methods are used to improve the appearance of an image by the application of a restoration process based on a mathematical model to explain the way the image was distorted by noise. Examples of types of degradation include blurring caused by motion or atmospheric disturbance, geometric distortion caused by imperfect lenses, superimposed interference patterns caused by mechanical systems, and noise induced by electronic sources.

Usually, it is assumed that the degradation model is either known or can be estimated from data. The general idea is to model the degradation process and then apply the inverse process to restore the original image. In cases when the available knowledge does not allow to adopt a reasonable model for the degradation mechanism it becomes necessary to extract information about the noise directed by data and then to use this information for restoration purposes. The knowledge about the particular generation process of the image is application specific. For example, it proves helpful to know how a specific lens distorts an image or how mechanical vibration from a satellite affects an image. This information can be gathered from the analysis of the image acquisition process and by applying image analysis techniques to samples of degraded images.

The restoration can be viewed as a process that attempts to reconstruct or recover a degraded image using some available knowledge about the degradation mechanism. Typically, the noise can be modeled with either a Gaussian, uniform or salt and pepper distribution. The restoration techniques are usually oriented toward modeling the type of degradation in order to infer the inverse process for recovering the given image. This approach usually involves the option for a criterion to numerically evaluate the quality of the resulted image and consequently the restoration process can be expressed in terms of an optimization problem.

The special filtering techniques of mean type prove particularly useful in reducing the normal/uniform noise component when the mean parameter is close to 0. In other words, the effects determined by the application of mean filters are merely the decrease of the local

variance corresponding to each processed window, and consequently to inhibit the variance component of the noise. The AMVR algorithm (Adaptive Mean Variance Removal) allows the removal of the normal/uniform noise whatever the mean of the noise is (Cocianu, State, & Vlamos, 2002). Similar to MMSE (Minimum Mean Square Error) filtering technique (Umbaugh, 1998) the application of the AMVR algorithm requires that the noise parameters and some additional features are known.

The multiresolution support set is a data structure suitable for developing noise removal algorithms. (Bacchelli & Papi, 2006; Balster et al., 2003). The multiresolution algorithms perform the restoration tasks by combining, at each resolution level, according to a certain rule, the pixels of a binary support image. Some others use a selective wavelet shrinkage algorithm for digital image denoising aiming to improve the performance. For instance Balster (Balster, Zheng & Ewing, 2003) proposes an attempt of this sort together with a computation scheme, the denoising methodology incorporated in this algorithm involving a two-threshold validation process for real time selection of wavelet coefficients.

A new solution of the denoising problem based on the description length of the noiseless data in the subspace of the basis is proposed in (Beheshti & Dahleh, 2003), where the desired description length is estimated for each subspace and the selection of the subspace corresponding to the minimum length is suggested.

In (Bacchelli & Papi, 2006), a method for removing Gaussian noise from digital images based on the combination of the wavelet packet transform and the PCA is proposed. The method leads to tailored filters by applying the Karhunen-Loeve transform in the wavelet packet domain and acts with a suitable shrinkage function on these new coefficients, allowing the noise removal without blurring the edges and other important characteristics of the images.

Wavelet thresholding methods modifying the noisy coefficients were proposed by several authors (Buades, Coll & Morel, 2005; Stark, Murtagh & Bijaoui, 1995). The attempts are based on the idea that images are represented by large wavelet coefficients that have to be preserved whereas the noise is distributed across the set of small coefficients that have to be canceled. Since the edges lead to a considerable amount of wavelet coefficients of lower values than the threshold, the cancellation of these wavelet coefficients may cause small oscillations near the edges resulting spurious wavelets in the restored image.

## 2. Mathematics behind the noise removal and image restoration algorithms

### 2.1 Principal Component Analysis (PCA) and Independent Component Analysis (ICA)

We assume that the signal is represented by a  $n$ -dimensional real-valued random vector  $X$  of 0 mean and covariance matrix  $\Sigma$ . The principal directions of the repartition of  $X$  are the directions corresponding to the maximum variability, where the variability is expressed in terms of the variance.

Definition. The vector  $\Psi_1 \in \mathbb{R}^n$  is the first principal direction if  $\|\Psi_1\| = 1$  and  $\text{var}(\Psi_1^T X) = \sup_{\substack{\Phi \in \mathbb{R}^n \\ \|\Phi\|=1}} \text{var}(\Phi^T X)$ .

The value  $\Psi_1^T X$  is referred as the first principal component of  $X$ .

Now, recursively, for any  $k$ ,  $2 \leq k \leq n$ , if we denote by  $L^\perp(\Psi_1, \dots, \Psi_{k-1})$  the linear subspace orthogonal on the linear subspace generated by the first  $(k-1)$  directions,  $\Psi_k \in \mathbb{R}^n$  is the  $k$ -th principal direction if  $\|\Psi_k\| = 1$  and  $\text{var}(\Psi_k^T X) = \sup_{\substack{\Phi \in L^\perp(\Psi_1, \dots, \Psi_{k-1}) \\ \|\Phi\|=1}} \text{var}(\Phi^T X)$ .

The value  $\Psi_k^T X$  is referred as the  $k$ -th principal component of the signal  $X$ .

Note that the principal directions  $\Psi_1, \dots, \Psi_n$  of any signal are an orthogonal basis of  $\mathbb{R}^n$ , and  $Y = \Psi^T X$  is the signal representation in terms of the principal directions, where  $\Psi = [\Psi_1, \dots, \Psi_n]$ . Obviously,  $\Psi \Psi^T = \Psi^T \Psi = I_n$ ,  $E(Y) = 0$  and  $\text{Cov}(Y, Y^T) = \Psi^T \Sigma \Psi$ . Consequently, if  $\Psi_1, \dots, \Psi_n$  are unit eigen vectors of  $\Sigma$ , then  $\text{Cov}(Y, Y^T) = \Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ , where  $\lambda_1, \dots, \lambda_n$  are the eigen values of  $\Sigma$ , that is the linear transform of matrix  $\Psi^T$  decorrelates the components of  $X$ . In the particular case of Gaussian signals,  $X \sim N(0, \Sigma)$ , the components of  $Y$  are also normal distributed,  $Y_i \sim N(0, \lambda_i)$ ,  $1 \leq i \leq n$ .

The fundamental result is given by the celebrated Karhunen-Loeve theorem:

**Theorem.** Let  $X$  be a  $n$ -dimensional real-valued random vector such that  $E(X) = 0$  and  $\text{Cov}(X, X^T) = \Sigma$ . If we denote by  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$  the eigen values of  $\Sigma$ , then, for any  $k$ ,  $1 \leq k \leq n$ , the  $k$ -th principal direction is an eigen vector of  $\Sigma$  associated to  $\lambda_k$ .

A series of approaches are based on the assumption that the signal results as a mixture of a finite number of hidden independent sources and noise. This sort of attempts are usually referred as techniques of Independent Component Analysis type. The simplest model is the linear one, given by  $X = AS + \eta$ , where  $A$  is an unknown matrix (mixing matrix),  $S$  is the  $n$ -dimensional random vector whose components are independent and  $\eta = (\eta_1, \eta_2, \dots, \eta_n)^T$  is a random vector representing the noise. The problem is to recover the hidden sources being given the signal  $X$  without knowing the mixing matrix  $A$ .

For simplicity sake, the noise model is of Gaussian type, that is  $\eta \sim N(0, \Sigma_\eta)$ . Then, if we denote  $V = AS$ , then, for any vector  $w \in \mathbb{R}^n$ ,  $w^T X = w^T V + w^T \eta$ . Consequently, the non-Gaussianity of  $w^T V$  can be maximized on the basis of  $w^T X$  if we use an expression that vanishes the component  $w^T \eta$ .

The kurtosis (the fourth-order cumulant) corresponding to a real-valued random variable  $Y$  is defined as  $\text{kurt}(Y) = E(Y^4) - 3(E(Y^2))^2$ . In case  $Y$  is normally distributed, we get  $\text{kurt}(Y) = 0$ . Since  $\eta \sim N(0, \Sigma_\eta)$ , for any  $w \in \mathbb{R}^n$ ,  $w^T \eta \sim N(0, w^T \Sigma_\eta w)$ , that is  $\text{kurt}(w^T \eta) = 0$ .

The non-Gaussianity can be also measured using the Shannon neg-entropy (mutual information). Being given the computational difficulty of evaluating the exact expression of neg-entropy, usually an approximation of it is used instead, for instance the approximation proposed in (Hyvarinen, Karhunen & Oja, 2001),  $J_G(w^T V) = [E(G(w^T V)) - E(G(v))]^2$ , where  $G$  is a non-polynomial function and  $v \sim N(0, 1)$ .

Usually the maximization of non-Gaussianity is performed on the pre-processed signal version  $\tilde{X}$ , applied in order to whiten the original clean signal. In case of the additive noise superposition model,  $X = X_0 + \eta$ , where  $X_0$  is the original clean signal (unknown) and

$\eta \sim N(0, \Sigma_\eta)$ . In case  $X_0$  and  $\eta$  are independent and  $\text{Cov}(\eta, \eta^T) = \Sigma_\eta$  is known, we get  $\text{Cov}(X_0, X_0^T) = \Sigma - \Sigma_\eta$ , where  $\text{Cov}(X, X^T) = \Sigma$  and the covariance matrix  $\Sigma$  corresponding to the observed signal  $X$  is assumed to be estimated from data. Then

$\tilde{X} = (\Sigma - \Sigma_\eta)^{-\frac{1}{2}} X = (\Sigma - \Sigma_\eta)^{-\frac{1}{2}} X_0 + (\Sigma - \Sigma_\eta)^{-\frac{1}{2}} \eta = (\Sigma - \Sigma_\eta)^{-\frac{1}{2}} X_0 + \tilde{\eta}$ , where  $(\Sigma - \Sigma_\eta)^{-\frac{1}{2}} X_0$  and  $\tilde{\eta}$  are independent and the covariance matrix of  $(\Sigma - \Sigma_\eta)^{-\frac{1}{2}} X_0$  is the unit matrix  $I_n$ . If  $X_0$  results by the linear transform of matrix  $A$  applied to the sources  $S$ ,  $X_0 = AS$ , then  $\tilde{X} = BS + \tilde{\eta}$ , where  $B = (\Sigma - \Sigma_\eta)^{-\frac{1}{2}} A$ . Consequently, the sources  $S$  are determined by maximizing the non-Gaussianity of  $\tilde{X} = BS + \tilde{\eta}$ . Usually, for simplicity sake, the matrix  $B$  is assumed to be orthogonal.

## 2.2 The use of concepts and tools of multiresolution analysis for noise removal and image restoration purposes

The multiresolution based algorithms perform the restoration tasks by combining, at each resolution level, according to a certain rule, the pixels of a binary support image. The values of the support image pixels are either 1 or 0 depending on their significance degree. At each resolution level, the contiguous areas of the support image corresponding to 1-value pixels are taken as possible objects of the image. The multiresolution support is the set of all support images and it can be computed using the statistically significant wavelet coefficients.

Let  $j$  be a certain multiresolution level. Then, for each pixel  $(x, y)$  of the input image  $I$ , the multiresolution support at the level  $j$  is  $M(I; j, x, y) = 1 \Leftrightarrow I$  contains significant information at the level  $j$  about the pixel  $(x, y)$ .

If we denote by  $\psi$  be the mother wavelet function, then the generic evaluation of the multiresolution support set results by computing the wavelet transform of the input image using  $\psi$  followed by the computation of  $M(I; j, x, y)$  on the basis of the statistically significant wavelet coefficients for each resolution level  $j$  and for each pixel  $(x, y)$ .

The computation of the wavelet transform of an one dimensional signal can be performed using the algorithm "À Trous" (Stark, Murtagh & Bijaoui, 1995). The algorithm can be extended to perform this computation in case of two-dimensional signals as, for instance, image signals

Using the resolution levels  $1, 2, \dots, p$ , where  $p$  is a selected level, the "À Trous" algorithm computes the wavelet coefficients according to the following scheme (Stark, Murtagh & Bijaoui, 1995).

