Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Information Fusion 8 (2007) 131–142

INFORMATION FUSION

www.elsevier.com/locate/inffus

Pixel-based and region-based image fusion schemes using ICA bases

Nikolaos Mitianoudis *, Tania Stathaki

Communications and Signal Processing Group, Imperial College London, Exhibition Road, SW7 2AZ London, UK

Received 14 January 2005; received in revised form 1 September 2005; accepted 1 September 2005
Available online 17 October 2005

Abstract

The task of enhancing the perception of a scene by combining information captured by different sensors is usually known as *image fusion*. The *pyramid decomposition* and the *Dual-Tree Wavelet Transform* have been thoroughly applied in image fusion as analysis and synthesis tools. Using a number of *pixel-based* and *region-based* fusion rules, one can combine the important features of the input images in the transform domain to compose an enhanced image. In this paper, the authors test the efficiency of a transform constructed using *Independent Component Analysis* (ICA) and *Topographic Independent Component Analysis* bases in image fusion. The bases are obtained by offline training with images of similar context to the observed scene. The images are fused in the transform domain using novel pixel-based or region-based rules. The proposed schemes feature improved performance compared to traditional wavelet approaches with slightly increased computational complexity.

Keywords: Image fusion; Image segmentation; Independent component analysis; Topographic ICA

1. Introduction

Let $I_1(x, y), I_2(x, y), \ldots, I_T(x, y)$ represent T images of size $M_1 \times M_2$ capturing the same scene. Each image has been acquired using different instrument modalities or capture techniques, allowing each image to have different characteristics, such as degradation, thermal and visual characteristics.

In this scenario, we usually employ multiple sensors that are placed relatively close and are observing the same scene. The images acquired by these sensors, although they should be similar, are bound to have some translational motion, i.e. miscorrespondence between several points of the observed scene. *Image registration* is the process of establishing point-by-point correspondence between a number of images, describing the same scene. In this study, we will assume that the input images $I_i(x, y)$ have negligible registration problems, which implies that the objects in all images are geometrically aligned [5].

The process of combining the important features from these T images to form a single enhanced image $I_f(x, y)$ is usually referred to as *image fusion*. Fusion techniques can be divided into *spatial domain* and *transform domain* techniques [6]. In spatial domain techniques, the input images are fused in the spatial domain, i.e. using localised spatial features. Assuming that $g(\cdot)$ represents the "fusion rule", i.e. the method that combines features from the input images, the spatial domain techniques can be summarised, as follows:

$$I_f(x,y) = g(I_1(x,y), \ldots, I_T(x,y)) \tag{1}$$

The main motivation behind moving to a transform domain is to work in a framework, where the image's salient features are more clearly depicted than in the spatial domain. Hence, the choice of the transform is very important. Let $\mathscr{T}\{\cdot\}$ represent a transform operator and $g(\cdot)$ the applied fusion rule. Transform-domain fusion techniques can then be outlined, as follows:

$$I_f(x,y) = \mathscr{T}^{-1}\{g(\mathscr{T}\{I_1(x,y)\}, \ldots, \mathscr{T}\{I_T(x,y)\})\} \tag{2}$$

The fusion operator $g(\cdot)$ describes the merging of information from the different input images. Many fusion rules

* Corresponding author. Tel.: +44 207 594 6199; fax: +44 207 594 6234.
E-mail address: n.mitianoudis@imperial.ac.uk (N. Mitianoudis).

1566-2535/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.inffus.2005.09.001

have been proposed in literature [14–16]. These rules can be categorised, as follows:

- *Pixel-based rules*: the information fusion is performed in a pixel-by-pixel basis either in the transform or spatial domain. Each pixel (x, y) of the T input images is combined with various rules to form the corresponding pixel (x, y) in the "fused" image I_T. Several basic transform-domain schemes were proposed [14], such as:
  - *fusion by averaging*: fuse by averaging the corresponding coefficients in each image ("mean" rule)
  $$\mathscr{T}\{I_f(x,y)\} = \frac{1}{T}\sum_{i=1}^{T}\mathscr{T}\{I_i(x,y)\} \tag{3}$$
  - *fusion by absolute maximum*: fuse by selecting the greatest in absolute value of the corresponding coefficients in each image ("max-abs" rule)
  $$\mathscr{T}\{I_f(x,y)\} = \operatorname{sgn}(\mathscr{T}\{I_i(x,y)\})\max_i|\mathscr{T}\{I_i(x,y)\}| \tag{4}$$
  - *fusion by denoising (hard/soft thresholding)*: perform simultaneous fusion and denoising by thresholding the transform's coefficients (sparse code shrinkage [10]).
  - *high/low fusion*, i.e. combining the "high-frequency" parts of some images with the "low-frequency" parts of some other images.

  The different properties of these fusion schemes will be explained later on. For a more complete review on pixel-based methods, one can always refer to Piella [15], Nikolov et al. [14] and Rockinger et al. [16].
- *Region-based fusion rules*: these schemes group image pixels to form contiguous regions, e.g. objects and impose different fusion rules to each image region. In [13], Li et al. created a binary decision map to choose between the coefficients using a majority filter, measuring activity in small patches around each pixel. In [15], Piella proposed several activity level measures, such as the absolute value, the median or the contrast to neighbours. Consequently, she proposed a region-based scheme using a local correlation measurement to performs fusion of each region. In [12], Lewis et al. produced a joint-segmentation map out of the input images. To perform fusion, they measured *priority* using *energy*, *variance*, or *entropy* of the wavelet coefficients to impose weighting on each region in the fusion process along with other heuristic rules.

In this paper, the authors examine the application of *Independent Component Analysis* (ICA) and *Topographic Independent Component Analysis* bases as an analysis tool for image fusion in both noisy and noiseless environments. The performance of the proposed transform in image fusion is compared to traditional fusion analysis tools, such as the *wavelet transform*. Common pixel-based fusion rules are tested together with a proposed "weighted combination" scheme, based on the $\mathscr{L}_1$-norm. Finally, a region-based approach that segments and fuses active and non-active areas of the image is introduced.

The paper is structured, as follows. In Section 2, we introduce the basics of the Independent Component Analysis technique and how it can be used to generate analysis/synthesis bases for image fusion. In Section 3, we describe the general method for performing image fusion using ICA bases. In Section 4, we present the proposed pixel-based weighted combination scheme and a combinatory region-based scheme. In Section 5, we benchmark the proposed transform and fusion schemes, using common fusion testbed. Finally, in Section 6, we outline the advantages and disadvantages of the proposed schemes together with some suggestions about future work.

2. ICA and topographic ICA bases

Assume an image $I(x, y)$ of size $M_1 \times M_2$ and a window W of size $N \times N$, centered around the pixel (x_0, y_0). An "image patch" is defined as the product between a $N \times N$ neighbourhood centered around pixel (x_0, y_0) and the window W

$$I_w(k,l) = W(k,l)I(x_0 - \lfloor N/2\rfloor + k, y_0 - \lfloor N/2\rfloor + l), \quad \forall k,l \in [0, N-1] \tag{5}$$

where $\lfloor\cdot\rfloor$ represents the lower integer part and N is odd. For the subsequent analysis, we will assume a rectangular window, i.e.

$$W(k,l) = 1, \quad \forall k,l \in [0,N-1] \tag{6}$$

2.1. Definition of bases

In order to uncover the underlying structure of an image, it is common practice in image analysis to express an image as the synthesis of several other *basis* images. These bases are chosen according to the image properties we aim to highlight with this analysis. A number of bases have been proposed in literature so far, such as cosine bases, complex cosine bases, Hadamard bases and wavelet bases. In this case, the bases are well defined in order to serve some specific analysis tasks. However, one can estimate arbitrary bases by training with a population of similar content images. The bases are estimated after optimising a cost function that defines the bases' desired properties.

