MULTIMODAL IMAGE FUSION IN SENSOR NETWORKS USING INDEPENDENT COMPONENT ANALYSIS

Nedeljko Cvejic, David Bull and Nishan Canagarajah

Department of Electrical and Electronic Engineering, University of Bristol
Merchant Venturers Building, Woodland Road, Bristol BS8 1UB, United Kingdom
{n.cvejic,dave.bull,nishan.canagarajah}@bristol.ac.uk

ABSTRACT

We present a novel image fusion algorithm based on ICA that has an improved performance over sensor networks. It employs segmentation to determine the most important regions in the input images and consequently fuses the ICA coefficients from the given regions. Sparse coding of the coefficients in ICA domain is used to minimize noise transferred from input images into the fused output. Experimental results confirm that the proposed method outperforms other state-of-the-art methods in the sensor network environment, characterized by JPEG 2000 compression and data packetisation.

Index Terms— image fusion, region-based fusion, independent component analysis, sensor networks, JPEG 2000

1. INTRODUCTION

As the size and cost of sensors decrease, sensor networks are increasingly becoming an attractive method to collect information in a given area [1]. However, there are still many technical challenges, mainly related to fusing the individual sensor data through an intelligent decision making process while reducing errors and compression noise. The task of the fusion algorithm is to combine the useful information from the input sensors to form a composite that represents the observed scene more adequately than using a single sensor [1].

Image and video fusion is a subarea of the more general topic of data fusion, dealing with image and video data. Instead of using a standard basis system, such as the DFT, one can train a set of bases that are suitable for a specific type of images. A training set of image patches, which are acquired randomly from images of similar content, can be used to train a set of statistically independent bases. Independent Component Analysis (ICA) is a widely used method that is able to identify statistically independent basis vectors in a linear generative model [2].

In this paper, we present a novel algorithm for fusion of multimodal images based on the ICA. It was tested in a sensors network environment and it has exhibited an improvement in performance in fusion of infrared (IR) and visible images over other state-of-the-art methods.

2. IMAGE ANALYSIS USING ICA

In order to obtain a set of statistically independent bases for image fusion in the ICA domain, training is performed with a predefined set of images. Training images are selected in such a way that the statistical properties are similar for the training images and the images to be fused. An input image $i(x, y)$ is randomly windowed using a rectangular window w of size $N \times N$, centered around the pixel $(m_0, n_0))$. The result of windowing is an "image patch" p:

$$p(m, n) = w(m, n) \cdot i(m_0 - N/2 + m, n_0 - N/2 + n) \quad (1)$$

where m and n take integer values from the interval $[0, N - 1]$. Each image patch $p(m, n)$ can be represented by a linear combination of a set of M basis patches $b_i(m, n)$ [3]:

$$p(m, n) = \sum_{i=1}^{M} v_i b_i(m, n) \quad (2)$$

where $v_1, v_2, ..., v_M$ stand for the projections of the original image patch on the basis patch, i.e. $v_i = \langle p(m, n), b_i(m, n)\rangle$. A 2D representation of the image patches can be simplified to a 1D representation, using lexicographic ordering [3]. This implies that an image patch $p(m, n)$ is reshaped into a vector $\underline{p}$, mapping all the elements from the image patch matrix to the vector in a row-wise fashion. Decomposition of image patches into a linear combination of basis patches can the be expressed as follows:

$$\underline{p}(t) = \sum_{i=1}^{M} v_i(t)\underline{b}_i = [\underline{b}_1\underline{b}_2...\underline{b}_M] \cdot [v_1(t)v_2(t)v_M(t)]^T \quad (3)$$

where t represents the image patch index. If we denote $B = [b_1b_2...b_M]$ and $v(t) = [v_1v_2...v_M]^T$, then equation (3) reduces to:

$$\underline{p}(t) = B\underline{v}(t) \quad (4)$$

$$\underline{v}(t) = B^{-1}\underline{p}(t) = A\underline{p}(t) \quad (5)$$

Thus, $B = [b_1b_2...b_M]^T$ represents an unknown mixing matrix (analysis kernel) and $A = [a_1a_2...a_M]^T$ the unmixing

