

ADAPTIVE IMAGE FUSION USING ICA BASES

Nikolaos Mitianoudis and Tania Stathaki

Imperial College London, Exhibition Road, SW7 2AZ London, UK

ABSTRACT

Image fusion can be viewed as a process that incorporates essential information from different modality sensors into a composite image. The use of bases trained using Independent Component Analysis (ICA) for image fusion has been highlighted recently. Common fusion rules can be used in the ICA fusion framework with promising results. In this paper, the authors propose an adaptive fusion scheme, based on the ICA fusion framework, that maximises the sparsity of the fusion image in the transform domain.

1. INTRODUCTION

Let $I_1(x, y), I_2(x, y), \dots, I_T(x, y)$ 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. In this study, we will assume that the objects in all input images are geometrically aligned. The process of combining important features from these T images to form a single enhanced image $I_f(x, y)$ is referred to as *image fusion*.

The problem of image fusion has been addressed widely in literature for military, biomedical or geographical imaging applications. Most approaches perform fusion in a transform domain that can highlight the salient features of the image. Pyramid decomposition and the Dual-Tree Wavelet Transform (DT-WT) are two common choices [1, 2], as they both model abrupt signal changes efficiently, i.e. edge information in an image. Many pixel-based or region-based fusion rules were proposed using these two image analysis frameworks.

In [3], we proposed to use bases trained using Independent Component Analysis (ICA) on similar-content images as analysis tools for image fusion. The main motivation was to use bases that can fit arbitrarily on the object types we are going to fuse. This framework can outperform generic analysis tools, such as wavelet analysis. One can use the same pixel-based and region-based fusion rules, that were employed before by the fusion community, in the ICA framework. In this paper, the authors propose a mechanism for a self-adaptive, unsupervised fusion scheme, that preserves and emphasises the local features in most fusion scenarios. This is achieved by optimising the sparseness of the fused image in the trans-

form domain. Two different probabilistic priors that model sparsity are employed with promising results.

2. IMAGE FUSION IN THE ICA DOMAIN

2.1. Image analysis and training using ICA bases

Assume an image $I(x, y)$ of size $M_1 \times M_2$. An “image patch” is defined as a $N \times N$ neighbourhood centered around the pixel (x_0, y_0) . Assume that we have a population of patches I_w , acquired randomly from the original image $I(x, y)$. Each image patch $I_w(k, l)$ is arranged into a vector \underline{I}_w , using lexicographic ordering. These vectors can be expressed as a linear combination of the bases vectors \underline{b}_j :

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

where t represents the t -th image patch selected from the original image. Equation (1) can be expressed, as follows:

$$\underline{I}_w(t) = B \underline{u}(t) \quad (2)$$

$$\underline{u}(t) = B^{-1} \underline{I}_w(t) = A \underline{I}_w(t) \quad (3)$$

where $B = [\underline{b}_1 \ \underline{b}_2 \ \dots \ \underline{b}_K]$ and $\underline{u}(t) = [u_1(t) \ u_2(t) \ \dots \ u_K(t)]^T$. In this case, $A = B^{-1} = [\underline{a}_1 \ \underline{a}_2 \ \dots \ \underline{a}_K]^T$ represents the *analysis* kernel and B the *synthesis* kernel. The estimation of these basis vectors is performed using a population of training image patches $\underline{I}_w(t)$ and a criterion (cost function) that selects the basis vectors. We can train analysis bases using *Independent Component Analysis* (ICA) and topographic ICA [4], as explained in [3]. The training procedure needs to be completed only once, as the estimated transform can be used for fusing similar content images.

A number of $N \times N$ patches (usually ~ 10000) are randomly selected from similar-content training images. We perform PCA on the selected patches and select the $K < N^2$ most important bases. It is always possible to keep the complete set of bases. Then, we iterate the ICA update rule in [5] or the topographical ICA rule in [4] for a chosen $L \times L$ neighbourhood until convergence. Each iteration, we orthogonalise the bases using a symmetric decorrelation scheme [5].

