Optimal Contrast for Color Image Fusion using ICA bases

Nikolaos Mitianoudis
Department of Electrical and Electronic Engineering
Imperial College
London, U.K.
Email: n.mitianoudis@imperial.ac.uk

Tania Stathaki
Department of Electrical and Electronic Engineering
Imperial College
London, U.K.
Email: t.stathaki@imperial.ac.uk

Abstract—Image Fusion is the procedure of combining useful features from multiple sensor image inputs to a single composite image. In this work, the authors revise the previously proposed Image Fusion framework, based on self-trained Independent Component Analysis (ICA) bases. In the original framework, equal importance was given to all input images in the reconstruction of the "fused" image's intensity. Even though this assumption holds for applications involving sensors of the same modality, it might not be valid in the case of multiple modality inputs of different intensity range. The authors propose a method of estimating the optimal intensity range (contrast) of the fused image via optimization of an image fusion index. The proposed approach is extended in the case of color visual inputs, in order to produce a color fused image.

Keywords: Image Fusion, Independent Component Analysis (ICA), ICA bases.

I. INTRODUCTION

Modern technology has enabled the development of low-cost, wireless sensors of various modalities that can be deployed to monitor a scene. In this paper, the case of multi-modal imaging sensors of known position, that are employed to monitor a scene, will be investigated. The information provided by multimodal sensors can be quite diverse. Each image has been obtained using different instruments or acquisition techniques, allowing each image to have different characteristics, such as degradation, thermal and visual characteristics. Multispectral sensors are increasingly being employed in military applications [1].

Let $x_1(i,j), \ldots, x_T(i,j)$ represent T images of size $M_1 \times M_2$ capturing the same scene, where i,j refer to the pixel coordinates in the image. The input images are assumed to have negligible registration problems. The process of combining the important features from the original T images to form a single enhanced image $f(i,j)$ is referred to as *Image Fusion*. Fusion techniques can be divided into *spatial domain* and *transform domain* techniques [2], depending on the domain, where the fusion procedure is performed. Several transformations were proposed to be used for image fusion, including the *Dual-Tree Wavelet Transform* [2], *Pyramid Decomposition* and self-trained Independent Component Analysis bases [3]. All these transformations project the input images onto localized bases, modeling sharp and abrupt transitions (edges) and therefore, transform the image into a more meaningful representation that can be used to detect and emphasize salient features, important for performing the task of image fusion.

The authors proposed a self-trained Image Fusion framework based on Independent Component Analysis, where the analysis transformation is estimated from a selection of images of similar content [3]. Several fusion rules were proposed in conjunction with this framework in [3]. The analysis framework is projecting the images into localized patches of small size. The local mean value of the patches is subtracted and stored in order to reconstruct the local means of the fused image. In [3], an average of the stored means was used to reconstruct the fused image. In [4], [5], it was demonstrated that this choice might not be optimal in several multi-modal cases and proposed an exhaustive search solution of the optimum performance in terms of the Piella and Heijmans Fusion Quality index [6] for the case of two input sensors. In this paper, the authors examine and provide a complete solution to this problem for the general T sensor color image fusion scenario, based on the Fusion Quality Index in [6].

II. INTRODUCTION TO IMAGE FUSION USING ICA BASES

Assume an image $x(i,j)$ of size $M_1 \times M_2$. An "image patch" x_w is defined as an $N \times N$ neighborhood centered around the pixel (i_0, j_0). Assume that there exists a population of patches x_w, acquired randomly from the image $x(i,j)$. Each image patch $x_w(k,l)$ is arranged into a vector $\underline{x}_w(t)$, using lexicographic ordering. The vectors $\underline{x}_w(t)$ are normalized to zero mean, producing unbiased vectors. These vectors can be expressed as linear combination of the bases vectors $\underline{b}_j$ with weights $u_i(t), i = 1, \ldots, K$:

$$\underline{x}_w(t) = \sum_{k=1}^{K} u_k(t)\underline{b}_k = [\underline{b}_1 \ \underline{b}_2 \ldots \underline{b}_K] \begin{bmatrix} u_1(t) \\ u_2(t) \\ \ldots \\ u_K(t) \end{bmatrix} \quad (1)$$