The $N \times N$ image patch $I_w(k, l)$ can be expressed as a linear combination of a set of K basis images $b_j(k, l)$, i.e.

$$I_w(k,l) = \sum_{j=1}^{K} u_j b_j(k,l) \tag{7}$$

where u_j are scalar constants. The two-dimensional (2D) representation can be simplified to an one-dimensional (1D) representation, by employing *lexicographic ordering*, in order to facilitate the analysis. In other words, the image patch $I_w(k, l)$ is arranged into a vector $\underline{I}_w$, taking all elements from matrix I_w in a row-wise fashion. Assume that we have a population of patches I_w, acquired randomly from the original image $I(x, y)$. These image patches can then be expressed in lexicographic ordering, as follows:

$$\underline{I}_w(t) = \sum_{j=1}^{K} u_j(t)\underline{b}_j = [\underline{b}_1 \quad \underline{b}_2 \quad \cdots \quad \underline{b}_K] \begin{bmatrix} u_1(t) \\ u_2(t) \\ \cdots \\ u_K(t) \end{bmatrix} \tag{8}$$

where t represents the tth image patch selected from the original image. The whole procedure of image patch selection and lexicographic ordering is depicted in Fig. 1. Let $B = [\underline{b}_1 \quad \underline{b}_2 \quad \cdots \quad \underline{b}_K]$ and $\underline{u}(t) = [u_1(t) \quad u_2(t) \quad \cdots \quad u_K(t)]^{\mathrm{T}}$. Then, Eq. (8) can be simplified, as follows:

$$\underline{I}_w(t) = B\underline{u}(t) \tag{9}$$

$$\underline{u}(t) = B^{-1}\underline{I}_w(t) = A\underline{I}_w(t) \tag{10}$$

In this case, $A = B^{-1} = [\underline{a}_1 \quad \underline{a}_2 \quad \cdots \quad \underline{a}_K]^{\mathrm{T}}$ represents the *analysis* kernel and B the *synthesis* kernel. This "transform" projects the observed signal $\underline{I}_w(t)$ on a set of basis vectors $\underline{b}_j$. The aim is to estimate a finite set of basis vectors that will be capable of capturing most of the signal's structure (energy). Essentially, we need N^2 bases for a *complete* representation of the N^2-dimensional signals $\underline{I}_w(t)$. However, with some energy compaction mechanism, we can have efficient *overcomplete* representations of the original signals using $K < N^2$ bases.

The estimation of these K vectors is performed using a population of training image patches $\underline{I}_w(t)$ and a criterion (cost function), which is going to be optimised in order to select the basis vectors. In the next paragraphs, we will estimate bases from image patches using several criteria.

2.1.1. Principal component analysis (PCA) bases

One of the transform's targets might be to analyse the image patches into uncorrelated components. *Principal component analysis* (PCA) can identify uncorrelated vector bases [8], assuming a linear generative model, like the one in (9). In addition, PCA can be used for dimensionality reduction to identify the K most important basis vectors. This is performed by eigenvalue decomposition of the data correlation matrix $C = \mathscr{E}\{\underline{I}_w\underline{I}_w^{\mathrm{T}}\}$. Assume that H is a matrix containing all the eigenvectors of C and D a diagonal matrix containing the eigenvalues of C. The eigenvalue at the ith diagonal element should correspond to the eigenvector at the ith column of H. Then, the rows of the following matrix V provide an orthonormal set of uncorrelated bases, which are called PCA bases

$$V = D^{-0.5}H^{\mathrm{T}} \tag{11}$$

The above set forms a *complete* set of bases, i.e. we have as many bases as the dimensionality of the problem (N^2). As PCA has good energy compaction properties, one can form a reduced (*overcomplete*) set of bases, based on the original ones. The eigenvalues can illustrate the significance of their corresponding eigenvector (basis vector). We can order the eigenvalues in the diagonal matrix D, in terms of decreasing absolute value. The eigenvector matrix H should be arranged accordingly. Then, we can select the first $K < N^2$ eigenvectors that correspond to the K most important eigenvalues and form reduced versions of $\hat{D}$ and $\hat{H}$. The reduced $K \times N^2$ PCA matrix $\hat{V}$ is calculated using (11) for $\hat{D}$ and $\hat{H}$. The input data can be mapped to the PCA domain via the transformation:

$$\underline{z}(t) = \hat{V}\underline{I}_w(t) \tag{12}$$

The number of bases K of the overcomplete set is chosen so that the computational load of a complete representation can be reduced. However, the overcomplete set should be able to provide an almost lossless representation of the original image. Therefore, the choice of K is usually a trade-off between computational complexity and image quality.

2.1.2. Independent component analysis (ICA) bases

A more strict criterion than uncorrelatedness is to assume that the basis vectors or equivalently the transform coefficients are *statistically independent*. *Independent Component Analysis* (ICA) can identify statistically independent

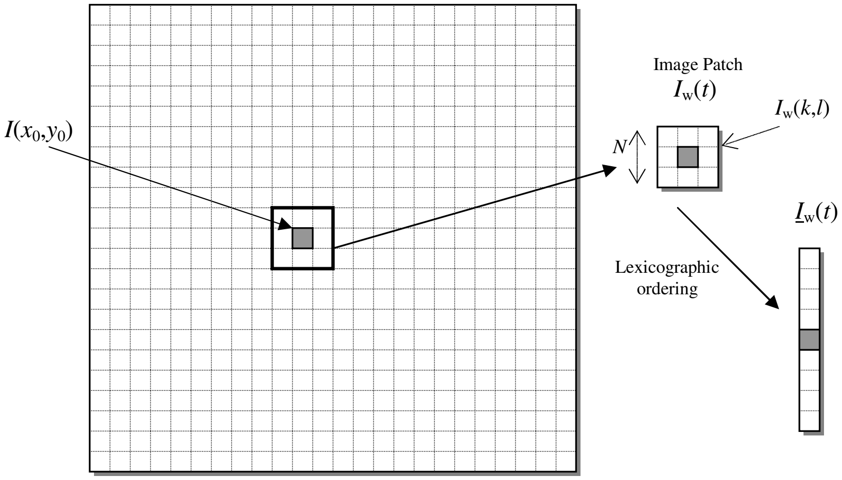


Fig. 1. Selecting an image patch I_w around pixel (x_0, y_0) and the lexicographic ordering.

basis vectors in a linear generative model [11]. A number of different approaches have been proposed to analyse the generative model in (9), assuming statistical independence between the coefficients u_i in the transform domain. Statistical independence can be closely linked with the non-Gaussianity. The *Central Limit Theorem* states that the sum of several independent random variables tends towards a Gaussian distribution. The same principal holds for any linear combination I_w of these independent random variables u_i. The Central Limit Theorem also implies that if we can find a combination of the observed signals in I_w with minimal Gaussian properties, then that signal will be one of the independent signals. Therefore, statistical independence and non-Gaussianity can be interchangeable terms.