This work is supported by the Data Information Fusion project of the Defence Technology Centre, UK.

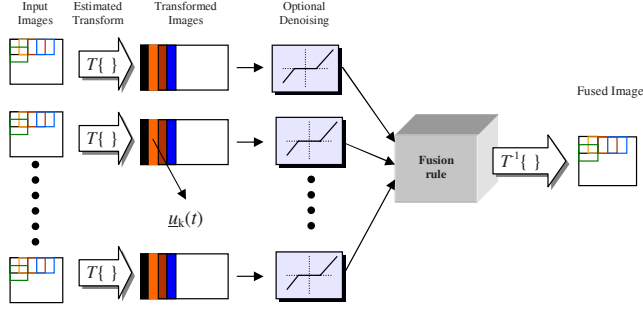


Fig. 1. The proposed fusion system using ICA / Topographical ICA bases.

2.2. Fusion in the ICA domain

After estimating a ICA or Topographic ICA transform $\mathcal{T}\{\cdot\}$, we can perform image fusion using ICA or Topographical ICA bases (see figure 1). Every possible $N \times N$ patch is isolated from each image $I_k(x, y)$ and is consequently rearranged to form a vector $\underline{I}_k(t)$. Each of the input vectors $\underline{I}_k(t)$ are transformed to the ICA or Topographic ICA domain representation $\underline{u}_k(t)$, using equation (3)

Optional denoising is also possible, by applying a “hard” threshold on the coefficients in the ICA domain [6]. The corresponding coefficients $\underline{u}_k(t)$ from each image are then combined to construct a new image $\underline{u}_f(t)$. The next step is to move back to the spatial domain, using the synthesis kernel B , and synthesise the image $I_f(x, y)$ by averaging the image patches $I_f(t)$ in the same order they were selected during the analysis step.

2.3. Various Pixel-based rules using ICA bases

In this section, we describe some generic rules that can be used for image fusion. The *fusion by absolute maximum* rule has been used widely by the image fusion community. It simply selects the greatest in absolute value of the corresponding coefficients in each image (“max-abs” rule). This rule seems to convey all the information about the edges to the fused image, however, the intensity information in constant background areas seems to be slightly distorted. In contrast, the *fusion by averaging* rule averages the corresponding coefficients (“mean” rule). This seems to preserve the correct contrast information, however, the edge details seem to get oversmoothed.

We can form a *Weighted Combination* (WC) pixel-based rule using the ICA framework [3]. We wish to use a “weighted combination” of the transform coefficients, i.e.

$$\underline{u}_f(t) = \sum_{k=1}^T w_k(t) \underline{u}_k(t) \quad (4)$$

To estimate the contributions $w_k(t)$ of each image to the “fused”

image, we can use the mean absolute value (\mathcal{L}_1 -norm) of each patch (arranged in a vector) in the transform domain, as an activity indicator:

$$E_k(t) = \|\underline{u}_k(t)\|_1 \quad k = 1, \dots, T \quad (5)$$

The weights $w_k(t)$ should emphasise sources with 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 k -th source image $\underline{u}_k(t)$ over the total contribution of all the T source images at patch t , in terms of activity.

$$w_k(t) = E_k(t) / \sum_{k=1}^T E_k(t) \quad (6)$$

In some patches, where $\sum_{k=1}^T E_k(t)$ might be very small, one can use the “max-abs” or “mean” fusion rule to avoid numerical instability.