Input: The sampled signal  $\{c_0(k)\}$

For  $j=0, 1, \dots, p$  do

**Step 1.**  $j=j+1$ ; compute,  $c_j(k) = \sum_l h(l) c_{j-1}(k + 2^{j-1}l)$ .

**Step 2.** Step 2. Compute  $\omega_j(k) = c_{j-1}(k) - c_j(k)$

End-for

Output: The set  $\{\omega_j(k), c_p\}_{j=1, \dots, p}$ .

Note that the computation of  $c_j(k)$  carried out in Step 1 imposes that either the periodicity condition  $c_j(k+N)=c_j(k)$  or the continuity property  $c_j(k+N)=c_j(N)$  holds.

Since the representation of the original sampled signal is  $c_0(k)=c_p(k)+\sum_{j=1}^p \omega_j(k)$ , in case of images, the values of  $c_0$  are computed for each pixel  $(x,y)$  as  $c_0(x,y)=c_p(x,y)+\sum_{j=1}^p \omega_j(x,y)$ .

If the input image  $I$  encodes a noise component  $\eta$ , then the wavelet coefficients also encode some information about  $\eta$ . A label procedure is applied to each  $\omega_j(x,y)$  in order to remove the noise component from the wavelet coefficients computed for  $I$ . In case for each pixel  $(x,y)$  of  $I$ , the distribution of the coefficients is available, the significance level corresponding to each component  $\omega_j(x,y)$  can be established using a statistical test. We say that  $I$  is local constant at the resolution level  $j$  in case the amount of noise in  $I$  at this resolution level can be neglected. Let  $H_0$  be the hypothesis  $H_0: I$  is local constant at the resolution level  $j$ . In case there is significant amount of noise in  $I$  at the resolution level  $j$ , we get that the alternative hypothesis  $\neg H_0: \omega_j(x,y) \sim N(0, \sigma_j^2)$ . In order to define the critical region  $W$  of the statistical test we proceed as follows. Let  $0 < \varepsilon < 1$  be the *a priori* selected significance level and let  $z_\varepsilon$  be such that when  $\neg H_0$  is true,

$$1 - \varepsilon = \text{Prob}(|\omega_j(x,y)| < z_\varepsilon) = \frac{1}{\sqrt{2\pi}\sigma_j} \int_{-z_\varepsilon}^{z_\varepsilon} \exp\left\{-\frac{t^2}{2\sigma_j^2}\right\} dt \quad (1)$$

In other words, the probability of rejecting  $\neg H_0$  (hence accept  $H_0$ ) when  $\neg H_0$  is true is  $\varepsilon$  and consequently, the critical region is  $W = [-z_\varepsilon, z_\varepsilon]$ . Accordingly, the significance level of the wavelet coefficients is given by the rule:  $\omega_j(x,y)$  is a significant coefficient if and only if  $\omega_j(x,y) \notin W$ .

Usually,  $z_\varepsilon$  is taken as  $k\sigma_j$ , where  $k$  is a selected constant  $k \approx 3$ , because

$$\begin{aligned} P(|\omega_j(x,y)| > k\sigma_j) &= P(\omega_j(x,y) > k\sigma_j) + P(\omega_j(x,y) < -k\sigma_j) = \\ &= 2P(\omega_j(x,y) > k\sigma_j) = 2(1 - P(\omega_j(x,y) \leq k\sigma_j)) \\ z_{k\sigma_j} < \varepsilon &\Rightarrow P(|\omega_j(x,y)| > k\sigma_j) \geq 2(1 - \varepsilon) \end{aligned}$$

Using the significance level, we set to 1 the statistically significant coefficient and respectively we set to 0 the non-significant ones. The restored image  $\tilde{I}$  is,

$$\tilde{I}(x,y) = c_p(x,y) + \sum_{j=1}^p g(\sigma_j, \omega_j(x,y)) \omega_j(x,y), \quad (2)$$

where  $g$  is defined by

$$g(\sigma_j, \omega_j(x, y)) = \begin{cases} 1, & |\omega_j(x, y)| \geq k\sigma_j \\ 0, & |\omega_j(x, y)| < k\sigma_j \end{cases}.$$

### 2.3 Information-based approaches in image restoration

The basics of the informational-based method for image restoration purposes are given by the following theoretical results (State, Cocianu & Vlamos, 2001).

**Lemma 1** Let  $X$  be a continuous  $n$ -dimensional random vector and  $A \in M_n(R)$  a non-singular matrix,  $Y = AX$ . Then,  $H(X) = H(Y) - \ln|A|$ , where

$$H(X) = - \int_{R^n} f(x) \ln f(x) dx$$

is the differential entropy (Shannon), and  $f$  is the density function of  $X$ .

**Lemma 2** Let  $X$  be a continuous  $n$ -dimensional normally distributed random vector,  $X \sim N(0, \Sigma)$  and let  $q$  be a natural number,  $1 \leq q < n$ . If  $X = \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix}$  where  $X^{(1)}$  is  $q$ -dimensional, then, for any  $x^{(2)} \in R^{n-q}$ ,  $H(X^{(1)} | X^{(2)} = x^{(2)}) = H(X^{(1)} - E(X^{(1)} | X^{(2)} = x^{(2)}))$ , where

$$E(X^{(1)} | X^{(2)} = x^{(2)})$$

is the regression function of  $X^{(1)}$  on  $X^{(2)} = x^{(2)}$ , and  $H(X^{(1)} | X^{(2)} = x^{(2)})$  is the conditional entropy of  $X^{(1)}$  on  $X^{(2)} = x^{(2)}$ .

Since  $H(X^{(1)} | X^{(2)} = x^{(2)})$  represents a measure of the amount of incertitude still remaining with respect to  $X^{(1)}$  when  $X^{(2)}$  is known, we get that the whole information carried by  $X^{(2)}$  with respect to  $X^{(1)}$  is concentrated on  $E(X^{(1)} | X^{(2)} = x^{(2)})$ .

If we denote  $\Sigma_{11} = \text{cov}(X^{(1)}, X^{(1)T})$ ,  $\Sigma_{22} = \text{cov}(X^{(2)}, X^{(2)T})$ ,  $\Sigma_{12} = \text{cov}(X^{(1)}, X^{(2)T})$ , we get  $E(X^{(1)} | X^{(2)} = x^{(2)}) = \Sigma_{12} \Sigma_{22}^{-1} x^{(2)}$  and  $Y^{(1)} \sim N(0, \Sigma_{11.2})$ ,

where  $Y^{(1)} = X^{(1)} - \Sigma_{12} \Sigma_{22}^{-1} x^{(2)} = X^{(1)} - E(X^{(1)} | X^{(2)} = x^{(2)})$ , and  $\Sigma_{11.2} = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} (\Sigma_{12})^T$ .

Consequently  $H(X^{(1)} - E(X^{(1)} | X^{(2)} = x^{(2)})) = H(X^{(1)} | X^{(2)} = x^{(2)}) = \frac{q}{2} \ln 2\pi e + \frac{1}{2} \ln |\Sigma_{11.2}|$ .

### 2.4 The image restoration method based on scatter matrices and on bounds on the probability of error

In statistical discriminant analysis, within-class, between-class and mixture scatter matrices are used to formulate criteria of class separability.

In case we need to discriminate between  $m$  classes  $H_i, i=1, m$  and  $\{X_1^{(i)}, \dots, X_N^{(i)}\}$  are examples of patterns coming respectively from these classes, the within-class scatter matrix

shows the scatter of samples around their class expected vectors and it is typically given by the expression  $S_w = \sum_{i=1}^m \xi_i \sum_{k=1}^N (X_k^{(i)} - \hat{\mu}_i)(X_k^{(i)} - \hat{\mu}_i)^T$ , where  $\hat{\mu}_i$  is the prototype of  $H_i$  and  $\xi_i$  is the *a priori* probability of  $H_i, i=1, m$ .

Very often, the *a priori* probabilities are taken  $\xi_i = \frac{1}{m}$  and each prototype is computed as the weighted mean of the patterns belonging to the respective class.

The between-class scatter matrix is the scatter of the expected vectors around the mixture mean as  $S_b = \sum_{i=1}^m \xi_i \sum_{k=1}^N (\hat{\mu}_i - \mu_0)(\hat{\mu}_i - \mu_0)^T$  where  $\mu_0$  represents the expected vector of the mixture distribution; usually  $\mu_0$  is taken as  $\mu_0 = \sum_{i=1}^m \xi_i \hat{\mu}_i$ .

The mixture scatter matrix is the covariance matrix of all samples regardless of their class assignments and it is defined by  $S_m = S_w + S_b$ . Note that all these scatter matrices are invariant under coordinate shifts.

In order to formulate criteria for class separability, these matrices should be converted into a number. This number should be larger when the between-class scatter is larger or the within-class scatter is smaller. Typical criteria are  $J_1 = \text{tr}(S_2^{-1}S_1)$ ,  $J_2 = \ln|S_2^{-1}S_1|$ , where  $(S_1, S_2) \in \{(S_b, S_w), (S_b, S_m), (S_w, S_m), (S_m, S_w)\}$  and their values can be taken as measures of overall class separability. Obviously, both criteria are invariant under linear non-singular transforms and they are currently used for feature extraction purposes [8]. When the linear feature extraction problem is solved on the base of either  $J_1$  or  $J_2$ , their values are taken as numerical indicators of the loss of information implied by the reduction of dimensionality and implicitly deteriorating class separability. Consequently, the best linear feature extraction is formulated as the optimization problem  $\arg\left(\inf_{A \in R^{n \times m}} |J_k(m, A) - J_k|\right)$  where  $m$  stands for the desired number of features,  $J_k(m, A)$  is the value of the criterion  $J_k, k=1, 2$  in the transformed  $m$ -dimensional space of  $Y = A^T X$ , where  $A$  is a  $n \times m$  matrix.

If the pattern classes are represented by the noisy image  $X^{(n)}$  and the filtered image  $F(X^{(n)})$  respectively, the value of each of the criteria  $J_k, k=1, 2$  is a measure of overall class separability as well as a measure of the amount of information discriminating between these classes. In other words,  $J_k, k=1, 2$  can be taken as measuring the effects of the noise removing filter expressing a measure of the quantity of information lost due to the use of the particular filter.

Lemma 3. For any  $m, 1 \leq m \leq n$ ,

$\arg\left(\inf_{A \in R^{n \times m}} |J_k(m, A) - J_k|\right) = \{A \Psi \mid A = (\Phi_1, \Phi_2, \dots, \Phi_m), \Psi \in R^{m \times m}, |\Psi| \neq 0\}$ , where  $\Phi_1, \dots, \Phi_m$  are unit eigenvectors corresponding to the  $m$  largest eigenvalues of  $S_2^{-1}S_1$  (Cocianu, State & Vlamos, 2004).

The probability of error is the most effective measure of a classification decision rule usefulness, but its evaluation involves integrations on complicated regions in high

dimensional spaces. When a closed-form expression for the error probability can not be obtained, we may seek either for approximate expressions, or upper/lower bounds for the error probability.