1-4244-0882-2/07/$20.00 ©2007 IEEE

matrix (synthesis kernel) [2]. This transform projects the observed signal $\underline{p}(t)$ on a set of basis vectors. The aim is to estimate a finite set of $K < N^2$ basis vectors that will be capable of capturing most of the input image properties and structure. In the first stage of basis estimation the Principal Component Analysis (PCA) is used for dimensionality reduction [2]. After the input image patches $\underline{p}(t)$ are transformed to their ICA domain representations $\underline{v}_k(t)$, we can perform image fusion in the ICA domain in the same manner as it is performed in e.g. the wavelet domain. After the composite image $v_f(t)$ is constructed in the ICA domain, we can move back to the spatial domain, using the synthesis kernel A, and synthesise the image $i_f(x, y)$.

3. PROPOSED FUSION METHOD

3.1. Separated Training Sets

In the proposed method, images used for training of the ICA bases are separated in two groups prior to the training process. Namely, all IR training images are grouped into a separate training subset, whereas all the visible training images constitute the second training subset. Introduction of separate training subsets provides us with two sets of ICA bases. The first ICA basis set is used to decompose the IR input image patches $v_i(t) = A_i p_i(t)$ and the second subset to transform the visible input image patches to ICA domain $v_v(t) = A_v p_v(t)$.

Separate ICA basis sets for decomposition of input images are more specifically trained to capture statistical properties of the specific modality of the input images (IR/visual). This enables the proposed method to outperform the standard method [3], in which images of both IR and visible modality are used for training which results in an "average" ICA bases set that is not able to take the full advantage of ICA decomposition. Fig. 1 confirms that when two separate training sets are used, the subjective quality of the fused image is increased considerably; e.g. fence detail is far more visible and person walking is brighter and less blurred. In Table I, subjective impression is confirmed by values obtained by Petrovic image fusion metric [4]. The metric is one of the most widespread tools for evaluation of image fusion algorithms. It uses the amount of edge information transferred from the source image to the fused image to give an estimation of the performance of a fusion algorithm [4]. It is clear that significantly higher metric values are obtained using separate training sets. Table I also shows that performance of the ICA fusion algorithm does not improve significantly when the number of training patches exceeds 10^3. Thus, the number of training patches has been fixed to 10^3 in order to make a trade-off between performance and computational complexity of the algorithm.

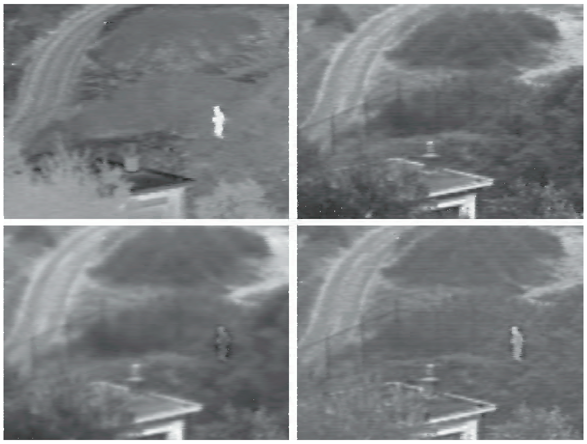

Fig. 1. Impact of the number of patches on the subjective quality of the fused images, image 1812, UN Camp sequence

Table 1. Fusion performance measured by Petrovic metric, the figures represent the mean value of the metric over 25 images, Octec sequence.