3. A GENERAL OPTIMISATION SCHEME FOR IMAGE FUSION

In this section, the focus is placed in finding an autonomous image fusion system, based on several desired properties of the fused image. As described earlier, the fused image \underline{u}_f in the transform domain will be given by a linear combination of the input images (equation (4)). We denote $\underline{w} = [w_1 \ w_2 \ \dots \ w_T]^T$. All elements of vector \underline{u}_i will contribute in the formation of the fused image, according to the weight w_i . Let us now define:

$$\underline{x}(n) = [u_1(n) \ u_2(n) \ \dots \ u_T(n)]^T \quad \forall n = 1, \dots, N^2 \quad (7)$$

Hence, the fusion procedure can be equivalently described by the following product:

$$\underline{u}_f(n) = \underline{w}^T \underline{x}(n) \quad \forall n = 1, \dots, N^2 \quad (8)$$

The problem of fusion can now be described as an optimisation problem of estimating \underline{w} , so that the fused image follows certain properties. It makes sense to assume that the fusion process enhances *sparsity* in the ICA domain. In other words, the fusion should emphasize the existence of strong coefficients in the transform, whilst suppress small values.

The connection between *sparsity* and ICA representations has been investigated thoroughly by Olshausen [7]. The basis functions, that emerge when adapted to static, whitened natural images under the assumption of statistical independence, resemble the Gabor-like spatial profiles of cortical simple-cell receptive fields. That is to say that the functions become spatially localised, oriented and bandpass. Because all of these properties emerge purely from the objective of finding sparse, independent components for natural images, the results suggest that the receptive fields of V1 neurons have been designed under the same principle. Therefore, it seems that

the actual non-distorted representation of the observed scene in the ICA domain should be more sparse than the distorted or individual sensor input. Consequently, an algorithm that maximises the sparsity of the fused image in the ICA domain should enhance its perceptive quality.

3.1. Laplacian priors

Assuming a Laplacian model for $u_f(n)$, we can perform Maximum Likelihood (ML) estimation of \underline{w} . The Laplacian probability density function is given below:

$$p(u_f) \propto e^{-\alpha|u_f|} \quad (9)$$

where α is a parameter that controls the width (variance) of the Laplacian. The likelihood expression for ML estimation can be given by:

$$L_n = -\log p(u_f|\theta_n) \propto \alpha|u_f| = \alpha|\underline{w}^T \underline{x}(n)| \quad (10)$$

Maximum Likelihood estimation can be performed by maximising the cost function $J(\underline{w}) = \mathcal{E}\{L_n\}$.

$$\max_{\underline{w}} \mathcal{E}\{\alpha|\underline{w}^T \underline{x}|\} \quad (11)$$

$$\text{subject to } \underline{e}^T \underline{w} = 1 \text{ and } \underline{w} > 0 \quad (12)$$

where $\underline{e} = [1 \ 1 \ \dots \ 1]^T$. The first derivative of the cost function is given by:

$$\frac{\partial J(\underline{w})}{\partial \underline{w}} = \frac{\partial}{\partial \underline{w}} \mathcal{E}\{\alpha|\underline{w}^T \underline{x}|\} = \alpha \mathcal{E}\{\text{sgn}(\underline{w}^T \underline{x}) \underline{x}\} \quad (13)$$

To solve the above optimisation problem, one has to consult methods for constrained optimisation. Using the Lagrange multipliers method for equality constraints and the Kuhn-Tucker conditions for inequality constraints is going to increase the computational complexity of the algorithm. In addition, the available data points for the estimation of the expectation are very limited to N^2 . Therefore, we propose to solve the unconstrained optimisation problem using a gradient ascent method and impose the constraints at each stage of the adaptation, by ensuring that the weights w_i remain always positive and they sum up to one.

$$\underline{w}^+ \leftarrow |\underline{w}|/(\underline{e}^T |\underline{w}|) \quad (14)$$

3.2. Verhulstian priors

Laplacian priors can often be unstable, as the $\text{sgn}(u)$ function in the update algorithm has a discontinuity at $u \rightarrow 0$ and may cause numerical instability during the update. Therefore, one can use alternate probabilistic priors that model sparsity, such as the *generalised Laplacian* or the *Verhulstian* distribution.