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

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

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

where $B = [\underline{b}_1 \ \underline{b}_2 \ldots \underline{b}_K]$ and $\underline{u}(t) = [u_1(t) \ u_2(t) \ldots u_K(t)]^T$. In this case, $A = B^{-1} =$

$[\underline{a}_1 \ \underline{a}_2 \ldots \underline{a}_K]^T$ represents the *analysis* kernel and B the *synthesis* kernel. The estimation of these basis vectors is performed using a population of training image patches $\underline{x}_w(t)$ and a criterion (cost function) that selects the basis vectors. Analysis/synthesis bases can be trained using *Independent Component Analysis* (ICA) and Topographic ICA, as explained in more detail in [3]. The training procedure needs to be performed only once, as the estimated transform can be used for fusing images with similar content as the training images.

A number of $N \times N$ patches (in the order of 10000 [7]) are randomly selected from similar-content training images. We perform Principal Component Analysis (PCA) on the selected patches in order to select the $K < N^2$ most important bases. Then, the ICA update rule or the topographical ICA rule in [3] for a chosen $L \times L$ neighborhood is iterated until convergence. In each iteration, the bases are orthogonalised using a symmetric decorrelation scheme. In contrast to [4], [5], sample patches from all multimodal inputs are selected to train the ICA bases.

A. Fusion in the ICA domain

After estimating an ICA or Topographic ICA transform, Image fusion using ICA or Topographical ICA bases is performed following the approach depicted in the generic diagram of Figure 1. Every possible $N \times N$ patch is isolated from each image $x_k(i, j)$ and is consequently re-arranged to form a vector $\underline{x}_k(t)$. These vectors $\underline{x}_k(t)$ are normalized to zero mean and the subtracted local mean $MN_k(t)$ is stored for the reconstruction process. Each of the input vectors $\underline{x}_k(t)$ is transformed to the ICA or Topographic ICA domain representation $\underline{u}_k(t)$, using equation (3). Optional denoising in the ICA representation is also possible, by applying sparse code shrinkage on the coefficients in the ICA domain [7], assuming Laplacian (generally sparse) priors for the ICA representation. The corresponding coefficients $\underline{u}_k(t)$ from each image are then combined to construct a composite image representation $\underline{u}_f(t)$ in the ICA domain. The next step is to move back to the spatial domain, using the synthesis kernel B, and synthesise the image $f(i, j)$ by averaging the image patches $\underline{u}_f(t)$ in the same order they were selected during the analysis step.

B. Various fusion rules using ICA bases

Some basic rules that can be used for image fusion are described in this section. Fusion by the *absolute maximum* rule simply selects the greatest in absolute value of the corresponding coefficients in each image ("max-abs" rule). This process seems to convey all the information about the edges to the fused image, however, the intensity information in constant background areas seems to be distorted. In contrast, fusion by the *averaging* rule averages the corresponding coefficients ("mean" rule). This process seems to preserve the correct contrast information, however, the edge details seem to get oversmoothed.

A *Weighted Combination* (WC) pixel-based rule can be established using the ICA framework [3]. The fused image
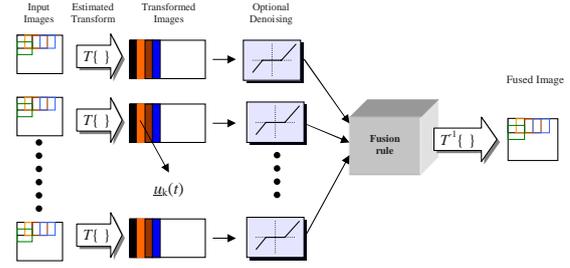

Figure 1. The proposed fusion system using ICA / Topographical ICA bases.

coefficients are constructed using a "weighted combination" of the input transform coefficients, i.e.

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

To estimate the contributions $w_k(t)$ of each image to the "fused" image, the mean absolute value ($\mathcal{L}_1$-norm) of each patch (arranged in a vector) in the transform domain can be employed as an activity indicator, because it fits a more general sparse profile of the ICA coefficients, denoted by a Laplacian distribution.

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

The weights $w_k(t)$ should emphasize 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.