Assume that the design of the Bayes classifier is intended to discriminate between two pattern classes and the available information is represented by mean vectors  $\mu_i$ ,  $i = 1, 2$  and the covariance matrices  $\Sigma_i$ ,  $i = 1, 2$  corresponding to the repartitions of the classes respectively. The Chernoff upper bounds of the Bayesian error (Fukunaga, 1990) are given by  $\varepsilon_s = \xi_1^s \xi_2^{1-s} \int (f_1(x))^s (f_2(x))^{1-s} dx$ ,  $s \in [0, 1]$ , where  $\xi = (\xi_1, \xi_2)$  is the *a priori* distribution and  $f_i$  is the density function corresponding to the  $i$ -th class,  $i = 1, 2$ . When both density functions are normal,  $f_i \sim N(\mu_i, \Sigma_i)$   $i = 1, 2$ , the integration can be carried out to obtain a closed-form expression for  $\varepsilon_s$ , that is  $\int (f_1(x))^s (f_2(x))^{1-s} dx = \exp(-\mu(s))$  where

$$\mu(s) = \frac{s(1-s)}{2} (\mu_2 - \mu_1)^T (s\Sigma_1 + (1-s)\Sigma_2)^{-1} (\mu_2 - \mu_1) + \frac{1}{2} \ln \frac{|s\Sigma_1 + (1-s)\Sigma_2|}{|\Sigma_1|^s |\Sigma_2|^{1-s}} \quad (3)$$

The upper bound  $\mu\left(\frac{1}{2}\right) = \frac{1}{8} (\mu_2 - \mu_1)^T \left(\frac{\Sigma_1 + \Sigma_2}{2}\right)^{-1} (\mu_2 - \mu_1) + \frac{1}{2} \ln \frac{|\Sigma_1 + \Sigma_2|}{\sqrt{|\Sigma_1| |\Sigma_2|}}$  is called the Bhattacharyya distance and it is frequently used as a measure of the separability between two repartitions. Using straightforward computations, the Bhattacharyya distance can be written as,

$$\mu\left(\frac{1}{2}\right) = \frac{1}{8} \text{tr} \left\{ \bar{\Sigma}^{-1} (\mu_2 - \mu_1) (\mu_2 - \mu_1)^T \right\} + \frac{1}{4} \ln |2I_n + \Sigma_1 \Sigma_2^{-1} + \Sigma_2 \Sigma_1^{-1}| - \frac{n}{4} \ln 2 \quad (4)$$

where

$$\bar{\Sigma} = \frac{\Sigma_1 + \Sigma_2}{2}$$

Note that one of the first two terms in (4) vanishes, when  $\mu_1 = \mu_2$ ,  $\Sigma_1 = \Sigma_2$  respectively, that is the first term expresses the class separability due to the mean-difference while the second one gives the class separability due to the covariance difference.

The Bhattacharyya distance can be used as criterion function as well to express the quality of a linear feature extractor of matrix  $A \in R^{n \times m}$ .

When  $\Sigma_1 = \Sigma_2 = \Sigma$ ,  $J = \mu\left(\frac{1}{2}\right) = \frac{1}{8} \text{tr} \left\{ \bar{\Sigma}^{-1} (\mu_2 - \mu_1) (\mu_2 - \mu_1)^T \right\}$  therefore  $J$  is a particular case of the criterion  $J_1$  for  $S_2 = \Sigma$  and  $S_1 = S_b = (\mu_2 - \mu_1) (\mu_2 - \mu_1)^T$ . Consequently the whole information about the class separability is contained by an unique feature  $\Phi_1 = \frac{\Sigma^{-1} (\mu_2 - \mu_1)}{\|\Sigma^{-1} (\mu_2 - \mu_1)\|}$ . When  $\mu_1 = \mu_2$  and  $\Sigma_1 \neq \Sigma_2$ ,



$$J = \frac{1}{4} \ln |2I_n + \Sigma_2^{-1} \Sigma_1 + \Sigma_1^{-1} \Sigma_2| - \frac{n}{4} \ln 2 = \frac{1}{4} \sum_{j=1}^n \left( 2 + \lambda_j + \frac{1}{\lambda_j} \right) - \frac{n}{4} \ln 2 \quad (5)$$

where  $\lambda_j, j=1, n$  are the eigenvalues of  $\Sigma_1^{-1} \Sigma_2$ .

If the linear feature extractor is defined by the matrix  $A \in R^{nm}$ , then the value of the Bhattacharyya distance in the transformed space  $Y = A^T X$  is given by,

$$J(m, A) = \frac{1}{4} \ln \left| 2I_m + (A^T \Sigma_2 A)^{-1} (A^T \Sigma_1 A) + (A^T \Sigma_1 A)^{-1} (A^T \Sigma_2 A) \right| - \frac{m}{4} \ln 2. \quad (6)$$

The critical points of  $J(m, A)$  are the solutions of the equation  $\frac{\partial J(m, A)}{\partial A} = 0$  that is,

$$B \{ \Sigma_2 A \Sigma_2^{-1}(m) \Sigma_1(m) \Sigma_2^{-1}(m) - \Sigma_1 A \Sigma_2^{-1}(m) \} + B \{ \Sigma_1 A \Sigma_2^{-1}(m) \Sigma_2(m) \Sigma_1^{-1}(m) - \Sigma_2 A \Sigma_1^{-1}(m) \} = 0 \quad (7)$$

where

$$\Sigma_i(m) = A^T \Sigma_i A, i=1, 2 \text{ and } B = \left[ (A^T \Sigma_1 A)^{-1} (A^T \Sigma_2 A) + (A^T \Sigma_2 A)^{-1} (A^T \Sigma_1 A) + 2I_m \right]^{-1}.$$

Suboptimal solutions can be identified as the solutions of the system

$$\begin{cases} \Sigma_2 A \Sigma_2^{-1}(m) \Sigma_1(m) \Sigma_2^{-1}(m) - \Sigma_1 A \Sigma_2^{-1}(m) = 0 \\ \Sigma_1 A \Sigma_1^{-1}(m) \Sigma_2(m) \Sigma_1^{-1}(m) - \Sigma_2 A \Sigma_1^{-1}(m) = 0 \end{cases} \quad (8)$$

or equivalently,  $\Sigma_2^{-1} \Sigma_1 A = A \Sigma_2^{-1}(m) \Sigma_1(m)$ .

Obviously the criterion function  $J$  is invariant with respect to non-singular transforms and, using standard arguments, one can prove that  $\Phi^{(m)} = (\Phi_1, \dots, \Phi_m)$  can be taken as the suboptimal linear feature extractor where  $\Phi_i, i=1, m$  are unit eigenvectors corresponding to the eigenvalues  $\lambda_1, \dots, \lambda_m$  of  $\Sigma_2^{-1} \Sigma_1$  such that  $\lambda_1 + \frac{1}{\lambda_1} \geq \dots \geq \lambda_m + \frac{1}{\lambda_m} \geq \dots \geq \lambda_n + \frac{1}{\lambda_n}$ .

But, in case of image restoration problem, each of the assumptions  $\mu_1 = \mu_2, \Sigma_1 = \Sigma_2$  is unrealistic, therefore, we are forced to accept the hypothesis that  $\mu_1 \neq \mu_2$  and  $\Sigma_1 \neq \Sigma_2$ . Since there is no known procedure available to optimize the criterion  $J$  when  $\Sigma_1 \neq \Sigma_2$  and  $\mu_1 \neq \mu_2$ , a series of attempts to find suboptimal feature extractors have been proposed instead (Fukunaga, 1990)

### 3. Noise removal algorithms

#### 3.1 Minimum mean-square error filtering (MMSE), and the adaptive mean-variance removal algorithm (AMVR)

The minimum mean-square error filter (MMSE) is an adaptive filter in the sense that its basic behavior changes as the image is processed. Therefore an adaptive filter could process

differently on different segments of an image. The particular MMSE filter may act as a mean filter on some windows of the image and as a median filter on other windows of the image. The MMSE filter allows the removal of the normal/uniform additive noise and its computation is carried out as

$$\bar{X}(l,c) = Y(l,c) - \frac{\sigma^2}{\sigma_{l,c}^2} [Y(l,c) - \mu_{l,c}],$$

for  $t \leq l \leq R-t, t \leq c \leq C-t$ , where  $Y$  is a  $R \times C$  noisy image,  $W_{l,c}$  is the  $n \times n$  window centered in  $(l,c)$ , where  $n = 2t+1, t \leq l \leq R-t, t \leq c \leq C-t$ ,  $\sigma^2$  is the noise variance,  $\sigma_{l,c}^2$  is the local variance (in the window  $W_{l,c}$ ), and  $\mu_{l,c}^2$  is the local mean (average in the window  $W_{l,c}$ ).

Note that since the background region of the image is an area of fairly constant value in the original uncorrupted image, the noise variance is almost equal to the local variance, and consequently the MMSE performs as a mean filter. In image areas where the local variances are much larger than the noise variance, the filter computes a value close to the pixel value corresponding to the unfiltered image data. The magnitude of the original and local means

respectively used to modify the initial image are weighted by  $\frac{\sigma^2}{\sigma_{l,c}^2}$ , the ratio of noise to local

variance. As the value of the ratio increases, implying primarily noise in the window, the filter returns primarily the value of the local average. As this ratio decreases, implying high local detail, the filter returns more of the original unfiltered image. Consequently, the MMSE filter adapts itself to the local image properties, preserving image details while removing noise. (Umbaugh, 1998).

The special filtering techniques of mean type prove particularly useful in reducing the normal/uniform noise component when the mean parameter is close to 0. In other words, the effects determined by the application of mean filters are merely the decrease of each processed window local variance and consequently the removal of the variance component of the noise.

The AMVR algorithm allows to remove the normal/uniform noise whatever the mean of the noise is. Similar to MMSE filtering technique in application of the AMVR algorithm, the noise parameters and features are known. Basically, the AMVR algorithm works in two stages, namely the removal of the mean component of the noise (Step 1 and Step 2), and the decrease of the variance of the noise using the adaptive filter MMSE. The description of the AMVR algorithm is (Cocianu, State & Vlamos, 2002) is,

**Input** The image  $Y$  of dimensions  $R \times C$ , representing a normal/uniform disturbed version of the initial image  $X$ ,  $Y(l,c) = X(l,c) + \eta_{l,c}^0$ ,  $1 \leq l \leq R, 1 \leq c \leq C$ , where  $\eta_{l,c}^0$  is a sample of the random variable  $\eta_{l,c}$  distributed either  $N(\mu_{l,c}, \sigma_{l,c}^2)$  or  $U(\mu_{l,c}, \sigma_{l,c}^2)$ .

**Step 1.** Generate the sample of images  $\{X_1, X_2, \dots, X_n\}$ , by subtracting the noise  $\eta_{l,c}$  from the processed image  $Y$ , where  $X_i(l,c) = Y(l,c) - \eta_{l,c}^i$ ,  $1 \leq l \leq R, 1 \leq c \leq C$  and  $\eta_{l,c}^i$  is a sample of the random variable  $\eta_{l,c}$ .

**Step 2.** Compute  $\bar{X}$ , the sample mean estimate of the initial image  $X$ , by averaging the pixel values,  $\bar{X}(l,c) = \frac{1}{n} \sum_{i=1}^n X_i(l,c)$ ,  $1 \leq l \leq R, 1 \leq c \leq C$ .

**Step 3.** Compute the estimate  $\hat{X}$  of  $X$  using the adaptive filter MMSE,  $\hat{X} = MMSE(\bar{X})$ .

Output The image  $\hat{X}$ .

### 3.2 Information-based algorithms for noise removal

Let us consider the following information transmission/processing system. The signal  $X$  representing a certain image is transmitted through a channel and its noise-corrupted version  $X^{(n)}$  is received. Next, a noise-removing binomial filter is applied to the output  $X^{(n)}$  resulting  $F(X^{(n)})$ . Finally, the signal  $F(X^{(n)})$  is submitted to a restoration process producing  $\bar{X}$ , an approximation of the initial signal  $X$ . In our attempt (Cocianu, State & Vlamos, 2004) we assumed that there is no available information about the initial signal  $X$ , therefore the restoration process should be based exclusively on  $X^{(n)}$  and  $F(X^{(n)})$ . We assume that the message  $X$  is transmitted  $N$  times and we denote by  $X_1^{(2)}, \dots, X_N^{(2)}$  the resulted outputs and by  $X_1^{(1)}, \dots, X_N^{(1)}$  their corresponding filtered versions.