The *Verhulstian* probability density function can be defined, as follows:

$$p(u) = e^{-\frac{u-m}{s}} / s \left(1 + e^{-\frac{u-m}{s}}\right)^2 \quad (15)$$

where m, s are parameters that control the mean and the standard deviation of the density function. In our case, we will assume zero mean and therefore $m = 0$. We can now derive the log-likelihood function for ML estimation:

$$L_n = \frac{1}{s} \underline{w}^T \underline{x} + \log s + 2 \log \left(1 + e^{-\frac{1}{s} \underline{w}^T \underline{x}}\right) \quad (16)$$

Maximum Likelihood estimation can be performed in a similar fashion to Laplacian priors, by maximising the cost function $J(\underline{w}) = \mathcal{E}\{L_n\}$. Again, we will perform a gradient ascent algorithm, as explained in the previous section with a correcting step that will constrain the solutions in the solution space that is permitted by the optimisation problem. The gradient is calculated, as follows:

$$\frac{\partial J(\underline{w})}{\partial \underline{w}} = \frac{1}{s} \mathcal{E} \left\{ \frac{1 - e^{-\frac{1}{s} \underline{w}^T \underline{x}}}{1 + e^{-\frac{1}{s} \underline{w}^T \underline{x}}} \underline{x} \right\} \quad (17)$$

3.3. Adaptive Fusion algorithm outline

Consequently, the adaptive fusion algorithm can be outlined as follows:

1. Initialise $\underline{w} = \underline{e}/T$. This implies that we give equal importance to all input patches initially ("mean" rule).
2. Update the weight vector, as follows:
(using Laplacian priors)

$$\underline{w}^+ \leftarrow \underline{w} + \eta \mathcal{E}\{\text{sgn}(\underline{w}^T \underline{x}) \underline{x}\} \quad (18)$$

(using Verhulstian priors)

$$\underline{w}^+ \leftarrow \underline{w} + \eta \mathcal{E} \left\{ \frac{1 - e^{-\frac{1}{s} \underline{w}^T \underline{x}}}{1 + e^{-\frac{1}{s} \underline{w}^T \underline{x}}} \underline{x} \right\} \quad (19)$$

where η represents the learning rate.

3. Apply the constraints, using the following update rule:

$$\underline{w}^+ \leftarrow |\underline{w}|/(\underline{e}^T |\underline{w}|) \quad (20)$$

4. Iterate steps 2, 3 until convergence.

4. EXPERIMENTS

We explore the performance of the proposed scheme 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 the Image Fusion Server [8]. 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. The concept of ground truth is not really meaningful in this case and therefore, we

can not have any numerical performance evaluation. We have tested the DTWT and the Topographic ICA (TopoICA) bases in various fusion rules. The maxabs rule in both cases manages to preserve most of the edge information, however, the two adaptive fusion scheme seem to balance between the details and the different contrast information that exists in the three different modality input images. The perception quality of the fused image using the adaptive schemes seems to have increased.

5. CONCLUSION

In this paper, the authors have made a step towards a more autonomous fusion system, where the fusion coefficients are selected by optimising several criteria. We have explored the use of *sparsity* as a criterion that can be used for fusion. Maximum likelihood estimation was used to calculate the coefficients that can lead to a sparser representation in the localised ICA domain. We experimented with both *Laplacian* and *Verhulstian* priors to represent sparsity during the Maximum Likelihood estimation with encouraging results.