We can briefly outline some of the different techniques that can be used to estimate independent coefficients u_i. Some approaches estimate u_i by minimising the *Kullback–Leibler* (KL) divergence between the estimated coefficients u_i and *several probabilistic priors* on the coefficients. Other approaches minimise the *mutual information* conveyed by the estimated coefficients or perform approximate diagonalisation of a *cumulant tensor* of $\underline{I}_w$. Finally, some methods estimate u_i by estimating the directions of the most non-Gaussian components using *kurtosis* or *negentropy*, as non-Gaussianity measures. For more on these techniques, one can refer to tutorial books on ICA, such as [1,11].

In this study, we will use an approach that optimises negentropy, as a non-Gaussianity measurement to identify the independent components u_i. This is also known as FastICA and was proposed by Hyvärinen and Oja [7]. In this technique, PCA is used as a preprocessing step to select the K most important vectors and orthonormalise the data using (12). Consequently, the statistical independent components can be identified using orthogonal projections $\underline{a}_i^T\underline{z}$. In order to estimate the projecting vectors $\underline{a}_i$, we have to minimise the following non-quadratic approximation of negentropy:

$$J_G(\underline{a}_i) = \left(\mathscr{E}\{G(\underline{a}_i^T\underline{z})\} - \mathscr{E}\{G(v)\}\right)^2 \tag{13}$$

where $\mathscr{E}\{\cdot\}$ denotes the expectation operator, v is a Gaussian variable of zero mean and unit variance and $G(\cdot)$ is practically any non-quadratic function. A couple of possible functions were proposed in [9]. In our analysis, we will use

$$G(x) = \alpha\sqrt{x+\epsilon} + \beta \tag{14}$$

where α, β are constants and ϵ is a small constant ($\epsilon \sim 0.1$) to tackle numerical instability, in the case that $x \to 0$. Hyvärinen and Oja produced a fixed-point method, optimising the above definition of negentropy, which is also known as the *FastICA* algorithm

$$\underline{a}_i^+ \leftarrow \mathscr{E}\{\underline{a}_i\phi(\underline{a}_i^T\underline{z})\} - \mathscr{E}\{\phi'(\underline{a}_i^T\underline{z})\}\underline{a}_i, \quad 1 \leqslant i \leqslant K \tag{15}$$

$$A \leftarrow A(A^TA)^{-0.5} \tag{16}$$

where $\phi(x) = -\partial G(x)/\partial x$. We randomly initialise the update rule in (15) for each projecting vector $\underline{a}_i$. The new updates are then orthogonalised, using the symmetric like orthogonalisation scheme in (16). These two steps are iterated, until $\underline{a}_i$ have converged.

2.1.3. Topographical independent component analysis (TopoICA) bases

In practical applications, one can very often observe clear violations of the independence assumption. It is possible to find couples of estimated components such that they are clearly dependent on each other. This dependence structure, however, is very informative and it would be useful to somehow estimate it [9].

Hyvärinen et al. [9] used the residual dependency of the "independent" components, i.e. dependencies that could not be cancelled by ICA, to define a *topographic* order between the components. Therefore, they modified the original ICA model to include a topographic order between the components, so that components that are near to each other in the topographic representation are relatively strongly dependent in the sense of higher-order correlations or mutual information. The proposed model is usually known as the *Topographic ICA* model. The topography is introduced using a neighbourhood function $h(i,k)$, which expresses the proximity between the ith and the kth component. A simple neighbourhood model can be the following:

$$h(i,k) = \begin{cases} 1, & \text{if } |i-k| \leqslant L \\ 0, & \text{otherwise} \end{cases} \tag{17}$$

where L defines the width of the neighbourhood. Consequently, the estimated coefficients u_i are no longer assumed independent, but can be modelled by some generative random variables d_k, f_i that are controlled by the neighbourhood function and shaped by a non-linearity $\phi(\cdot)$ (similar to the one in the FastICA algorithm). The topographic source model, proposed by Hyvärinen et al. [9], is the following:

$$u_i = \phi\left(\sum_{k=1}^{K} h(i,k)d_k\right)f_i \tag{18}$$

Assuming a fixed-width neighbourhood $L \times L$ and that the input data are preprocessed by PCA, Hyvärinen et al. performed Maximum Likelihood estimation of the synthesis kernel B using the linear model in (9) and the topographic source model in (18), making several assumptions for the generative random variables d_k and f_i. Optimising an approximation of the derived log-likelihood, they formed the following gradient-based topographic ICA rule:

$$\underline{a}_i^+ \leftarrow \underline{a}_i + \eta\mathscr{E}\{\underline{z}(\underline{a}_i^T\underline{z})r_i\}, \quad 1 \leqslant i \leqslant K \tag{19}$$

$$A \leftarrow A(A^TA)^{-0.5} \tag{20}$$

where η defines the learning rate of the gradient optimisation scheme and

$$r_i = \sum_{k=1}^{K} h(i,k)\phi\left(\sum_{j=1}^{K} h(j,k)(\underline{a}_i^{\mathrm{T}}\underline{z})^2\right) \tag{21}$$

As previously, we randomly initialise the update rule in (19) for each projecting vector $\underline{a}_i$. The new updates are then orthogonalised and the whole procedure is iterated, until $\underline{a}_i$ have converged. For more details on the definition and derivation of the topographic ICA model, one can always refer to the original work by Hyvärinen et al. [9].

2.2. Training ICA bases

In this paragraph, we describe the training procedure of the ICA and topographic ICA bases more thoroughly. We have to stress that the training procedure needs to be completed only once. After we have successfully trained the desired bases, the estimated transform can be used for fusion of similar content images.

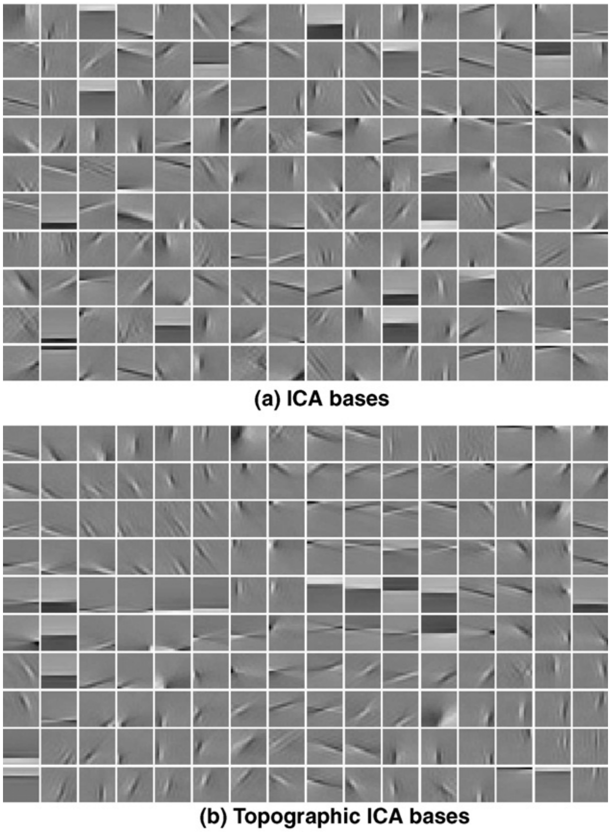
We select a set of images with similar content to the ones that will be used for image fusion. A number of $N \times N$ patches (usually ∼10000) are randomly selected from the training images. We apply lexicographic ordering to the selected images patches. We perform PCA on the selected patches and select the $K < N^2$ most important bases, according to the eigenvalues corresponding the bases. It is always possible to keep the complete set of bases. Then, we iterate the ICA update rule in (15) or the topographical ICA rule in (19) for a chosen $L \times L$ neighbourhood until convergence. Each iteration, we orthogonalise the bases using the scheme in (16).