If we denote the given image by  $X$ , then we model  $\{X_1^{(2)}, \dots, X_N^{(2)}\}$  as a Bernoullian sample of the random  $r \times c$ -dimensional vector  $X^{(n)} = X + \eta$  where  $\eta \sim N(\mu, \Sigma)$  and  $\{X_1^{(1)}, \dots, X_N^{(1)}\}$  is a sample of the filtered random vector  $F(X^{(n)})$ . Obviously,  $X^{(n)}$  and  $F(X^{(n)})$  are normally distributed. Let us denote  $\mu^{(1)} = E(F(X^{(n)}))$ ,  $\mu^{(2)} = E(X^{(n)})$  and let  $\Sigma_{11}, \Sigma_{22}$  be their covariance matrices. We consider the working assumption that the  $2 \times r \times c$ -dimensional vector  $(X^{(n)}, F(X^{(n)}))$  is also normally distributed, therefore the conditional distribution of  $F(X^{(n)})$  on  $X^{(n)}$  is  $N(\mu^{(1)} + \Sigma_{12}\Sigma_{22}^{-1}(X^{(n)} - \mu^{(2)}), \Sigma_{11.2})$ , where

$$E(F(X^{(n)})|X^{(n)}) = \mu^{(1)} + \Sigma_{12}\Sigma_{22}^{-1}(X^{(n)} - \mu^{(2)}) \quad (9)$$

is the regression function of  $F(X^{(n)})$  on  $X^{(n)}$ ,

and  $\Sigma_{12} = \text{cov}(F(X^{(n)}), X^{(n)})$   $\Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}(\Sigma_{12})^T$  (see § 2.3).

It is well known (Anderson, 1958) that  $F(X^{(n)}) - E(F(X^{(n)})|X^{(n)})$  minimizes the variance and maximizes the correlation between  $F(X^{(n)})$  and  $X^{(n)}$  in the class of linear functions of  $X^{(n)}$ . Moreover,  $E(F(X^{(n)})|X^{(n)})$  is  $X^{(n)}$ -measurable and, since  $F(X^{(n)}) - E(F(X^{(n)})|X^{(n)})$  and  $X^{(n)}$  are independent, the whole information carried by  $X^{(n)}$  with respect to  $F(X^{(n)})$  is contained by  $E(F(X^{(n)})|X^{(n)})$ .

As a particular case, using the conclusions established by the lemmas 1 and 2 (§ 2.3), we can conclude that  $H(F(X^{(n)}) - E(F(X^{(n)})|X^{(n)})) = H(E(F(X^{(n)})|X^{(n)}))$  and  $E(F(X^{(n)})|X^{(n)})$  contains the

whole information existing in  $X^{(\eta)}$  with respect to  $F(X^{(\eta)})$  a part of it being responsible for the initial existing noise  $\eta$  and another component being responsible for the quality degradation.

According to our regression-based algorithm, the rows of the restored image  $\bar{X}$  are computed sequentially on the basis of the samples  $\{X_1^{(2)}, \dots, X_N^{(2)}\}$  and  $\{X_1^{(1)}, \dots, X_N^{(1)}\}$  representing the available information about  $X^{(\eta)}$  and  $F(X^{(\eta)})$

If we denote by  $X_k^{(p)}(i)$  the  $i$ -th row of  $X_k^{(p)}$ ,  $p = 1, 2, 1 \leq k \leq N$ , then the mean vectors  $\mu^{(p)}$  are estimated by the corresponding sample means  $\hat{\mu}^{(p)} = (\hat{\mu}^{(p)}(1), \dots, \hat{\mu}^{(p)}(r))$ , where  $\hat{\mu}^{(p)}(i) = \frac{1}{N} \sum_{k=1}^N X_k^{(p)}(i)$ ,  $1 \leq i \leq r$  and the covariance matrices  $\Sigma_{ts}(i)$ ,  $t, s = 1, 2$  are estimated respectively by their sample covariance matrices counterparts,  $\hat{\Sigma}_{ts}(i) = \frac{1}{N-1} \sum_{k=1}^N (X_k^{(t)}(i) - \hat{\mu}^{(t)}(i))(X_k^{(s)}(i) - \hat{\mu}^{(s)}(i))^T$ . Frequently enough it happens that the matrices  $\hat{\Sigma}_{tt}(i)$ ,  $t = 1, 2$  are ill conditioned, therefore in our method the Penrose pseudoinverse  $(\hat{\Sigma}_{tt}(i))^+$  is used instead of  $(\hat{\Sigma}_{tt}(i))^{-1}$ .

Since the aim is to restore as much as possible the initial image  $X$ , we have to find out ways to improve the quality of  $F(X^{(\eta)})$  in the same time preventing the introduction additional noise.

Obviously,

$$\mu^{(1)} = E(F(X + \eta)) = F(X + E(\eta)) = F(X) + F(E(\eta)) = F(X) + E(\eta) \quad (10)$$

$$\mu^{(2)} = E(X + \eta) = X + E(\eta) \quad (11)$$

hence  $\mu^{(1)} - \mu^{(2)} = F(X) - X$  and  $X^{(\eta)} - \mu^{(1)} = X - F(X) + \eta - E(\eta)$ . In other words,  $\mu^{(1)} - \mu^{(2)}$  can be viewed as measuring the effects of the noise  $\eta$  as well as the quality degradation while the term  $X^{(\eta)} - \mu^{(1)}$  retains more information about the quality of image and less information about  $\eta$  (Cocianu, State & Vlamos, 2004). This argument entails the heuristic used by our method (Step 4), the restored image being obtained by applying a threshold filter to  $\mu^{(1)}$  and adding the correction term  $\rho \Sigma_{12} \Sigma_{22}^{-1} (\mu^{(2)} - \mu^{(1)})$ ,  $\bar{X} = T(\hat{\mu}^{(1)}(i)) + \rho \Sigma_{12} \Sigma_{22}^{-1} (\mu^{(2)} - \mu^{(1)})$ .

The heuristic regression-based algorithm (HRBA) for image restoration (Cocianu, State & Vlamos, 2004)

Input: The sample  $\{X_1^{(2)}, \dots, X_N^{(2)}\}$  of noise corrupted versions of the  $r \times c$ -dimensional image  $X$

**Step 1.** Compute the sample  $\{X_1^{(1)}, \dots, X_N^{(1)}\}$  by applying the binomial filter of mask

$$M_t = \frac{1}{12+t} \begin{bmatrix} 1 & 2 & 1 \\ 2 & t & 2 \\ 1 & 2 & 1 \end{bmatrix}, \quad t \geq 4 \quad \text{to each component of } \{X_1^{(2)}, \dots, X_N^{(2)}\},$$

$$X_i^{(1)} = F(X_i^{(2)}), 1 \leq i \leq N.$$

**Step 2.** For each row  $1 \leq i \leq r$ , compute  $\hat{\mu}^{(p)}(i) = \frac{1}{N} \sum_{k=1}^N X_k^{(p)}(i), 1 \leq i \leq r, p = 1, 2,$

$$\hat{\Sigma}_{ts}(i) = \frac{1}{N-1} \sum_{k=1}^N (X_k^{(t)}(i) - \hat{\mu}^{(t)}(i)) (X_k^{(s)}(i) - \hat{\mu}^{(s)}(i))^T, t, s = 1, 2.$$

**Step 3.** For each row  $1 \leq i \leq r$ , compute  $T(\hat{\mu}^{(1)}(i))$  by applying a threshold filter to  $\hat{\mu}^{(1)}(i)$ .

**Step 4.** Compute the rows  $\bar{X}(i)$  of the restored image  $\bar{X}$ ,  $\bar{X}(i) = T(\hat{\mu}^{(1)}(i)) + \rho \hat{\Sigma}_{12}^{-1}(i) (\hat{\Sigma}_{22}(i))^{-1} (\hat{\mu}^{(2)}(i) - \hat{\mu}^{(1)}(i))$ , where  $\rho$  is a noise-preventing constant conveniently determined to prevent the restoration of the initial noise. By experimental arguments, the recommended range of  $\rho$  is  $[1.5, 5.5]$ .

Note that since the regression function can be written as,

$$E(F(X^{(\eta)})) = \mu^{(1)} + \Sigma_{12} \Sigma_{22}^{-1} (X^{(\eta)} - \mu^{(2)}) = \Sigma_{12} \Sigma_{22}^{-1} (X^{(\eta)} - \mu^{(1)}) + \mu^{(1)} - \Sigma_{12} \Sigma_{22}^{-1} (\mu^{(2)} - \mu^{(1)}) \quad (12)$$

the correction term used at Step 4 is precisely the sample mean estimation of the  $E(\Sigma_{12} \Sigma_{22}^{-1} (X^{(\eta)} - \mu^{(1)}))$ .

The idea of our attempt is to use the most informative features discriminating between  $X^{(\eta)}$  and  $F(X^{(\eta)})$  for getting correction terms in restoring the filtered images  $F(X^{(\eta)})$ . The attempt is justified by the argument that besides information about the removed noise, the most informative features discriminating between  $X^{(\eta)}$  and  $F(X^{(\eta)})$  would contain appropriate information allowing quality improvement of the image  $F(X^{(\eta)})$  (Cocianu, State & Vlamos, 2004). Let  $\{X_1^{(2)}, \dots, X_N^{(2)}\}$  be the sample of noise corrupted versions of the  $r \times c$ -dimensional image  $X$  and  $\{X_1^{(1)}, \dots, X_N^{(1)}\}$  their filtered versions,  $X_i^{(1)} = F(X_i^{(2)}), 1 \leq i \leq N$ . We assume  $\xi_1 = \xi_2 = 0.5$ , therefore the scatter matrices become  $S_w = \hat{\Sigma}_1 + \hat{\Sigma}_2$ ,  $S_b = (\hat{\mu}^{(1)} - \hat{\mu}^{(2)}) (\hat{\mu}^{(1)} - \hat{\mu}^{(2)})^T$  and  $S_m = S_w + S_b$  where

$$\hat{\mu}^{(i)} = \frac{1}{N} \sum_{k=1}^N X_k^{(i)}, \quad \hat{\Sigma}_i = \frac{1}{N-1} \sum_{k=1}^N (X_k^{(i)} - \hat{\mu}^{(i)}) (X_k^{(i)} - \hat{\mu}^{(i)})^T, \quad i = 1, 2$$

Since  $\text{rank}(S_b) = 1$ , we get  $\text{rank}(S_w^{-1} S_b) = 1$ , that is when  $S_2 = S_w$  and  $S_1 = S_b$ , the matrix  $S_2^{-1} S_1$  has an unique positive eigenvalue, one of its unit eigenvectors being given by

$$\Phi_1 = \frac{S_w^{-1}(\mu^{(1)} - \mu^{(2)})}{\|S_w^{-1}(\mu^{(1)} - \mu^{(2)})\|}.$$

The heuristic scatter matrices-based algorithms (HSBA) for image restoration (Cocianu, State & Vlamos, 2004)

The idea of our attempt is to use the most informative features discriminating between  $X^{(\eta)}$  and  $F(X^{(\eta)})$  for getting correction terms in restoring the filtered images  $F(X^{(\eta)})$ . The attempt is justified by the argument that besides information about the removed noise, the most informative features discriminating between  $X^{(\eta)}$  and  $F(X^{(\eta)})$  would contain appropriate information allowing quality improvement of the image  $F(X^{(\eta)})$  (Cocianu, State & Vlamos, 2004). Let  $\{X_1^{(2)}, \dots, X_N^{(2)}\}$  be the sample of noise corrupted versions of the  $r \times c$ -dimensional image  $X$  and  $\{X_1^{(1)}, \dots, X_N^{(1)}\}$  their filtered versions,  $X_i^{(1)} = F(X_i^{(2)})$ ,  $1 \leq i \leq N$ . We assume  $\xi_1 = \xi_2 = 0.5$ , therefore the scatter matrices are

$$S_w = \hat{\Sigma}_1 + \hat{\Sigma}_2, \quad S_b = (\hat{\mu}^{(1)} - \hat{\mu}^{(2)})(\hat{\mu}^{(1)} - \hat{\mu}^{(2)})^T \quad \text{and} \quad S_m = S_w + S_b$$

where

$$\hat{\mu}^{(i)} = \frac{1}{N} \sum_{k=1}^N X_k^{(i)}, \quad \hat{\Sigma}_i = \frac{1}{N-1} \sum_{k=1}^N (X_k^{(i)} - \hat{\mu}^{(i)})(X_k^{(i)} - \hat{\mu}^{(i)})^T, \quad i = 1, 2.$$