6. REFERENCES

- [1] G. Piella, "A general framework for multiresolution image fusion: from pixels to regions," *Information Fusion*, vol. 4, pp. 259–280, 2003.
- [2] S.G. Nikolov, J.J. Lewis, R.J. O'Callaghan, D.R. Bull, and C.N. Canagarajah, "Hybrid fused displays: between pixel- and region-based image fusion," in *Proc. 7th Int. Conf. on Information Fusion*, Stockholm, Sweden, 2004, pp. 1072–1079.
- [3] N. Mitianoudis and T. Stathaki, "Pixel-based and region-based image fusion schemes using ICA bases," *Information Fusion*, in press. Available online from Science Direct., 2006.
- [4] A. Hyvärinen, P. O. Hoyer, and M. Inki, "Topographic independent component analysis," *Neural Computation*, vol. 13, no. 7, pp. 1527–1558, 2001.
- [5] A. Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. on Neural Networks*, vol. 10, no. 3, pp. 626–634, 1999.
- [6] A. Hyvärinen, P. O. Hoyer, and E. Oja, "Image denoising by sparse code shrinkage," in *Intelligent Signal Processing*, S. Haykin and B. Kosko, Eds. IEEE Press, 2001.
- [7] B.A. Olshausen, *Sparse Codes and Spikes*, In: Probabilistic Models of the Brain: Perception and Neural Function., MIT Press, 2002.
- [8] The Image fusion server, "<http://www.imagefusion.org/>,"

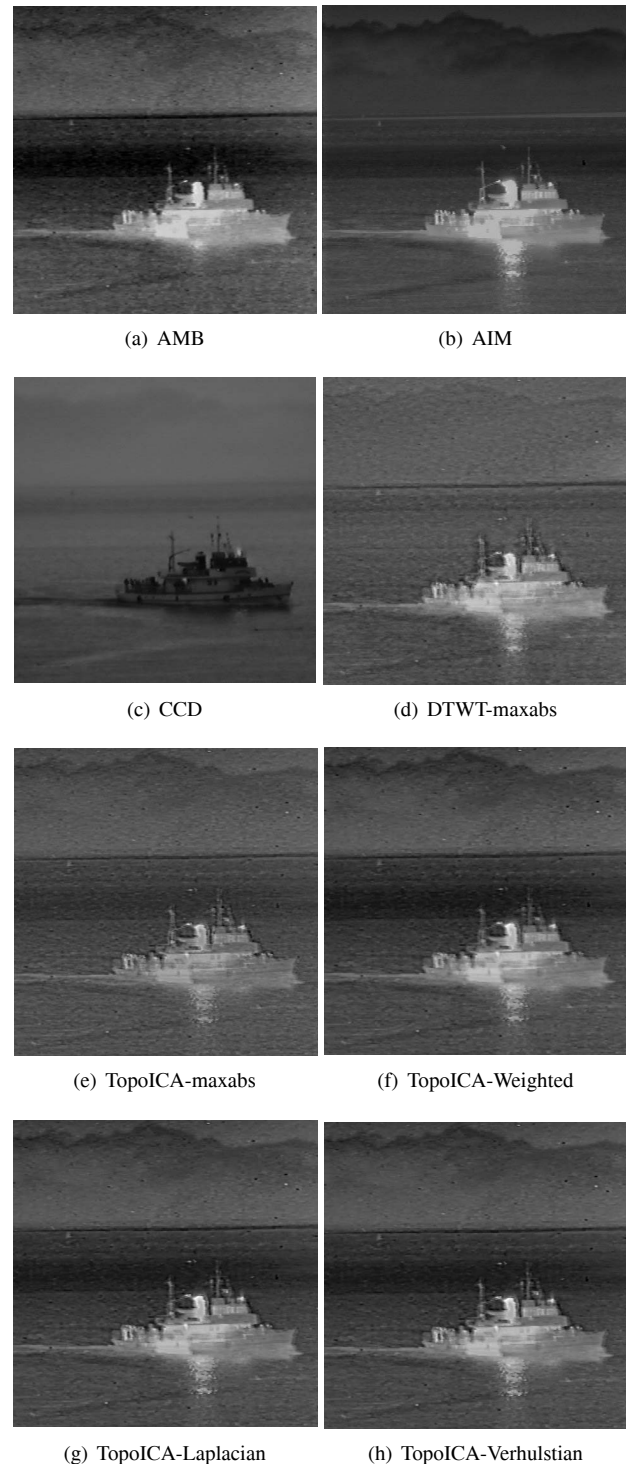


Fig. 2. Multi-modal image fusion: Three images acquired through different modality sensors and various fusion results with various transforms and fusion rules.