Some examples from trained ICA and topographic ICA bases are depicted in Fig. 2. We randomly selected 10000 16×16 patches from natural landscape images. Using PCA, we selected the 160 most important bases out of the 256 bases available. In Fig. 2(a), we can see the ICA bases estimated using FastICA (15). In Fig. 2(b), we can the set of the estimated topographic ICA bases using the rule in (19) and assuming a 3×3 neighbourhood for the topographic model.

2.3. Properties of the ICA bases

Let us explore some of the properties of the ICA and the topographical ICA bases and the transforms they constitute. Both transforms are *invertible*, i.e. they guarantee perfect reconstruction. Using the symmetric orthogonalisation step $A \leftarrow A(A^{\mathrm{T}}A)^{-0.5}$, the bases remain orthogonal, i.e. the transform is *orthogonal*.

We can examine the estimated example set of ICA and topographical ICA bases in Fig. 2. The ICA and topographical ICA basis vectors seem to be closely related to wavelets and Gabor functions, as they represent similar features in different scales. However, these bases have more degrees of freedom than wavelets [9]. The Discrete Wavelet transform has only two orientations and the Dual-Tree wavelet transform can give six distinct subbands at each level with orientation ±15°, ±45°, ±75°. The ICA bases can get arbitrary orientations to fit the training patches.


(a) ICA bases

(b) Topographic ICA bases

Fig. 2. Comparison between ICA and the topographical ICA bases trained on the same set of image patches. We can observe the local correlation of the bases induced by the "topography".

One basic drawback of these transforms is that they are not *shift invariant*. This property is generally mentioned to be very important for image fusion in literature [14]. Piella [15] states that the fusion result will depend on the location or orientation of objects in the input sources in the case of misregistration problems or when used for image sequence fusion. As we assume that the observed images are all registered, the lack of shift invariance should not necessarily be a problem. In addition, Hyvärinen et al. proposed to approximate shift invariance in these ICA schemes, by employing a *sliding window* approach [10]. This implies that the input images are not divided into distinct patches, but instead every possible $N \times N$ patch in the image is analysed. This is similar to the *spin cycling* method, proposed by Coifman and Donoho [2]. This will also increase the computational complexity of the proposed framework. We have to stress that the sliding window approach is only necessary for the fusion part and not for the estimation of bases.

The basic difference between ICA and topographic ICA bases is the "topography", as introduced in the latter bases. The introduction of some local correlation in the ICA model enables the algorithm to uncover some connections between the independent components. In other words, topographic bases provide an ordered representation of the data, compared to the unordered representation of the ICA bases. In an image fusion framework, "topography" can identify groups of features that can characterise certain objects in the image. One can observe the ideas comparing Fig. 2(a) and (b). Topographic ICA seems to offer a more comprehensive representation compared to the general ICA model.

Another advantage of the ICA bases is that the estimated transform can be tailored to the needs of the application. Several image fusion applications work with specific types of images. For example, military applications work with images of airplanes, tanks, ships etc. Biomedical applications employ Computed Tomography (CT), Positron Emission Tomography (PET), ultra-sound scan images etc. Consequently, one can train bases for specific application areas. These bases should be able to analyse the trained data types more efficiently than a generic transform.

3. Image fusion using ICA bases

In this section, we describe the whole procedure of performing image fusion using ICA or topographical ICA bases, which is summarised in Fig. 3. We assume that a ICA or topographic ICA transform $\mathcal{T}\{\cdot\}$ is already estimated, as described in a previous section. Also, we assume that we have T $M_1 \times M_2$ registered sensor images $I_k(x,y)$ that need to be fused. From each image we isolate every possible $N \times N$ patch and using lexicographic ordering, we transform it to a vector $\underline{I}_k(t)$. The patches' size N should be the same as the one used in the transform estimation. Therefore, each image $I_k(x,y)$ is now represented by a population of $(M_1-N)(M_2-N)$ vectors $\underline{I}_k(t)$, $\forall t \in [1,(M_1-N)(M_2-N)]$. Each of these representations $\underline{I}_k(t)$ is transformed to the ICA or topographic ICA domain representation $\underline{u}_k(t)$. Assuming that A is the estimated analysis kernel, we have

$$\underline{u}_k(t) = \mathcal{T}\{\underline{I}_k(t)\} = A\underline{I}_k(t) \tag{22}$$

Once the image representations are in the ICA domain, one can apply a "hard" threshold on the coefficients and perform optional denoising (sparse code shrinkage), as proposed by Hyvärinen et al. [10]. Then, one can perform image fusion in the ICA or topographic ICA domain in the same manner that is performed in the wavelet or dual-tree wavelet domain. The corresponding coefficients $\underline{u}_k(t)$ from each image are combined in the ICA domain to construct a new image $\underline{u}_f(t)$. The method $g(\cdot)$ that combines the coefficients in the ICA domain is called "fusion rule"

$$\underline{u}_f(t) = g(\underline{u}_1(t), \ldots, \underline{u}_k(t), \ldots, \underline{u}_T(t)) \tag{23}$$

We can use one of the many proposed rules for fusion, as they were analysed in the introduction section and in literature [15,14]. Therefore, the "max-abs" and the "mean" rules can be two very common options. However, one can use more efficient fusion rules, as we will see in the next section. Once the composite image $\underline{u}_f(t)$ is constructed in the ICA domain, we can move back to the spatial domain, using the synthesis kernel B, and synthesise the image $I_f(x,y)$ by averaging the image patches $\underline{I}_f(t)$ in the same order they were selected during the analysis step. The whole procedure can be summarised as follows:

(1) Segment all input images $I_k(x,y)$ into every possible $N \times N$ image patch and transform them to vectors $\underline{I}_k(t)$ via lexicographic ordering.
(2) Move the input vectors to the ICA/topographic ICA domain, and get the corresponding representation $\underline{u}_k(t)$.
(3) Perform optional thresholding of $\underline{u}_k(t)$ for denoising.
(4) Fuse the corresponding coefficient using a fusion rule and form the composite representation $\underline{u}_f(t)$.
(5) Move $\underline{u}_f(t)$ to the spatial domain and reconstruct the image $I_f(x,y)$ by averaging the overlapping image patches.

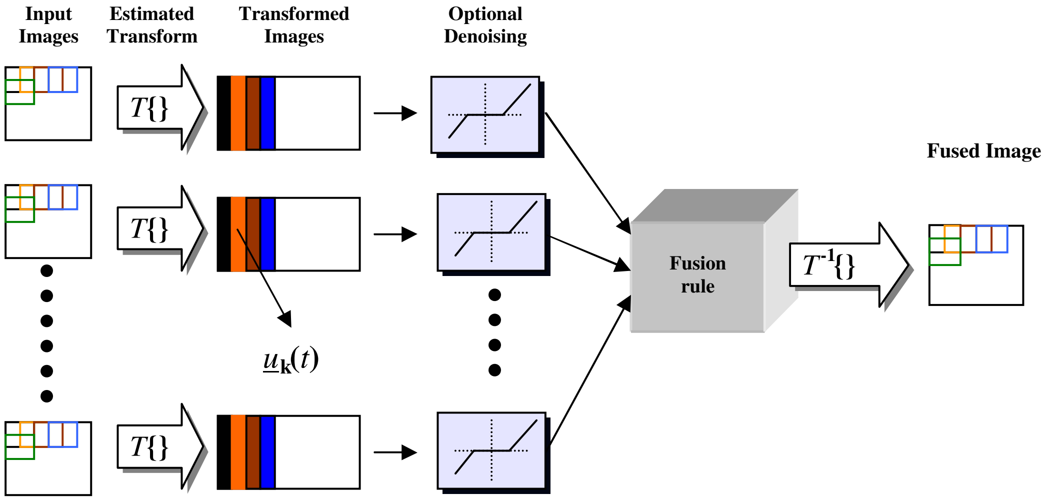


Fig. 3. The proposed fusion system using ICA/topographical ICA bases.