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

A *regional* approach can also be established, by dividing the observed area into areas of "low" and "high" activity, using the $\mathcal{L}_1$-norm based $E_k(t)$ measurement. The areas containing salient information can be heuristically labeled as "high" activity areas, if $E_k(t) > 2\text{mean}_t\{E_k(t)\}$ and can be fused using a "max-abs" or a "weighted-combination" fusion rule. The remaining areas of "low-activity" contain background information and can be fused using the "mean" rule. Another regional approach can be to use alternative segmentations of the observed scene, based on the input sensor images and consequently fuse the different regions independently [4], [5]. In this work, the "max-abs" or the weighted combination rule will be used for ICA-based fusion for simplicity.

III. A NOVEL APPROACH FOR AUTOMATED CONTRAST CORRECTION

The next step is to estimate the spatial-domain representation $f(i, j)$ of the fused image. To reconstruct the image in the spatial domain, the process described in Section II is inverted. The vectors $\underline{u}_f(t)$ are re-transformed to the local $N \times N$ patches $u_f(k, l)$. The local mean of each patch is

restored using the stored patches means $MN_k(t)$ and add it to $u_f(k,l)$. This restoration appears not to be a trivial issue. There exist T local intensity values for each patch of the reconstructed image, each belonging to the corresponding input sensor. In the case of performing multi-focus image fusion, the local intensities from all input sensors will be similar for all corresponding patches, thus, the local means can be reconstructed by averaging the $MN_k(t)$, in terms of k. In the case of multi-modal image fusion, the problem of reconstructing the local intensities of the fused image becomes more serious, since the T input images are acquired from sensors with different intensity range. The fused image is an artificial image, that does not exist in nature, and it is therefore difficult to find a criterion that can dictate the most efficient way of combining the input sensors intensity range. The details from all input images will be transferred to the fused image by the fusion algorithm. However, an inappropriate choice of local means may render some of the local details totally invisible in the fused image and therefore deteriorate the fusion performance.

In [4], [5], the authors demonstrated that this might not be an optimal solution to the problem, where optimality is defined in terms of the Piella and Heijmans index [6][1]. They proposed a method to estimate a weighted averaging of the corresponding local means in a two-sensor fusion scenario. In the two-sensor case, the proposed 2D optimization problem can be reduced to an 1D problem, since the weights should sum to unity. The 1D optimization was solved by numerically assessing the Piella index for quantized values of $q_1 \in [0,1]$ with a step value of 0.1 to infer approximately about the value of q_1 that maximizes the Piella index. This approach relies on a valid concept, however, it is rather computationally expensive to expand the proposed implementation in [4], [5] in a multiple sensor scenario. In addition, the approach in [4], [5] assumes the existence of a single global optimum of the cost function based on the Piella Index, an assumption for which no theoretical justification was provided.

A. Optimizing Piella's Index

In this section, we propose a general estimation technique of the optimal local mean of the fused image patches. The proposed technique can be applied to a N-D input scenario and performs iterative estimation of the optimal local solution that does not require exhaustive search of the solution space. Assume that $m_{x_1}, m_{x_2}, \ldots, m_{x_T}$ are the means of the input sensors images, m_f is the mean of the fused image and $q_1, q_2, \ldots, q_T$ are the weights that will be used to estimate m_f via the equation:

$$m_f = q_1 m_{x_1} + q_2 m_{x_2} + \ldots + q_T m_{x_T} \quad (7)$$

The Wang and Bovik Image Quality Index [8] of an estimated image f in relation to the source image x is given by the

[1]For simplicity we will refer to this index as the Piella Index for the rest of the document

following formula:

$$Q(x,f) = \frac{2\sigma_{xf}}{\sigma_x^2 + \sigma_f^2} \frac{2m_x m_f}{m_x^2 + m_f^2} \quad (8)$$

where σ_{xf} represents the correlation between the two images and σ_f, σ_x represent the standard deviations of the two images respectively. Let $Q_\sigma(x,f) = \frac{2\sigma_{xf}}{\sigma_x^2 + \sigma_f^2}$ and $Q_m(x,f) = \frac{2m_x m_f}{m_x^2 + m_f^2}$. It is straightforward to see that the Image Quality Index can be factorized into the term Q_σ that is dependent on correlation/variance figures and the term Q_m that is dependent on mean values. More specifically,

$$Q(x,f) = Q_\sigma(x,f)Q_m(x,f) = Q_\sigma(x,f)\frac{2m_x m_f}{m_x^2 + m_f^2} \quad (9)$$