Since  $\text{rank}(S_b) = 1$ , we get  $\text{rank}(S_w^{-1}S_b) = 1$ , that is when  $S_2 = S_w$  and  $S_1 = S_b$ , the matrix  $S_2^{-1}S_1$  has an unique positive eigenvalue, its unit eigenvector being given by

$$\Phi_1 = \frac{S_w^{-1}(\mu^{(1)} - \mu^{(2)})}{\|S_w^{-1}(\mu^{(1)} - \mu^{(2)})\|}$$

a. The variant of the HSBA when  $S_2 = S_w$  and  $S_1 = S_b$  (Cocianu, State & Vlamos, 2004)

Input : The sample  $\{X_1^{(2)}, \dots, X_N^{(2)}\}$  of noise corrupted versions of the  $r \times c$ -dimensional image  $X$

**Step 1.** Compute the sample  $\{X_1^{(1)}, \dots, X_N^{(1)}\}$  by applying the binomial filter as in Step 1 of HRBA.

**Step 2.** For each row  $1 \leq i \leq r$ , do Step 3 until Step 7

**Step 3.** Compute  $\hat{\mu}^{(p)}(i) = \frac{1}{N} \sum_{k=1}^N X_k^{(p)}(i)$ ,  $\hat{\Sigma}_p(i) = \frac{1}{N-1} \sum_{k=1}^N (X_k^{(p)}(i) - \hat{\mu}^{(p)}(i))(X_k^{(p)}(i) - \hat{\mu}^{(p)}(i))^T$ ,  
 $p = 1, 2$

**Step 4.** Compute  $S_b(i) = (\hat{\mu}^{(1)}(i) - \hat{\mu}^{(2)}(i))(\hat{\mu}^{(1)}(i) - \hat{\mu}^{(2)}(i))^T$  and the Penrose pseudoinverse

$$S_w^+(i) \text{ of the matrix } S_w(i) = \hat{\Sigma}_1(i) + \hat{\Sigma}_2(i)$$

**Step 5.** Compute the optimal linear feature extractor  $\Phi_1(i) = \frac{S_w^{-1}(\mu^{(1)}(i) - \mu^{(2)}(i))}{\|S_w^{-1}(\mu^{(1)}(i) - \mu^{(2)}(i))\|}$

containing the information about class separability between  $\{X_1^{(2)}(i), \dots, X_N^{(2)}(i)\}$  and

$\{X_1^{(1)}(i), \dots, X_N^{(1)}(i)\}$  expressed in terms of the criterion function  $J_1$  (see §2.4)

**Step 6.** Compute  $T(\hat{\mu}^{(1)}(i))$  by applying a threshold filter to  $\hat{\mu}^{(1)}(i)$  and the correction term

$$Y(i) = \Phi_1^T \hat{\mu}^{(1)}(i)$$

**Step 7.** Compute the row  $\bar{X}(i)$  of the restored image  $\bar{X}$  by correcting the filtered image  $T(\hat{\mu}^{(1)}(i))$  using the most informative feature,  $\bar{X}(i) = T(\hat{\mu}^{(1)}(i)) + \sigma Y(i) \Phi_1$ , where  $\sigma$  is a noise-preventing constant conveniently determined to prevent the restoration of the initial noise,  $0 < \sigma < 1$ .

Note that at Step 4, the computation of  $S_w^+(i)$  is carried out instead of  $S_w^{-1}(i)$ , this modification being needed because the matrix  $S_w(i)$  could happen to be ill-conditioned.

b. The variant of the HSBA when  $S_2 = S_m$  and  $S_1 = S_w$  (Cocianu, State & Vlamos, 2004)

Let  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$  be the eigenvalue matrix of  $S_m$  and  $\Phi$  the matrix having as columns the corresponding unit eigenvectors. According to the algorithm of simultaneous diagonalization (Duda & Hart, 1973), the optimal linear feature extractor is given by

$A = \Phi \Lambda^{-\frac{1}{2}} \Psi$  where  $\Psi$  is an orthogonal matrix whose columns are unit eigenvectors of  $K = \left( \Phi \Lambda^{-\frac{1}{2}} \right)^T S_w \Phi \Lambda^{-\frac{1}{2}}$ . The most informative features about the separability of the classes

represented by the samples  $\{X_1^{(2)}, \dots, X_N^{(2)}\}$  and  $\{X_1^{(1)}, \dots, X_N^{(1)}\}$  are the entries of  $Y = A^+ \hat{\mu}^{(1)}$ , therefore the restoration can be performed by adding the correction term  $\sigma AY$  to  $T(\hat{\mu}^{(1)}(i))$  the filtered prototype of  $\{X_1^{(1)}, \dots, X_N^{(1)}\}$ .

The number of significant features is either pre-established or dynamically determined by the magnitude of the eigenvalues of  $S_2^{-1}S_1$ .

The variant of HSBA when  $S_2 = S_m$  and  $S_1 = S_w$  can be described as follows.

Input : The sample  $\{X_1^{(2)}, \dots, X_N^{(2)}\}$  of noise corrupted versions of the  $r \times c$ -dimensional image  $X$

**Step 1.** Compute the sample  $\{X_1^{(1)}, \dots, X_N^{(1)}\}$  as described in Step 1 of the variant (a) of the HSBA

**Step 2.** For each row  $1 \leq i \leq r$ , do Step 3 until Step 8

**Step 3.** Compute  $\hat{\mu}^{(p)}(i) = \frac{1}{N} \sum_{k=1}^N X_k^{(p)}(i)$ ,  $\hat{\Sigma}_p(i) = \frac{1}{N-1} \sum_{k=1}^N (X_k^{(p)}(i) - \hat{\mu}^{(p)}(i))(X_k^{(p)}(i) - \hat{\mu}^{(p)}(i))^T$ ,  
 $p = 1, 2$

**Step 4.** Compute  $S_w(i) = \hat{\Sigma}_1(i) + \hat{\Sigma}_2(i)$  and  $S_m(i) = S_w(i) + (\hat{\mu}^{(1)}(i) - \hat{\mu}^{(2)}(i))(\hat{\mu}^{(1)}(i) - \hat{\mu}^{(2)}(i))^T$

**Step 5.** Compute the eigenvalues  $(\lambda_1(i), \dots, \lambda_n(i))$  and the corresponding unit eigenvectors  $(\Phi_1(i), \dots, \Phi_n(i))$  of  $S_m(i)$ . Select the largest  $t$  eigenvalues and let  
 $\Lambda_t(i) = \text{diag}(\lambda_1(i), \dots, \lambda_t(i))$ ,  $\Phi^{(t)}(i) = (\Phi_1(i), \dots, \Phi_t(i))$

$$K(i) = \left( \Phi^{(t)}(i) \Lambda_t^{-\frac{1}{2}}(i) \right)^T S_w \left( \Phi^{(t)}(i) \Lambda_t^{-\frac{1}{2}}(i) \right).$$

**Step 6.** Compute  $\Psi(i)$  a matrix whose columns are unit eigenvectors of  $K(i)$ . The most informative feature vectors responsible for the class separability between  $\{X_1^{(2)}(i), \dots, X_N^{(2)}(i)\}$  and  $\{X_1^{(1)}(i), \dots, X_N^{(1)}(i)\}$  are the columns of

$$A(i) = \Phi^{(t)}(i) \Lambda_t^{-\frac{1}{2}}(i) \Psi(i).$$

**Step 7.** Compute  $T(\hat{\mu}^{(1)}(i))$  by applying a threshold filter to  $\hat{\mu}^{(1)}(i)$  and the correction term  
 $Y(i) = A^+(i) T(\hat{\mu}^{(1)}(i))$

**Step 8.** Compute the row  $\bar{X}(i)$  of the restored image  $\bar{X}$  by correcting the filtered image  $T(\hat{\mu}^{(1)}(i))$  using the information contained by the selected features,  $\bar{X}(i) = T(\hat{\mu}^{(1)}(i)) + \sigma A(i) Y(i)$ , where  $\sigma$  is a noise-preventing constant conveniently determined to prevent the restoration of the initial noise,  $0 < \sigma < 1$ .

c. The variant of the HSBA when  $S_2 = S_w$  and  $S_1 = S_m$  (Cocianu, State & Vlamos, 2004)

In case we take  $S_2 = S_w$  and  $S_1 = S_m$  we obtain a variant of the HSBA similar to the variant (b). In our approach, for each row  $1 \leq i \leq r$  of the processed images, the most informative features used in getting the correction term are determined using the matrix  $M(i)$  whose columns are eigenvectors of  $S_w^+ S_m$  such that  $M(i) M^T(i) = S_w^+(i)$ .

Our image restoration algorithm based on the Bhattacharyya distance can be described as follows.

The HBA for image restoration (Cocianu, State & Vlamos, 2004)

Input : The sample  $\{X_1^{(2)}, \dots, X_N^{(2)}\}$  of noise corrupted versions of the  $r \times c$ -dimensional image  $X$  and the number  $k$  of desired features.

**Step 1.** Compute the sample  $\{X_1^{(1)}, \dots, X_N^{(1)}\}$  as described in Step 1 of the variant (a) of the HSBA

**Step 2.** For each row  $1 \leq i \leq r$ , do Step 3 until Step 8



**Step 3.** Compute

$$\hat{\mu}^{(p)}(i) = \frac{1}{N} \sum_{k=1}^N X_k^{(p)}(i), \quad \hat{\Sigma}_p(i) = \frac{1}{N-1} \sum_{k=1}^N \left( X_k^{(p)}(i) - \hat{\mu}^{(p)}(i) \right) \left( X_k^{(p)}(i) - \hat{\mu}^{(p)}(i) \right)^T, \quad p = 1, 2$$

**Step 4.** Compute

$$\mu\left(\frac{1}{2}, i\right) = \frac{1}{8} \left( \hat{\mu}^{(2)}(i) - \hat{\mu}^{(1)}(i) \right)^T \left( \frac{\hat{\Sigma}_1(i) + \hat{\Sigma}_2(i)}{2} \right)^{-1} \left( \hat{\mu}^{(2)}(i) - \hat{\mu}^{(1)}(i) \right) + \frac{1}{2} \ln \frac{\left| \frac{\hat{\Sigma}_1(i) + \hat{\Sigma}_2(i)}{2} \right|}{\sqrt{\left| \hat{\Sigma}_1(i) \right| \left| \hat{\Sigma}_2(i) \right|}}$$

**Step 5.** Compute the eigenvalues  $(\lambda_1, \dots, \lambda_c)$  of the matrix  $\hat{\Sigma}_2^{-1}(i) \hat{\Sigma}_1(i)$  and a matrix  $A(i)$  whose columns  $\Phi_i$  are eigenvectors of  $\hat{\Sigma}_2^{-1}(i) \hat{\Sigma}_1(i)$  such that  $A(i) A^T(i) = \hat{\Sigma}_2^{-1}(i)$ ,  $i = 1, c$ .

**Step 6.** Arrange the eigenvalues such that for any  $1 \leq s < j \leq c$ ,

$$\frac{\left[ \Phi_s^T \left( \hat{\mu}^{(2)}(i) - \hat{\mu}^{(1)}(i) \right) \right]^2}{1 + \lambda_s} + \ln \left( 2 + \lambda_s + \frac{1}{\lambda_s} \right) \geq \frac{\left[ \Phi_j^T \left( \hat{\mu}^{(2)}(i) - \hat{\mu}^{(1)}(i) \right) \right]^2}{1 + \lambda_j} + \ln \left( 2 + \lambda_j + \frac{1}{\lambda_j} \right)$$

and select the feature matrix  $M(i) = (\Phi_1, \dots, \Phi_k)^T$ .