4. Pixel-based and region-based fusion rules using ICA bases

In this section, we describe two proposed fusion rules for ICA bases. The first one is an extension of the "max-abs" pixel-based rule, which we will refer to as the *Weight Combination* (WC) rule. The second one is a combination of the WC and the "mean" rule in a region-based scenario.

4.1. A weight combination (WC) pixel-based method

An alternative to common fusion methods, is to use a "weighted combination" of the transform coefficients, i.e.

$$\mathscr{T}\{\underline{I}_f(t)\} = \sum_{k=1}^{T} w_k(t)\mathscr{T}\{\underline{I}_k(t)\} \tag{24}$$

There are several parameters that can be employed in the estimation of the contribution $w_k(t)$ of each image to the "fused" one. In [15], Piella proposed several *activity measures*. Following the general ideas proposed in [15], we propose the following scheme. As we process each image in $N \times N$ patches, we can use the mean absolute value ($\mathscr{L}_1$-norm) of each patch (arranged in a vector) in the transform domain, as an activity indicator in each patch.

$$E_k(t) = \|\underline{u}_k(t)\|_1 \quad k = 1, \dots, T \tag{25}$$

The weights $w_k(t)$ should emphasise sources that feature more intense activity, as represented by $E_k(t)$. Consequently, the weights $w_k(t)$ for each patch t can be estimated by the contribution of the kth source image $\underline{u}_k(t)$ over the total contribution of all the T source images at patch t, in terms of activity. Hence, we can choose

$$w_k(t) = E_k(t) \Bigg/ \sum_{k=1}^{T} E_k(t) \tag{26}$$

There might be some cases, where $\sum_{k=1}^{T} E_k(t)$ is very small, denoting small energy activity in the corresponding patch. As this can cause numerical instability, we can use the "max-abs" or "mean" fusion rule for those patches.

4.2. Region-based image fusion using ICA bases

In this section, we will use the analysis of the input images in the estimated ICA domain to perform some regional segmentation and then we will fuse these regions using different rules, i.e. perform *region-based* image fusion. During, the proposed analysis methodology, we have already divided the image in small $N \times N$ patches (i.e. regions). Using the splitting/merging philosophy of region-based segmentation [17], we can find a criterion to merge the pixels corresponding to each patch in order to form contiguous areas of interest.

One could use the energy activity measurement, as introduced by (25), to infer the existence of edges in the corresponding frame. As the ICA bases tend to focus on the edge information, it is clear that great values for $E_k(t)$, correspond to great activity in the frame, i.e. the existence of edges. In contrast, small values for $E_k(t)$ denote the existence of almost constant background in the frame. Using this idea, we can segment the image in two regions: (i) "active" regions containing details and (ii) "non-active" regions containing background information. The threshold that will be used to characterise a region as "active" or "non-active" can be set heuristically to $2\text{mean}_t\{E_k(t)\}$. Since we are not interested in creating the most accurate edge-detector, we can allow some tolerance around the real edges of the image. As a result, we form the following segmentation map $m_k(t)$ from each input image:

$$m_k(t) = \begin{cases} 1, & \text{if } E_k(t) > 2\text{mean}_t\{E_k(t)\} \\ 0, & \text{otherwise} \end{cases} \tag{27}$$

The segmentation map of each input image is combined to form a single segmentation map, using the logical OR operator. As mentioned earlier, we are not interested in forming a very accurate edge detection map, but instead we have to ensure that our segmentation map contains all the edge information

$$m(t) = \text{OR}\{m_1(t), m_2(t), \dots, m_T(t)\} \tag{28}$$

Now that we have segmented the image into "active" and "non-active" regions, we can fuse these regions using different pixel-based fusion schemes. For the "active" region, we can use a fusion scheme that preserves the edges, i.e. the "max-abs" scheme or the weighted combination scheme and for the "non-active" region, we can use a scheme that preserves the background information, i.e. the "mean" or "median" scheme. Consequently, this could form a more accurate fusion scheme, that pays attention to the structure of the image itself, rather than fuse information generically.

5. Experiments

In this section, we test the performance of the proposed image fusion schemes based on ICA bases. It is not our intention to provide an exhaustive comparison of the many different transforms and fusion schemes that exist in literature. Instead, a comparison with fusion schemes using *wavelet packets* analysis and the *Dual-Tree* (*Complex*) *Wavelet Transform* are performed. In these examples we will test the "fusion by absolute maximum" (max-abs), the "fusion by averaging" (mean), the Weighted Combination (weighted) and the Region-based (Regional) fusion, where applicable.

We present three experiments, using both artificial and real image data sets. In the first experiment, we have the *Ground Truth* image $I_{gt}(x, y)$, which enable us to perform numerical evaluation of the fusion schemes. We assume that the input images $I_i(x, y)$ are processed by the fusion schemes to create the "fused" image $I_f(x, y)$. To evaluate the scheme's performance, we can use the following *Signal-to-Noise Ratio* (SNR) expression to compare the ground truth image with the fused image

$$\mathrm{SNR}_{(\mathrm{dB})} = 10\log_{10}\frac{\sum_x\sum_y I_{\mathrm{gt}}(x,y)^2}{\sum_x\sum_y (I_{\mathrm{gt}}(x,y) - I_f(x,y))^2} \tag{29}$$

As traditionally employed by the fusion community, we can also use the *Image Quality Index* Q_0, as a performance measure [19]. Assume that m_I represents the mean of the image $I(x,y)$ and all images are of size $M_1 \times M_2$. As $-1 \leqslant Q_0 \leqslant 1$, the value of Q_0 that is closer to 1, indicates better fusion performance

$$Q_0 = \frac{4\sigma_{I_{\mathrm{gt}}I_f} m_{I_{\mathrm{gt}}} m_{I_f}}{(m_{I_{\mathrm{gt}}}^2 + m_{I_f}^2)(\sigma_{I_{\mathrm{gt}}}^2 + \sigma_{I_f}^2)} \tag{30}$$

where

$$\sigma_I^2 = \frac{1}{M_1M_2 - 1}\sum_{x=1}^{M_1}\sum_{y=1}^{M_2}(I(x,y) - m_I)^2 \tag{31}$$

$$\sigma_{IJ} = \frac{1}{M_1M_2 - 1}\sum_{x=1}^{M_1}\sum_{y=1}^{M_2}(I(x,y) - m_I)(J(x,y) - m_J) \tag{32}$$

We trained the ICA and the topographic ICA bases using 10 000 8×8 image patches selected randomly from 10 images of similar content to the ground truth or the observed scene. We used 40 out of the 64 possible bases to perform the transformation in either case. We compared the performance of the ICA and topographic ICA transforms (topoICA) with a Wavelet Packet decomposition[1] and the Dual-Tree Wavelet Transform.[2] For the Wavelet Packet decomposition (WP), we used Symmlet-7 (Sym7) bases, with 5 level-decomposition using Coifman–Wickerhauser entropy. For the Dual-Tree Wavelet Transform (DTWT), we used 4 levels of decomposition and the filters included in the package. In the next pages, we will present some of the resulting fusion images. However, the visual differences between the fused images may not be very clear in the printed version of the paper, due to limitation in space. Consequently, the reader is prompted to acquire the whole set either by download[3] or via email to us.