Fusion	Number of ICA training patches				
method	100	200	10^3	10^4	$4{\cdot}10^4$
Standard ICA	0.472	0.533	0.581	0.588	0.592
Proposed ICA	0.320	0.355	0.396	0.406	0.406

3.2. The segmentation algorithm

Our experiments showed that important objects in the IR input images (e.g. a person or a smaller object) are often masked by textured high-energy background in the visual image. In this case the important objects from the IR image become blurred or, in extreme cases, completely masked. Therefore, we perform segmentation in the spatial domain and then fuse patches from separate regions separately. This differs from the methods in [3, 5] where the fusion was performed on a more general, pixel level. The quality of the segmentation algorithm is of vital importance to the fusion process. An adapted version of the combined morphological−spectral unsupervised image segmentation algorithm is used, which is described in [6], enabling it to handle multi-modal images.

The algorithm works in two stages. The first stage produces an initial segmentation by using both textured and non-textured regions. The detail coefficients of the DT-CWT are used to process texture. The gradient function is applied to all levels and orientations of the DT-CWT coefficients and up-sampled to be combined with the gradient of the intensity information to give a perceptual gradient. The larger gradients indicate possible edge locations. The watershed transform of the perceptual gradient gives an initial segmentation. The second stage uses these primitive regions to produce a graph representation of the image which is processed using a spectral clustering technique.

The method can use either intensity information or textural information or both to obtain the segmentation map. This flexibility is useful for multi-modal fusion where some a priori information of the sensor types is known. For example, IR images tend to lack textural information with most features having a similar intensity value throughout the region. Therefore, we used an intensity only segmentation map, as it gives better results than a texture based segmentation.

The segmentation can be performed either separately or jointly. For separate segmentation, each of the input images generates an independent segmentation map for each image.

$$S_1 = \sigma(i_1, D_1), \ldots, S_N = \sigma(i_N, D_N) \qquad (6)$$

where D_n represent detail coefficients of the DT-CWT used in segmentation. Alternatively, information from all images could be used to produce a joint segmentation map.

$$S_{joint} = \sigma(i_1 \cdots i_N, D_1 \cdots D_N) \qquad (7)$$

In general, jointly segmented images work better for fusion. This is because the segmentation map will contain a minimum number of regions to represent all the features in the scene most efficiently. A problem can occur for separately segmented images, where different images have different features or features which appear as slightly different sizes in different modalities. Where regions partially overlap, if the overlapped region is incorrectly dealt with, artefacts will be introduced and the extra regions created to deal with the overlap will increase the time taken to fuse the images. After the images are jointly segmented it is essential to determine the importance of regions in each of the input images. We have decided to use the normalized Shannon entropy of a region as the priority. Thus, the priority $P(r_{t_n})$ is given as:

$$P(r_{t_n}) = \frac{1}{|r_{t_n}|} \sum_{\forall\theta, \forall l, (x,y) \in r_{t_n}} d^2_{n(\theta,l)}(x,y) \log d^2_{n(\theta,l)}(x,y) \qquad (8)$$

with the convention $0 \log(0) = 0$, where $|r_{t_n}|$ is the size of the region r_{t_n} in input image n and $d_{n(\theta,l)}(x,y) \in D_{n(\theta,l)}$ detail coefficients of the DT-CWT used in segmentation. Finally, a mask M is generated that determines which image each region should come from in the fused image. An example of the IR input image, visual input image, performed joint segmentation and the image fusion mask is given in Fig. 2.

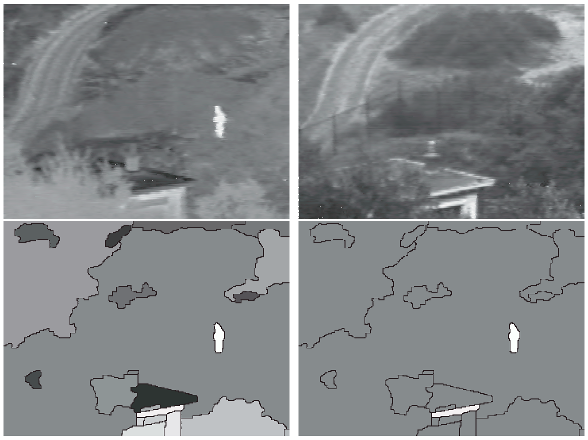

Fig. 2. Segmentation and region selection prior to fusion. Top: IR input image (left), visible input image (right). Bottom: Regions obtained by joint segmentation of input images (left), image mask: white from IR, grey from visible (right).