**Step 7.** Compute  $T(\hat{\mu}^{(1)}(i))$  by applying a threshold filter to  $\hat{\mu}^{(1)}(i)$

$$\text{and the correction term } Y(i) = M^T(i) T(\hat{\mu}^{(1)}(i))$$

**Step 8.** Compute the row  $\bar{X}(i)$  of the restored image  $\bar{X}$  by correcting the filtered image  $T(\hat{\mu}^{(1)}(i))$  using the information contained by the selected features,  $\bar{X}(i) = T(\hat{\mu}^{(1)}(i)) + \sigma M(i) Y(i)$

where  $\sigma$  is a noise-preventing constant conveniently determined to prevent the restoration of the initial noise,  $0 < \sigma < 1$ .

### 3.3 Wavelet-based denoising

The multiresolution support provides a suitable framework for noise filtering and image restoration by noise removal. Briefly, the idea is to determine a set of statistically significant wavelet coefficients from which the multiresolution support is extracted, that is the procedure is mainly based on an underlying statistical image model governing the whole process. The multiresolution support is the basis of subsequent filtering process.

We extend the MNR algorithm to the algorithm GMNR to allow the noise removal in more general cases when the noise mean can be any real number, and compare the performances of the resulted method against the most frequently used restoration algorithms (MMSE and AMVR). Briefly, the MNR algorithm is described as follows (Stark, Murtagh & Bijaoui, 1995). The parameter  $k$  used in Step 2 controls the width of the confidence interval, its value being set to a value around 3.

Input: The image  $X_0$ , the number of resolution levels  $p$  and the heuristic threshold  $k$ .

**Step 1.** Compute the sequence of image variants  $\{X_j\}_{j=1}^N$  using a discrete low-pass filter  $h$  and get the wavelet coefficients by applying the "À Trous" algorithm

$$X_j(r, c) = \sum_l \sum_k h(l, k) X_{j-1}(r + 2^{j-1}l, c + 2^{j-1}k), \quad \omega_j(r, c) = X_{j-1}(r, c) - X_j(r, c),$$

**Step 2.** Select the significant coefficients where  $\omega_j(r, c)$  is taken as being significant if  $|\omega_j(r, c)| \geq k\sigma_j$ , for  $j = 1, \dots, p$

**Step 3.** Use the filter  $g$  defined by  $g(\sigma_j, \omega_j(r, c)) = \begin{cases} 1, & |\omega_j(r, c)| \geq k\sigma_j \\ 0, & |\omega_j(r, c)| < k\sigma_j \end{cases}$  to compute the

$$\text{restored image, } \tilde{X}(r, c) = X_p(r, c) + \sum_{j=1}^p g(\sigma_j, \omega_j(r, c)) \omega_j(r, c),$$

Output The restored image  $\tilde{X}$ .

In the following, the algorithm GMNR is an extension of the MNR algorithm aiming to get the multiresolution support set in case of arbitrary noise mean, and to use this support set for noise removal purposes (Cocianu, State, Stefanescu, Vlamos, 2004). Let us denote by  $X$  the original "clean" image, and  $\eta \sim N(m, \sigma^2)$  be the additively superimposed noise, that is the image to be processed is  $Y = X + \eta$ . Using the two-dimensional filter  $\phi$ , the sampled variants of  $X$ ,  $Y$  and  $\eta$  result by convoluting them with  $\phi$  respectively,

$$c_0(x, y) = \langle Y(l, c), \phi(x-l, y-c) \rangle, \quad I_0(x, y) = \langle X(l, c), \phi(x-l, y-c) \rangle,$$

$$E_0(x, y) = \langle \eta(l, c), \phi(x-l, y-c) \rangle, \quad c_0 = I_0 + E_0.$$

The wavelet coefficients of  $c_0$  computed by the algorithm "À Trous" are,

$$\omega_j^{c_0}(x, y) = \omega_j^{I_0}(x, y) + \omega_j^{E_0}(x, y), \quad \text{where } \frac{1}{2}\psi\left(\frac{x}{2}\right) = \varphi(x) - \frac{1}{2}\varphi\left(\frac{x}{2}\right).$$

For any pixel  $(x, y)$ , we get  $c_p(x, y) = I_p(x, y) + E_p(x, y)$ , where  $p$  stands for the number of the resolution levels, and the image  $c_0$  is

$$c_0(x, y) = I_p(x, y) + E_p(x, y) + \sum_{j=1}^p \omega_j^{I_0}(x, y) + \sum_{j=1}^p \omega_j^{E_0}(x, y) \quad (13)$$

The noise mean can be inhibited by applying the following "white-wall" type technique.

**Step 1.** Get the set of images  $E^{(i)}$ ,  $1 \leq i \leq n$ , by additively superimposing noise  $N(m, \sigma^2)$  on a "white wall" image.

**Step 2.** Compute  $c_j$ ,  $E_j^{(i)}$ , and the coefficients  $\omega_j^{c_0}$ ,  $\omega_j^{E^{(i)}}$ ,  $1 \leq i \leq n$ ,  $1 \leq j \leq p$ , by applying the "À Trous" algorithm.

**Step 3.** Get the image  $\tilde{I}$  by averaging the resulted versions,

$$\tilde{I}(x, y) = \frac{1}{n} \sum_{i=1}^n \left[ c_p(x, y) - E_p^{(i)}(x, y) + \sum_{j=1}^p \left( \omega_j^{c_0}(x, y) - \omega_j^{E^{(i)}}(x, y) \right) \right].$$

**Step 4.** Compute an approximation of the original image  $I_0$  using the multiresolution filtering based on the statistically significant wavelet coefficients.

Note that  $\tilde{I}$  computed at Step 3 is  $\tilde{I} = I_0 + E'$ , where  $E' \sim N(m', \sigma'^2)$ ,  $m' \approx 0$  and  $E(\sigma'^2) \approx \sigma^2$ .

### 3.4 A combined noise removal method based on PCA and shrinkage functions

In the following, the data  $\mathbf{X}$  is a collection of image representations modeled as a sample  $\mathbf{X}_0$  coming from a multivariate wide sense stationary stochastic process of mean  $\boldsymbol{\mu}$  and covariance matrix  $\boldsymbol{\Sigma}$ , each instance being affected by additively superimposed random noise. In general, the parameters  $\boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}$  can not be assumed as been known and they are estimated from data. The most frequently used model for the noise component  $\boldsymbol{\eta}$  is also a wide sense multivariate stationary stochastic process of Gaussian type. Denoting by  $n$  the dimensionality of the image representations, the simplest noise model is the “white” model, that is  $\boldsymbol{\eta} = (\eta_t, t \in [0, \infty))$ , where for any  $t \geq 0$ ,  $\eta_t \sim N(0, \sigma^2 \mathbf{I}_n)$ . Consequently, the mathematical model for the noisy image versions is,  $\mathbf{X} = \mathbf{X}_0 + \boldsymbol{\eta}$ .

The aim is to process the data  $\mathbf{X}$  using estimates of  $\boldsymbol{\mu}$ ,  $\boldsymbol{\Sigma}$  and  $\sigma^2$  to derive accurate approximations of  $\mathbf{X}_0$ .

The data are preprocessed to get normalized and centered representations. The preprocessing step is needed to enable the hypotheses that  $0 < \sigma^2 < 1$ . If  $\mathbf{Y} = \mathbf{X}_0 - \boldsymbol{\mu} + \boldsymbol{\eta}$ , then  $\text{Cov}(\mathbf{Y}, \mathbf{Y}^T) = \boldsymbol{\Sigma} + \sigma^2 \mathbf{I}_n$ . Let  $\theta_1 \geq \theta_2 \geq \dots \geq \theta_n$  be the eigen values of  $\boldsymbol{\Sigma}$ ,  $\boldsymbol{\Phi} = (\boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2, \dots, \boldsymbol{\Phi}_n)$  an orthogonal matrix whose columns are unit eigen vectors of  $\boldsymbol{\Sigma}$ , and  $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$  the diagonal matrix whose entries are  $\lambda_i = 1 + \frac{\sigma^2}{\theta_i}$ .

We apply the linear transform of matrix  $\mathbf{A}^T = \boldsymbol{\Lambda}^{-\frac{1}{2}} \boldsymbol{\Phi}^T$  to  $\mathbf{Y}$  and get the representation  $\mathbf{Z} = \mathbf{A}^T \mathbf{Y} = \mathbf{A}^T (\mathbf{X}_0 - \boldsymbol{\mu}) + \mathbf{A}^T \boldsymbol{\eta}$ . Using the assumptions concerning the noise,  $\mathbf{A}^T \boldsymbol{\eta} \sim N(0, \sigma^2 \boldsymbol{\Lambda}^{-1})$  and consequently, the components of  $\mathbf{A}^T \boldsymbol{\eta}$  are independent, each component being of Gaussian type.

By applying the shrinkage function  $g(y) = \text{sign}(y) \max\left(0, |y| - \sqrt{2} \frac{\sigma^2}{\lambda_i}\right)$  to  $\mathbf{Z}$  (Hyvarinen, 2001), we get the estimate  $\tilde{\mathbf{Z}}_0$  of  $\mathbf{Z}_0 = \mathbf{A}^T (\mathbf{X}_0 - \boldsymbol{\mu})$ . Finally, using the equation  $\mathbf{A} \mathbf{A}^T = \boldsymbol{\Sigma}^{-1}$ , we get the estimate  $\hat{\mathbf{X}}_0 = \boldsymbol{\mu} + \boldsymbol{\Sigma} \mathbf{A} \tilde{\mathbf{Z}}_0$  of  $\mathbf{X}_0$ .

In the following, we combine the above described estimation process with a compression/decompression scheme, in order to remove the noise in a feature space of less dimensionality. For given  $m, 1 \leq m \leq n$ , we denote by  $\boldsymbol{\Phi}^m = (\boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2, \dots, \boldsymbol{\Phi}_m)$  and  $\boldsymbol{\Lambda}_m = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_m)$  the matrix having as columns the first  $m$  columns of  $\boldsymbol{\Phi}$  and the diagonal matrix having as entries the first  $m$  entries of  $\boldsymbol{\Lambda}$  respectively.

The noise removal process in the  $m$ -dimensional feature space applied to  $F = (\Lambda_m)^{-\frac{1}{2}} (\Phi^m)^T \mathbf{Y}$  produces the cleaned version  $F_0$ , and consequently the estimate of  $\mathbf{X}_0$  results by decompressing

$$F_0, \hat{\mathbf{X}}_0 = \left( (\Lambda_m)^{-\frac{1}{2}} (\Phi^m)^T \right)^+ F_0.$$

The model-free version of CSPCA is a learning from data method that computes estimates of the first and second order statistics on the basis of a series of  $n$ -dimensional noisy images  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_N, \dots$  (State, Cocianu, Sararu, Vlamos, 2009). Also, the estimates of the eigen values and eigen vectors of the sample covariance matrix are obtained using first order approximations derived in terms of perturbation theory. The first and second order statistics are computed in a classical way, that is

$$\boldsymbol{\mu}_N = \frac{1}{N} \sum_{i=1}^N \mathbf{X}_i \text{ and } \boldsymbol{\Sigma}_N = \frac{1}{N-1} \sum_{i=1}^N (\mathbf{X}_i - \boldsymbol{\mu}_N)(\mathbf{X}_i - \boldsymbol{\mu}_N)^T.$$

Using straightforward computation, the following recursive equations can be derived

$$\boldsymbol{\mu}_{N+1} = \frac{N}{N+1} \boldsymbol{\mu}_N + \frac{1}{N+1} \mathbf{X}_{N+1}, \quad \boldsymbol{\Sigma}_{N+1} = \boldsymbol{\Sigma}_N - \frac{1}{N} \boldsymbol{\Sigma}_N + \frac{1}{N+1} (\mathbf{X}_{N+1} - \boldsymbol{\mu}_N)(\mathbf{X}_{N+1} - \boldsymbol{\mu}_N)^T.$$

Denoting by  $\Delta \boldsymbol{\Sigma}_N = \boldsymbol{\Sigma}_{N+1} - \boldsymbol{\Sigma}_N$ , in case the eigen values of  $\boldsymbol{\Sigma}_N$  are pairwise distinct, using arguments of perturbation theory type, the recursive equations for the eigen values and eigen vectors can be also derived (State, Cocianu, Vlamos, Stefanescu, 2006)

$$\lambda_i^{N+1} = \lambda_i^N + (\boldsymbol{\Psi}_i^N)^T \Delta \boldsymbol{\Sigma}_N \boldsymbol{\Psi}_i^N, \quad \boldsymbol{\Psi}_i^{N+1} = \boldsymbol{\Psi}_i^N + \sum_{\substack{j=1 \\ j \neq i}}^n \frac{(\boldsymbol{\Psi}_j^N)^T \Delta \boldsymbol{\Sigma}_N \boldsymbol{\Psi}_i^N}{\lambda_i^N - \lambda_j^N} \boldsymbol{\Psi}_j^N.$$

Assume that the information is represented by  $\boldsymbol{\mu}_N, \boldsymbol{\Sigma}_N, \Lambda_N, \Phi_N$ , and a new noisy image  $\mathbf{X}_{N+1}$  is presented as input. Then the cleaned version  $\hat{\mathbf{X}}_{N+1}$  of  $\mathbf{X}_{N+1}$  is computed and supplied as output followed by the updating  $\boldsymbol{\mu}_{N+1}, \boldsymbol{\Sigma}_{N+1}, \Lambda_{N+1}, \Phi_{N+1}$ . The updated values of these parameters are fed into the restoration module and they will be used for the next test.