5.1. Experiment 1: Artificially distorted images

In the first experiment, we have created three images of an "airplane" using different localised artificial distortions. The introduced distortions can model several different types of degradation that may occur in visual sensor imaging, such as motion blur, out-of-focus blur and finally pixelate or shape distortion, due to low bit-rate transmission or channel errors. This synthetic example can be a good starting point for evaluation, as there are no registration errors between the input images and we can perform numerical evaluation, as we have the ground truth image. We applied all possible combinations of transforms and the fusion rules (the "Weighted" and "Regional" fusion rules cannot be applied in the described form for the WP and DTWT transforms). Some results are depicted in Fig. 5, whereas the full numerical evaluation is presented in Table 1.

We can see that using the ICA and the TopoICA bases, we can get better fusion results both in visual quality and metric quality (PSNR, Q_0). We observe the ICA bases provide an improvement of ~1.5–2 dB, compared to the wavelet transforms, using the "max-abs" rule. The topoICA bases seem to score slightly better than the normal ICA bases, mainly due to better adaptation to local features. In terms of the various fusion schemes, the "max-abs" rule seems to give very low performance in this example using visual sensors. This can be explained, due to the fact that this scheme seems to highlight the important features of the images, however, it tends to lose some constant background information. On the other hand, the "mean" rule gives the best performance (especially for the wavelet coefficient), as it seems to balance the high detail with the low-detail information. However, the "fused" image in this case seems quite "blurry", as the fusion rule has oversmoothed the image details. Therefore, the high SNR has to be cross-checked with the actual visual quality and image perception, where we can clearly that the salient features have been filtered. The "weighted combination" rule seems to balance the pros and cons of the two previous approaches, as the results feature high PSNR and Q_0 (inferior to the "mean" rule), but the "fused" images seem sharper with correct constant background information. In Fig. 4, we can see the segmentation map created by (27) and (28). The proposed region-based scheme manages to capture most of the salient areas of the input images. It performs reasonably well as an edge detector, however, it produces thicker edges, as the objective is to identify areas around the edges, not the edges themselves. The region-based fusion scheme produces similar results to the "Weighted" fusion scheme. However, it seems to produce better visual quality in constant background areas, as the "mean" rule is more suitable for the "non-active"regions (Fig. 5).

5.2. Experiment 2: The "Toys" dataset

In the second experiment, we use the "Toys" example, which is a real visual sensor example provided by Lehigh Image Fusion group [4]. In this example, we have three registered images with different focus points, observing the same scene of toys (Fig. 6). In the first image, we have focused on left part, in the second on the center part and in the third image on the right part of the image. The ground truth image is not available, which is very common in many multi-focus examples. Therefore, SNR-type measurements are not available in this case.

Here, we can see that the ICA and TopoICA bases perform slightly better than wavelet-based approaches. Also, we can see that the "max-abs" rule performs slightly better than any other approach, with almost similar performance

[1] We used WaveLab v8.02, as available at http://www-stat.stanford.edu/~wavelab/.

[2] Code available online by the Polytechnic University of Brooklyn, NY at http://taco.poly.edu/WaveletSoftware/.

[3] http://www.commsp.ee.ic.ac.uk/~nikolao/ElsevierImages.zip.

Table 1
Performance comparison of several combinations of transforms and fusion rules in terms of PSNR (dB)/Q_0 using the "airplane" example

	WP (Sym7)	DT-WT	ICA	TopoICA
Max-abs	13.66/0.8247	13.44/0.8178	14.48/0.8609	14.80/0.8739
Mean	22.79/0.9853	22.79/0.9853	17.41/0.9565	17.70/0.9580
Weighted	–	–	17.56/0.9531	17.70/0.9547
Regional	–	–	17.56/0.9533	17.69/0.9549

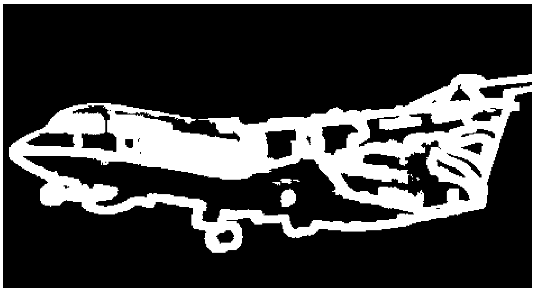
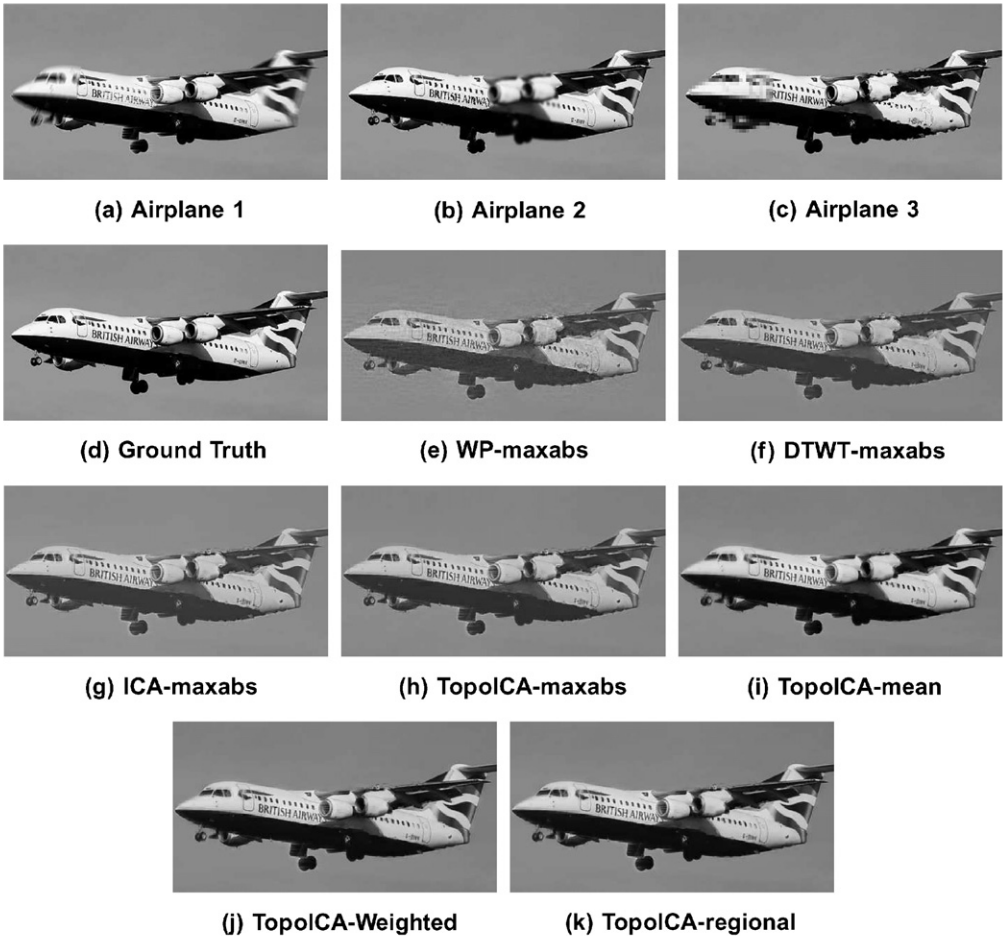


Fig. 4. Region mask created for the region-based image fusion scheme. The white areas represent "active" segments and the black areas "non-active" segments.

from the "Weighted" scheme. The reason might be that the three images have the same colour information, however, most parts of each image are blurred. Therefore, the "max-abs" that identifies the greatest activity seems more suitable for a multi-focus example.