3.3. Nonlinear shrinkage of coefficients in ICA domain

However, in the case when the images to be fused and set of training images are corrupted with noise, it is crucial to determine the ICA coefficients to be used in the reconstruction of the fused image so that the noise transferred from input images into the fused output is minimized. Thus, we decided to use an approach similar to image denoising algorithms in the ICA domain [7] to reduce noise in the fused image.

Assume that we observe an $N-$dimensional vector x as: $x = s + n$, where s is the vector of the original signal and n is Gaussian white noise. The goal of signal denoising is to find $\overline{s} = g(x)$ such that $\overline{n}$ is close to n in some well-defined sense. The ICA algorithms we have described in Section 2 do not include the presence of noise. In principle one could apply the ICA method to noisy data and assume that the method would work as before, but for noisy images it is optimal to use advanced methods, such as [7]. We use the modified version of the algorithm in [7], summarized as following:

1. Use database of noiseless data z for training process, different from the images to be fused, but with similar statistical properties. Use FastICA to get ICA transformation matrix B.
2. For each component $s_i = b_i^T z$, estimate a density model (usually supergaussian) and a nonlinear function g_i. Essentially g_i represent a shrinkage function that is commonly used in wavelet image processing [7].
3. Transform noisy vector x, to a sparse basis $c = Wx$.
4. Perform componentwise nonlinear processing $\hat{s}_i = g_i(c_i)$.
5. Inverse transform, $\hat{v} = W^{-1}\hat{s}$ to get denoised image.

The density model for s_i that we used is a mixture of a normal and the Laplacian density, $p(s_i) = Ce^{(-as_i^2/2 - b|s_i|)}$ where C is a normalizing constant. Parameters b and a can be estimated from statistics (i.e., the mean and variance) of s_i [7]. The nonlinear function that we used in the experiments was $g(x) = \frac{1}{1+\sigma^2 a} sign(x) \max(0, |x| - b\sigma^2)$, where σ is the noise variance.

4. EXPERIMENTAL RESULTS

The proposed image fusion method was tested in a surveillance scenario with two input images: infrared and visible.

The images used in experiments are surveillance images from TNO Human Factors and Octec Ltd., publicly available at the Image Fusion web site. We compared the proposed method with a simple averaging method, the contrast pyramid (CP) method, ratio pyramid (RP) method and the dual-tree complex wavelet transform (DT-CWT)[1]. CP, RP and DT-CWT methods have been chosen for comparison because they have been previously reported to obtain excellent performance in multimodal image fusion [1, 4]. Before performing image fusion using the proposed algorithm, the ICA bases were trained using a set of images with content comparable to the test set. The number of rectangular patches $(N = 8)$ used for training was 1000, randomly selected from the training set. Obtained ICA coefficients are combined using the principle described in Section 3, while reconstruction of the fused image was done using optimisation based on the Petrovic metric [4].

Table 2. Fusion performance measured by Petrovic metric, mean value of the metric over 25 images, Octec sequence.