The restoring algorithm can be described as follows. Assuming that  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_{N_0}$  is the initial collection of noisy images, we evaluate  $\boldsymbol{\mu}_{N_0}, \boldsymbol{\Sigma}_{N_0}, \Lambda_{N_0}, \Phi_{N_0}$ . On the basis of these information, a number of  $M$  noisy images  $\mathbf{X}_{N_0+1}, \dots, \mathbf{X}_{N_0+M}$  are next processed according to the following scheme.

$k \leftarrow 1$

While ( $k \leq M$ )

Get  $\mathbf{X}_{N_0+k}$

Compute  $\boldsymbol{\mu}_{N_0+k}, \boldsymbol{\Sigma}_{N_0+k}, \Lambda_{N_0+k}, \Phi_{N_0+k}$  (M1)

Compute  $\hat{\mathbf{X}}_{N_0+k}$  (M2)

Output  $\hat{\mathbf{X}}_{N_0+k}$

$k = k + 1$

Endwhile

The computations carried out in the module M1 involve the stored parameters  $\boldsymbol{\mu}_{N_0+k-1}, \boldsymbol{\Sigma}_{N_0+k-1}, \boldsymbol{\Lambda}_{N_0+k-1}, \boldsymbol{\Phi}_{N_0+k-1}$  and the noisy current image  $\mathbf{X}_{N_0+k}$  to update the new values of the parameters  $\boldsymbol{\mu}_{N_0+k}, \boldsymbol{\Sigma}_{N_0+k}, \boldsymbol{\Lambda}_{N_0+k}, \boldsymbol{\Phi}_{N_0+k}$ . The new values of the parameters are fed into the module M2 and they are used to clean the input. The computation of the new values of the parameters is performed as,

$$\begin{aligned}\boldsymbol{\mu}_{N_0+k} &= \frac{N_0+k-1}{N_0+k} \boldsymbol{\mu}_{N_0+k-1} + \frac{1}{N_0+k} \mathbf{X}_{N_0+k} \\ \boldsymbol{\Sigma}_{N_0+k} &= \boldsymbol{\Sigma}_{N_0+k-1} - \frac{1}{N_0+k-1} \boldsymbol{\Sigma}_{N_0+k-1} + \frac{1}{N_0+k} (\mathbf{X}_{N_0+k} - \boldsymbol{\mu}_{N_0+k-1})(\mathbf{X}_{N_0+k} - \boldsymbol{\mu}_{N_0+k-1})^T \\ \lambda_i^{N_0+k} &= \lambda_i^{N_0+k-1} + (\boldsymbol{\Psi}_i^{N_0+k-1})^T \Delta \boldsymbol{\Sigma}_{N_0+k-1} \boldsymbol{\Psi}_i^{N_0+k-1} \\ \boldsymbol{\Psi}_i^{N_0+k} &= \boldsymbol{\Psi}_i^{N_0+k-1} + \sum_{\substack{j=1 \\ j \neq i}}^n \frac{(\boldsymbol{\Psi}_j^{N_0+k-1})^T \Delta \boldsymbol{\Sigma}_{N_0+k-1} \boldsymbol{\Psi}_i^{N_0+k-1}}{\lambda_i^{N_0+k-1} - \lambda_j^{N_0+k-1}} \boldsymbol{\Psi}_j^{N_0+k-1} \\ \boldsymbol{\Lambda}_{N_0+k} &= \text{diag}\{\lambda_1^{N_0+k}, \dots, \lambda_n^{N_0+k}\} \\ \boldsymbol{\Phi}_{N_0+k} &= (\boldsymbol{\Psi}_1^{N_0+k}, \dots, \boldsymbol{\Psi}_n^{N_0+k})\end{aligned}$$

The cleaned version  $\hat{\mathbf{X}}_{N_0+k}$  of each input  $\mathbf{X}_{N_0+k}$  is obtained by applying the previously described shrinkage technique, as follows.

$$\begin{aligned}\mathbf{Y}_{N_0+k} &= \mathbf{X}_{N_0+k} - \boldsymbol{\mu}_{N_0+k} \\ \mathbf{A}_{N_0+k} &= \boldsymbol{\Phi}_{N_0+k} \left( \boldsymbol{\Lambda}_{N_0+k} \right)^{-\frac{1}{2}} \\ \mathbf{Z}_{N_0+k} &= (\mathbf{A}_{N_0+k})^T \mathbf{Y}_{N_0+k}\end{aligned}$$

Compute  $\tilde{\mathbf{Z}}_{N_0+k}$  by applying the shrinkage function to  $\mathbf{Z}_{N_0+k}$ .

$$\hat{\mathbf{X}}_{N_0+k} = \boldsymbol{\mu}_{N_0+k} + \boldsymbol{\Sigma}_{N_0+k} \mathbf{A}_{N_0+k} \tilde{\mathbf{Z}}_{N_0+k}$$

Let us assume that the  $L$  gray levels of the initial image  $X$  are affected by noise of Gaussian type  $\eta \sim N(0, \Sigma)$ , and denote by  $\Phi$  an orthogonal  $n \times n$  matrix whose columns are unit eigen vectors of  $\Sigma$ , where  $n$  is the dimension of the input space. If  $\Sigma$  is known, the matrix  $\Phi$  can be computed by classical techniques, respectively in cases when  $\Sigma$  is not known, the columns of  $\Phi$  can be learned adaptively by PCA networks (Rosca, State, Cocianu, 2008).

We denote by  $Y$  the resulted image,  $Y = X + \eta$ . The images are represented by  $R \times C$  matrices, they are processed by blocks of size  $R \times C_1$ ,  $C = pC_1$ ,  $2 \leq p < C$ . In the preprocessing step, using the matrix  $\Phi$ , the noise is removed by applying the MNR algorithm to the de-correlated blocks of  $Y' = \Phi^T Y$ .

The restoration process of the image  $Y$  using the learned features is performed as follows (State, Cocianu, 2007)

- Step 1.** Compute the image  $Y'$  on the basis of the initial image by de-correlating the noise component  $Y'_{i,j} = \Phi^T Y_{i,j} = \Phi^T X_{i,j} + \eta'_{i,j}$ ,  $1 \leq i \leq R$ ,  $1 \leq j \leq C_1$ , where  $\Phi$  is a matrix of unit eigen vectors of the noise covariance matrix. Then  $\eta' = \Phi^T \eta \sim N(0, \Lambda)$  because  $\Phi^T \Sigma \Phi = \Lambda = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ , where  $\{\lambda_1, \lambda_2, \dots, \lambda_n\}$  are the eigen values of  $\Sigma$ .
- Step 2.** Remove the noise  $\eta'$  from the image  $Y'$  using its multiresolution support. The image  $Y''$  results by labeling each wavelet coefficient of each pixel.

$$Y''_{i,j} = \text{MNR}(Y'_{i,j}) \cong \Phi^T X_{i,j}, \quad \forall i = 1, \dots, R, \quad j = 1, \dots, C_1,$$

- Step 3.** Compute an approximation  $\tilde{X} \approx X$  of the initial image  $X$  by applying the linear transform of matrix  $\Phi$  to  $Y''$ ,  $\tilde{X}_{i,j} = \Phi Y''_{i,j} \cong \Phi \Phi^T X_{i,j} = X_{i,j}$ ,  $\forall i = 1, \dots, R$ ,  $j = 1, \dots, C_1$

#### 4. Conclusions and experimental comparative analysis on the performances of some noise removal and restoration algorithms

In order to evaluate the performance of the proposed noise removal algorithms, a series of experiments were performed on different 256 gray level images. We compare the performance of our algorithm NFPCA against MMSE, AMVR, and GMNR. The implementation of the GMNR algorithm used the masks

$$h_1 = \begin{pmatrix} \frac{1}{256} & \frac{1}{64} & \frac{3}{128} & \frac{1}{64} & \frac{1}{256} \\ \frac{1}{64} & \frac{1}{16} & \frac{3}{32} & \frac{1}{16} & \frac{1}{64} \\ \frac{3}{128} & \frac{3}{32} & \frac{9}{64} & \frac{3}{32} & \frac{3}{128} \\ \frac{1}{64} & \frac{1}{16} & \frac{3}{32} & \frac{1}{16} & \frac{1}{64} \\ \frac{1}{256} & \frac{1}{64} & \frac{3}{128} & \frac{1}{64} & \frac{1}{256} \end{pmatrix} \quad \text{and} \quad h_2 = \begin{pmatrix} \frac{1}{20} & \frac{1}{10} & \frac{1}{20} \\ \frac{1}{10} & \frac{2}{5} & \frac{1}{10} \\ \frac{1}{20} & \frac{1}{10} & \frac{1}{20} \end{pmatrix}.$$

Some of the conclusions experimentally derived concerning the comparative analysis of the restoration algorithms presented in the paper against some similar techniques are presented

in Table 1, and Table 2. The aims of the comparative analysis were to establish quantitative indicators to express both the quality and efficiency of each algorithm. The values of the variances in modeling the noise in images processed by the NFPCA represent the maximum of the variances per pixel resulted from the decorrelation process. We denote by  $U(a,b)$  the uniform distribution on the interval  $[a,b]$  and by  $N(\mu, \sigma^2)$  the Gaussian distribution of mean  $\mu$  and variance  $\sigma^2$ .

It seems that the AMVR algorithm proves better performances from the point of view of mean error per pixel in case of uniform distributed noise as well as in case of Gaussian type noise. Also, it seems that at least for 0-mean Gaussian distributed noise, the mask  $h_2$  provides less mean error per pixel when the restoration is performed by the MNR algorithm.

Several tests were performed to investigate the potential of the proposed CSPCA. The tests were performed on data represented by linearized monochrome images decomposed in blocks of size  $8 \times 8$ . The preprocessing step was included in order to get normalized, centered representations. Most of the tests were performed on samples of volume 20, the images of each sample sharing the same statistical properties. The proposed method proved good performance for cleaning noisy images keeping the computational complexity at a reasonable level. An example of noisy image and its cleaned version respectively are presented in Figure 1.