5.3. Experiment 3: Multi-modal image fusion

In the third example, we explore the performance in multi-modal image fusion. In this case, the input images are acquired from different modality sensors to unveil different components in the observed scene. We have used some surveillance images from TNO Human Factors, provided by Toet [18]. More of these can be found in the Image Fusion Server [3]. The images are acquired by three kayaks approaching the viewing location from far away (Fig. 7). As a result, their corresponding image size



(a) Airplane 1
(b) Airplane 2
(c) Airplane 3
(d) Ground Truth
(e) WP-maxabs
(f) DTWT-maxabs
(g) ICA-maxabs
(h) TopoICA-maxabs
(i) TopoICA-mean
(j) TopoICA-Weighted
(k) TopoICA-regional

Fig. 5. Three artificially-distorted input images and various fusion results using various transforms and fusion rules.

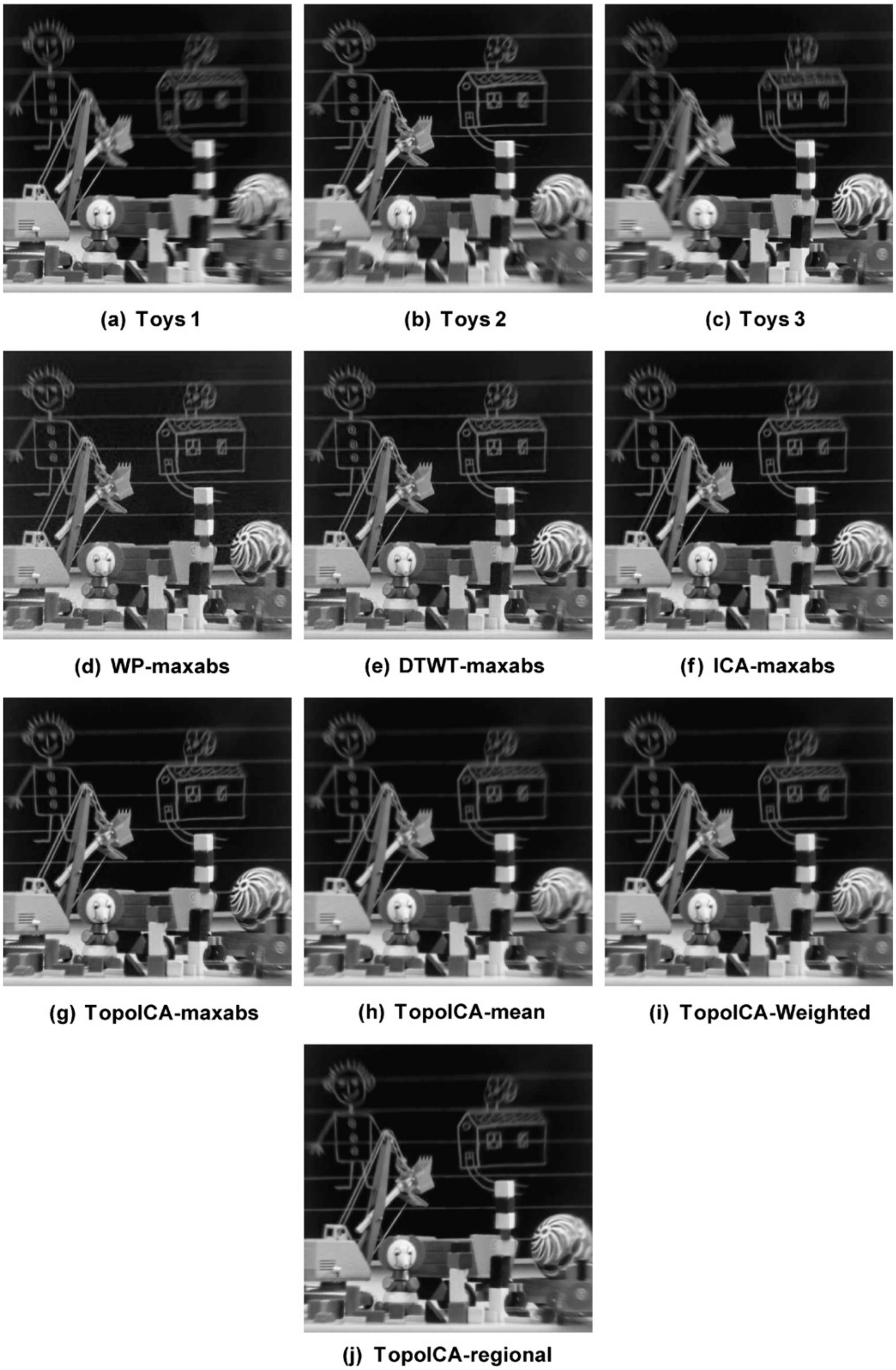
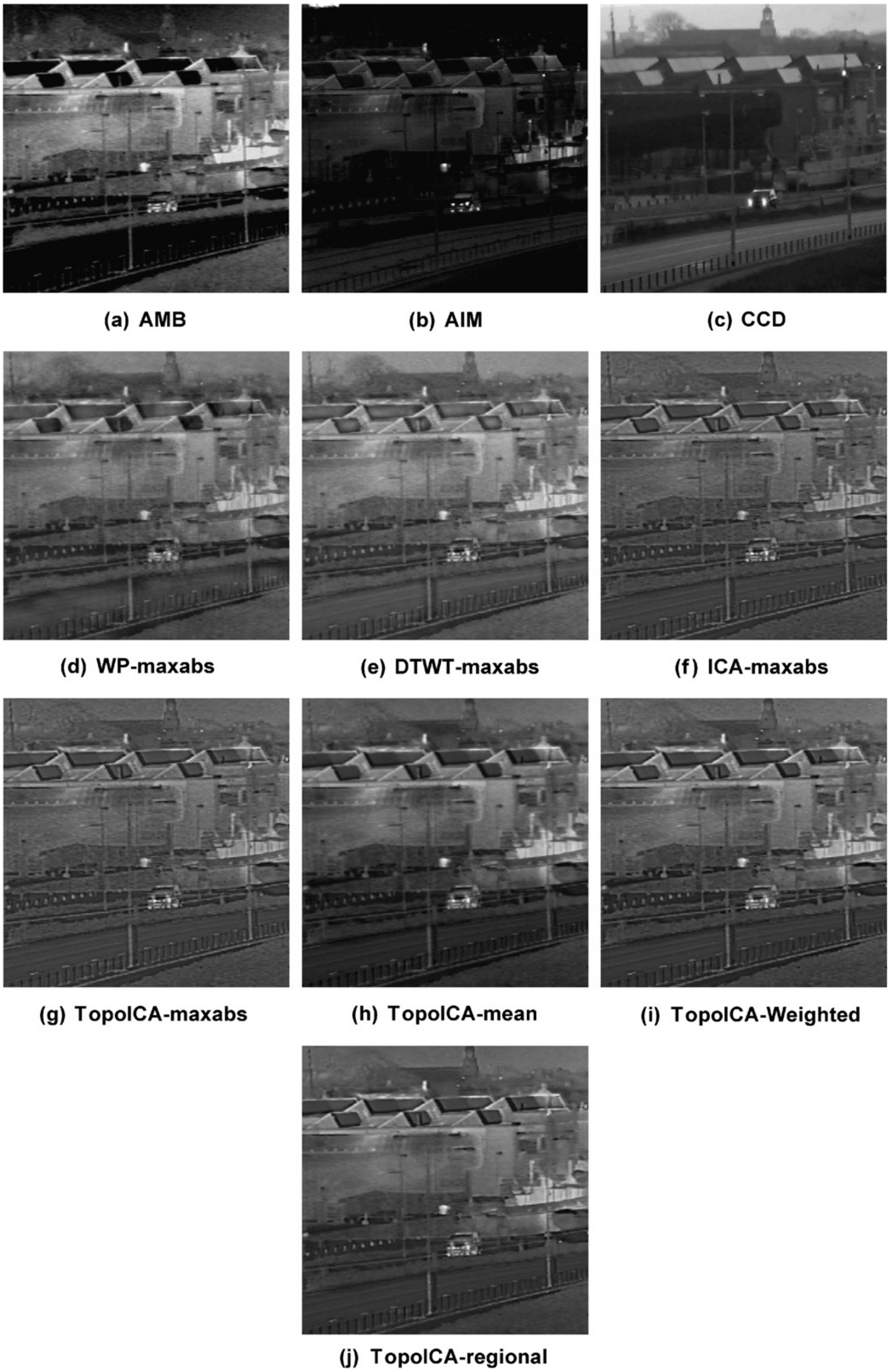