Fusion	Packet length (bytes)					
method	1000	500	200	100	50	P_{pl}
DT-CWT	0.506	0.507	0.503	0.441	0.413	10^{-3}
Ratio	0.415	0.415	0.416	0.375	0.358	
Contrast	0.465	0.466	0.463	0.416	0.378	
Average	0.349	0.348	0.348	0.319	0.311	
ICA	0.588	0.589	0.587	0.490	0.476	
DT-CWT	0.451	0.414	0.351	0.298	0.211	10^{-2}
Ratio	0.378	0.351	0.317	0.284	0.230	
Contrast	0.419	0.391	0.330	0.275	0.196	
Average	0.322	0.294	0.285	0.267	0.229	
ICA	0.512	0.452	0.410	0.344	0.266	
DT-CWT	0.437	0.351	0.256	0.216	0.169	10^{-1}
Ratio	0.385	0.306	0.249	0.225	0.186	
Contrast	0.414	0.335	0.242	0.200	0.155	
Average	0.319	0.262	0.243	0.221	0.200	
ICA	0.458	0.370	0.260	0.257	0.210	

In order to evaluate performance of the image fusion algorithms in the sensor network environment, input images were first compressed using JPEG 2000. Sensor network transmission was simulated by dividing the image into data packets which were transmitted at a given probability of packet loss (P_{pl}). After the recovered packets at the receiver side were recomposed into images, these images were used as inputs for the fusion algorithms. The performance of methods was measured using the Petrovic metric and the results are given in Table 2 and Fig. 3. Results in Table 2 and subjective quality of the fused images in Fig. 3 show that the proposed algorithms performs significantly better than the other state-of-the-art methods in the sensor network environment, for all data packet lengths and probabilities of packet loss.

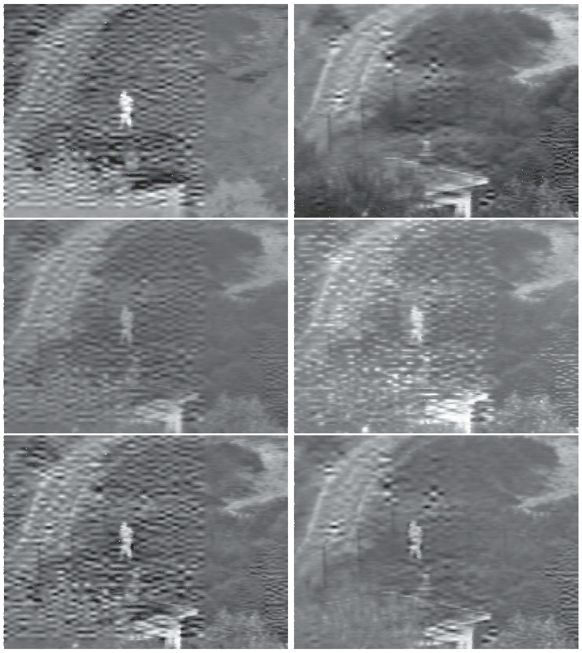

Fig. 3. Subjective fusion results. Top: input IR image (left), input visible image (right). Middle: fused image using averaging (left) and ratio pyramid (right). Bottom: fused image using DT-CWT (left) and the proposed ICA method (right)

5. REFERENCES

[1] R. Blum and Z. Liu, *Multi-sensor Image Fusion And Its Applications*, CRC Press, London, UK, 2005.

[2] A. Hyvrinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, John Wiley, London, UK, 2001.

[3] N. Mitianoudis and T. Stathaki, “Pixel-based and region-based image fusion schemes using ICA bases,” *Information Fusion*, to appear.

[4] V. Petrovic and C. Xydeas, “Objective evaluation of signal-level image fusion performance,” *Optical Engineering*, vol. 44, no. 8, pp. 087003, 2005.

[5] N. Cvejic, D. Bull, and N. Canagarajah, “A novel ICA domain multimodal image fusion algorithm,” in *Proc. SPIE*. 2006, pp. 62420W1–8, Orlando, FL.

[6] R. O’Callaghan and D. Bull, “Combined morphological-spectral unsupervised image segmentation,” *IEEE Transactions on Image Processing*, vol. 14, pp. 49–62, 2005.

[7] S. Chettri and W. Campbell, “Denoising remotely sensed digital imagery,” in *Proc. IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data*. 2006, pp. 193–201, Washington, DC.