Restoration algorithm	Type of noise	Mean error/pixel
MMSE	U(30,80)	52.08
AMVR		10.94
MMSE	U(40,70)	50.58
AMVR		8,07
MMSE	N(40,200)	37.51
AMVR		11.54
GMNR		14.65
NFPCA		12.65
MMSE	N(50,100)	46.58
AMVR		9.39
GMNR		12.23
NFPCA		10.67

Table 1. Comparative analysis on the performance of the proposed algorithms

Restoration algorithm	Type of noise	Mean error/pixel
MNR ( $h_1$ )	N(0,100)	11.6
MNR ( $h_2$ )		9.53
MNR ( $h_1$ )	N(0,200)	14.16
MNR ( $h_2$ )		11.74

Table 2. Comparative analysis on MNR

The tests performed on new sample of images pointed out good generalization capacities and robustness of CSPCA. The computational complexity of CSPCA method is less than the complexity of the ICA code shrinkage method.

A synthesis of the comparative analysis on the quality and efficiency corresponding to the restoration algorithms presented in section 3.2 is supplied in Table 3.

So far, the tests were performed on monochrome images only. Some efforts that are still in progress aim to adapt and extend the proposed methodology to colored images. Although the extension is not straightforward and some major modifications have to be done, the already obtained results encourage the hopes that efficient variants of these algorithms can be obtained for noise removal in case of colored images too.

The tests on the proposed algorithms were performed on images of size 256x256 pixels, by splitting the images in blocks of smaller size, depending on the particular algorithm. For instance, in case of algorithms MNR and GMNR, the images are processed pixel by pixel, and the computation of the wavelet coefficients by the "A Trous" algorithm is carried out using 3x3 and 5x5 masks. The tests performed on NFPCA, CSPCA, and the model free version of CSPCA processed blocks of 8x8 pixels.

<i>Restoration algorithm</i>	<i>Mean error/pixel Noise distributed N(30,150)</i>	<i>Mean error/pixel Noise distributed N(50,200)</i>
Mean	9.422317	12.346784
HRBA	9.333114	11.747860
HSBA	9.022712	11.500245
HBA	9.370968	11.484837

Table 3. Comparative analysis on the performance of the proposed algorithms



The initial noisy image



The cleaned version of the initial image

Fig. 1. The performance of model-free version of CSPCA

The comparison of the proposed algorithm NFPCA and the currently used approaches MMSE and AMVR points out better results of NFPCA in terms of the mean error per pixel. Some of the conclusions are summarized in Table 1 and Table 2, where the noise was modeled using the uniform and normal distributions. As it is shown in Table 1, in case of the AMVR algorithm the mean error per pixel is slightly less than in case of using NFPCA, but the AMVR algorithm induces some blur effect in the image while the use of the NFPCA seems to assure reasonable small errors without inducing any annoying side effects.



The tests performed on new sample of images pointed out good generalization capacities and robustness of CSPCA. The computational complexity of CSPCA method is less than the complexity of the ICA code shrinkage method. The authors aim to extend the work from both, methodological and practical points of view. From methodological point of view, some refinements of the proposed procedures and their performances are going to be evaluated on standard large size image databases are in progress. From practical point of view, the procedures are going to be extended in solving specific GIS tasks.

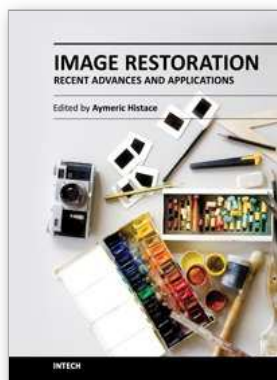
So far, the tests were performed on monochrome images only. Some efforts that are still in progress aim to adapt and extend the proposed methodology to colored images. Although the extension is not straightforward and some major modifications have to be done, the already obtained results encourage the hopes that efficient variants of these algorithms can be obtained for noise removal in case of colored images too.

The tests on the proposed algorithms were performed on images of size 256x256 pixels, by splitting the images in blocks of smaller size, depending on the particular algorithm. For instance, in case of the algorithms MNR and GMNR, the images are processed pixel by pixel, and the computation of the wavelet coefficients by the "A Trous" algorithm is carried out using 3x3 and 5x5 masks. The tests performed on NFPCA, CSPCA, and the model free version of CSPCA processed blocks of 8x8 pixels.

## 5. References

- Anderson, T.W. (1958) *An Introduction to Multivariate Statistical Analysis*, John Wiley & Sons
- Bacchelli, S., Papi S. (2006). Image denoising using principal component analysis in the wavelet domain. *Journal of Computational and Applied Mathematics, Volume 189, Issues 1-2*, 1 May 2006, 606-62
- Balster, E. J., Zheng, Y. F., Ewing, R. L. (2003). Fast, Feature-Based Wavelet Shrinkage Algorithm for Image Denoising. In: *International Conference on Integration of Knowledge Intensive Multi-Agent Systems. KIMAS '03: Modeling, Exploration, and Engineering Held in Cambridge, MA on 30 September-October 4*
- Beheshti, S. and Dahleh, M.A. , (2003 ) *A new information-theoretic approach to signal denoising and best basis selection* In: *IEEE Trans. On Signal Processing*, Volume: 53, Issue:10, Part 1, 2003: 3613- 3624
- Buades, A., Coll, B., and Morel, J.-M., (2005) *A non-local algorithm for image denoising*. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2005. Volume 2, Issue 2*, 20-25 June 2005 Page(s): 60 - 65 vol. 2
- Chatterjee, C., Roychowdhury, V.P., Chong, E.K.P. (1998). On Relative Convergence Properties of Principal Component Analysis Algorithms. In: *IEEE Transaction on Neural Networks*, vol.9, no.2
- Chellappa, R. , Jinchi, H. (1985). A nonrecursive Filter for Edge Preserving Image Restoration In: *Proc. Intl. Conf. on Acoustic, Speech and Signal Processing, Tampa 1985*
- Cocianu, C., State, L., Stefanescu, V., Vlamos, P. (2004) On the Efficiency of a Certain Class of Noise Removal Algorithms in Solving Image Processing Tasks. In: *Proceedings of the ICINCO 2004, Setubal, Portugal*
- Cocianu, C., State, L., Vlamos, P. (2002). On a Certain Class of Algorithms for Noise Removal in Image Processing: A Comparative Study, In: *Proceedings of the Third IEEE Conference on Information Technology ITCC-2002, Las Vegas, Nevada, USA, April 8-10, 2002*

- Cocianu, C., State, L., Vlamos, P. (2007). Principal Axes - Based Classification with Application in Gaussian Distributed Sample Recognition. *Economic Computation and Economic Cybernetics Studies and Research*, Vol. 41, No 1-2/2007, 159-166
- Deng, G., Tay, D., Marusic, S. (2007). *A signal denoising algorithm based on overcomplete wavelet representations and Gaussian models*. *Signal Processing Volume 87, Issue 5 (May 2007)*, 866-876
- Diamantaras, K.I., Kung, S.Y. (1996). *Principal Component Neural Networks: theory and applications*,. John Wiley & Sons
- Duda, R.O., Hart, P.E. *Pattern Classification and Scene Analysis*, Wiley & Sons, 1973
- Fukunaga K., *Introduction to Statistical Pattern Recognition*, Academic Press, Inc. 1990
- Gonzales, R., Woods, R. (2002) *Digital Image Processing*, Prentice Hall
- Haykin, S. (1999) *Neural Networks A Comprehensive Foundation*, Prentice Hall, Inc.
- Hyvarinen, A., Hoyer, P., Oja, P. (1999). *Image Denoising by Sparse Code Shrinkage*, [www.cis.hut.fi/projects/ica](http://www.cis.hut.fi/projects/ica),
- Hyvarinen, A., Karhunen, J., Oja, E. (2001) *Independent Component Analysis*, John Wiley & Sons
- J. Karhunen, E. Oja. (1982). *New Methods for Stochastic Approximations of Truncated Karhunen-Loeve Expansions*, In: *Proceedings 6<sup>th</sup> International Conference on Pattern Recognition*, Springer Verlag
- Jain, A. K., Kasturi, R., Schnuck, B. G. (1995) *Machine Vision*, McGraw Hill
- Pitas, I. (1993). *Digital Image Processing Algorithms*, Prentice Hall
- Rioul, O., Vetterli, M. (1991). *Wavelets and signal processing*. *IEEE Signal Processing Mag.* 8(4), 14-38
- Rosca, I., State, L., Cocianu, C. (2008) *Learning Schemes in Using PCA Neural Networks for Image Restoration Purposes*. *WSEAS Transactions on Information Science and Applications*, Vol. 5, July 2008, 1149-1159,
- Sanger, T.D. (1989). *An Optimality Principle for Unsupervised Learning*, *Advances in Neural Information Systems*, ed. D.S. Touretzky, Morgan Kaufmann
- Sonka, M., Hlavac, V. (1997). *Image Processing, Analyses and Machine Vision*, Chapman & Hall Computing
- Stark, J.L., Murtagh, F., Bijaoui, A. (1995). *Multiresolution Support Applied to Image Filtering and Restoration*, Technical Report
- State L, Cocianu C, Panayiotis V., *Attempts in Using Statistical Tools for Image Restoration Purposes*, The Proceedings of SCI2001, Orlando, USA, July 22-25, 2001
- State, L, Cocianu, C, Vlamos, P. (2008). *A New Unsupervised Learning Scheme to Classify Data of Relative Small Volume*. *Economic Computation and Economic Cybernetics Studies and Research*, 2008, 109-120
- State, L, Cocianu, C, Vlamos, P., Stefanescu, V. (2006) *PCA-Based Data Mining Probabilistic and Fuzzy Approaches with Applications in Pattern Recognition*. In: *Proceedings of ICSoft 2006*
- State, L., Cocianu, C., Săraru, C., Vlamos, P. (2009) *New Approaches in Image Compression and Noise Removal*, *Proceedings of the First International Conference on Advances in Satellite and Space Communications, SPACOMM 2009*, IEEE Computer Society Press, 2009
- State, L., Cocianu, C., Vlamos, P. (2007) *The Use of Features Extracted from Noisy Samples for Image Restoration Purposes*, *Informatica Economică*, Nr. 41/2007
- State, L., Cocianu, C., Vlamos, P., Stefanescu, V. (2005) *Noise Removal Algorithm Based on Code Shrinkage Technique*, *Proceedings of the 9th World Multiconference on Systemics, Cybernetics and Informatics (WMSCI 2005)*, Orlando, USA
- Umbugh, S. (1998). *Computer Vision and Image Processing*, Prentice Hall



## **Image Restoration - Recent Advances and Applications**

Edited by Dr Aymeric Histace

ISBN 978-953-51-0388-2

Hard cover, 372 pages

**Publisher** InTech

**Published online** 04, April, 2012

**Published in print edition** April, 2012

This book represents a sample of recent contributions of researchers all around the world in the field of image restoration. The book consists of 15 chapters organized in three main sections (Theory, Applications, Interdisciplinarity). Topics cover some different aspects of the theory of image restoration, but this book is also an occasion to highlight some new topics of research related to the emergence of some original imaging devices. From this arise some real challenging problems related to image reconstruction/restoration that open the way to some new fundamental scientific questions closely related with the world we interact with.

### **How to reference**

In order to correctly reference this scholarly work, feel free to copy and paste the following:

State Luminița, Cătălina-Lucia Cocianu and Vlamos Panayiotis (2012). Statistical-Based Approaches for Noise Removal, Image Restoration - Recent Advances and Applications, Dr Aymeric Histace (Ed.), ISBN: 978-953-51-0388-2, InTech, Available from: <http://www.intechopen.com/books/image-restoration-recent-advances-and-applications/statistical-based-approaches-for-noise-removal>

**INTech**  
open science | open minds

### **InTech Europe**

University Campus STeP Ri  
Slavka Krautzeka 83/A  
51000 Rijeka, Croatia  
Phone: +385 (51) 770 447  
Fax: +385 (51) 686 166  
[www.intechopen.com](http://www.intechopen.com)

### **InTech China**

Unit 405, Office Block, Hotel Equatorial Shanghai  
No.65, Yan An Road (West), Shanghai, 200040, China  
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元  
Phone: +86-21-62489820  
Fax: +86-21-62489821