Fig. 6. The "Toys" data-set demonstrating several out-of-focus examples and various fusion results with various transforms and fusion rules.

varies from less than 1 pixel to almost the entire field of view, i.e. they are minimal registration errors. The first sensor (AMB) is a Radiance HS IR camera (Raytheon), the second (AIM) is an AIM 256 microLW camera and the third is a Philips LTC500 CCD camera. Consequently, we get three different modality inputs for the same observed scene. However, the concept of ground truth is not really meaningful in this case and therefore, we cannot have any numerical performance evaluation for this example.



(a) AMB
(b) AIM
(c) CCD
(d) WP-maxabs
(e) DTWT-maxabs
(f) ICA-maxabs
(g) TopoICA-maxabs
(h) TopoICA-mean
(i) TopoICA-Weighted
(j) TopoICA-regional

Fig. 7. Multi-modal image fusion: three images acquired through different modality sensors and various fusion results with various transforms and fusion rules.

In this example, we can witness some effects of misregistration in the fused image. We can see that all four transforms seem to have included most salient information from the input sensor images, especially in the "max-abs" and "weighted" schemes. However, it seems that the fused image created using the ICA and the TopoICA bases looks sharper and less blurry.

6. Conclusion

In this paper, the authors have introduced the use of ICA and topographical ICA bases for image fusion applications. These bases seem to construct very efficient tools, which can compliment common techniques used in image fusion, such as the Dual-Tree Wavelet Transform. The proposed method can outperform wavelet approaches. The topographical ICA bases offer more accurate directional selectivity, thus capturing the salient features of the image more accurately. A weighted combination image fusion rule seemed to improve the fusion quality over traditional fusion rules in several cases. In addition, a region-based approach was introduced. At first, segmentation into "active" and "non-active" areas is performed. The "active" areas are fused using the pixel-based weighted combination rule and the "non-active" areas are fused using the pixel-based "mean" rule.

The proposed schemes seem to increase the computational complexity of the image fusion framework. The extra computational cost is not necessarily introduced by the estimation of the ICA bases, as this task is performed only once. The bases can be trained offline using selected image samples and then employed constantly by the fusion applications. The increase in complexity comes from the "sliding window" technique that is introduced to achieve shift invariance. Implementing this fusion scheme in a more computationally efficient framework than MATLAB will decrease the time needed for the image analysis and synthesis part of the algorithm.

For future work, the authors would be looking at evolving to a more autonomous fusion system. The fusion system should be able to select the essential coefficients automatically, by optimizing several criteria, such as activity measures and region information. In addition, the authors would like to explore the nature of "topography", as introduced by Hyvärinen et al., and form more efficient activity detectors, based on topographic information.

Acknowledgements

This work is supported by the Data Information Fusion project of the Defence Technology Centre, UK. The authors would like to thank the three anonymous reviewers for their insightful comments and suggestions.

References

[1] A. Cichocki, S.I. Amari, Adaptive Blind Signal and Image Processing. Learning Algorithms and Applications, John Wiley & Sons, 2002.

[2] R.R. Coifman, D.L. Donoho, Translation-invariant de-noising, Technical report, Department of Statistics, Stanford University, Stanford, California, 1995.

[3] The Image fusion server. Available from: <http://www.imagefusion.org/>.

[4] Lehigh fusion test examples. Available from: <http://www.ece.lehigh.edu/spcrl/>.

[5] A. Goshtasby, 2-D and 3-D Image Registration: For Medical, Remote Sensing, and Industrial Applications, John Wiley & Sons, 2005.

[6] P. Hill, N. Canagarajah, D. Bull, Image fusion using complex wavelets, in: Proceedings of the 13th British Machine Vision Conference, Cardiff, UK, 2002.

[7] A. Hyvärinen, Fast and robust fixed-point algorithms for independent component analysis, IEEE Transactions on Neural Networks 10 (3) (1999) 626–634.

[8] A. Hyvärinen, Survey on independent component analysis, Neural Computing Surveys 2 (1999) 94–128.

[9] A. Hyvärinen, P.O. Hoyer, M. Inki, Topographic independent component analysis, Neural Computing 13 (2001).

[10] A. Hyvärinen, P.O. Hoyer, E. Oja, Image denoising by sparse code shrinkage, in: S. Haykin, B. Kosko (Eds.), Intelligent Signal Processing, IEEE Press, 2001.

[11] A. Hyvärinen, J. Karhunen, E. Oja, Independent Component Analysis, John Wiley & Sons, 2001.

[12] J.J. Lewis, R.J. O'Callaghan, S.G. Nikolov, D.R. Bull, C.N. Canagarajah, Region-based image fusion using complex wavelets, in: Proceedings of the 7th International Conference on Information Fusion, Stockholm, Sweden, 2004, pp. 555–562.

[13] H. Li, S. Manjunath, S. Mitra, Multisensor image fusion using the wavelet transform, Graphical Models and Image Processing 57 (3) (1995) 235–245.

[14] S.G. Nikolov, D.R. Bull, C.N. Canagarajah, M. Halliwell, P.N.T. Wells, Image fusion using a 3-d wavelet transform, in: Proceedings of the 7th International Conference on Image Processing and its Applications, 1999, pp. 235–239.

[15] G. Piella, A general framework for multiresolution image fusion: from pixels to regions, Information Fusion 4 (2003) 259–280.

[16] O. Rockinger, T. Fechner, Pixel-level image fusion: the case of image sequences, SPIE Proceedings 3374 (1998) 378–388.

[17] M. Sonka, V. Hlavac, R. Boyle, Image Processing, Analysis and Machine Vision, second ed., Brooks/Cole Publishing Company, 1999.

[18] A. Toet, Targets and Backgrounds: Characterization and Representation VIII, The International Society for Optical Engineering, 2002, pp. 118–129.

[19] Z. Wang, A.C. Bovik, A universal image quality index, IEEE Signal Processing Letters 9 (3) (2002) 81–84.