The first version of the Piella Index is based on the Wang and Bovik index. The Piella Index simply segments the input sensor images and the fused image into multiple overlapping patches and estimates the contribution of each input patch to the fused image in terms of the Image Quality Index. These scores are scaled according to the local information quality λ_i of each patch, i.e. local variance, in order to emphasize the proper scores concerning the information that was actually used during the fusion process. Note that λ_i is normalized to the total local information quality (saliency) of the corresponding input patches. This implies that $\sum_{i=1}^{T} \lambda_i = 1$. Therefore for a single patch, the Piella Index is defined as:

$$\begin{aligned} Q_p^n &= \lambda_1 Q(x_1, f) + \lambda_2 Q(x_2, f) + \ldots + \lambda_T Q(x_T, f) \\ &= \sum_{i=1}^{T} \lambda_i Q(x_i, f) \end{aligned} \quad (10)$$

To estimate the expected Piella index for all the extracted patches of the images, one has to estimate the following expression:

$$Q_p = \mathcal{E}\{\sum_{i=1}^{T} \lambda_i Q(x_i, f)\} \quad (11)$$

The expectation in the above equation will be approximated by an average of all patches, producing the first version of the Piella Index. This is similar to attributing equal probability for each patch, i.e. assuming a uniform prior and consequently the expectation is approximated by the sample mean. There is also a second version of Piella's index, where this expectation is approximated by a weighted sum of the individual terms. The imposed weights are based on the importance of each frame, in terms of maximum input sensor saliency, compared to the total maximum saliency of the samples. A third version was also proposed in by estimating the second version of the index for the edge maps of the input sensors and fused images and multiplying it with the original second index version [6]. In this optimization the first version is considered for simplicity.

The next step is to optimise Q_p in terms of $\underline{q}$ in order to estimate the mean m_f and essentially the weights $\underline{q}$. This adaptation will not necessarily affect the energy, i.e. the local activity of the patch in the fused image and its comparison to the activity input sensor patches. The bias (i.e. the local

mean) is the only factor that is changed during the adaptation of m_f. Let $A_{\sigma_{x_i}} = \lambda_i Q_\sigma(x_i, f)$, then the above equation can be expressed as:

$$Q_p(\underline{q}) = \mathcal{E}\left\{\sum_{i=1}^{T} A_{\sigma_{x_i}} \frac{2 m_{x_i} m_f}{m_{x_i}^2 + m_f^2}\right\} \qquad (12)$$

The objective is to estimate $\underline{q} = [q_1 \ q_2 \ \ldots \ q_T]^T$ by maximizing Q_p. Obviously,

$$\frac{\partial Q_p}{\partial \underline{q}} = \begin{bmatrix} \frac{\partial Q_p}{\partial q_1} \\ \frac{\partial Q_p}{\partial q_2} \\ \ldots \\ \frac{\partial Q_p}{\partial q_T} \end{bmatrix} \qquad (13)$$

The term $\partial Q_p / \partial q_i$ can be expressed, as follows:

$$\frac{\partial Q_p}{\partial q_i} = \frac{\partial Q_p}{\partial m_f} \frac{\partial m_f}{\partial q_i} = m_{x_i} \frac{\partial Q_p}{\partial m_f} \qquad (14)$$

Consequently,

$$\frac{\partial Q_p}{\partial \underline{q}} = \frac{\partial Q_p}{\partial m_f} \begin{bmatrix} m_{x_1} \\ m_{x_2} \\ \ldots \\ m_{x_T} \end{bmatrix} = \frac{\partial Q_p}{\partial m_f} \underline{m}_x \qquad (15)$$

$$\frac{\partial Q_p}{\partial m_f} = \mathcal{E}\left\{\sum_{i=1}^{T} A_{\sigma_{x_i}} m_{x_i} \frac{m_{x_i}^2 - m_f^2}{(m_{x_i}^2 + m_f^2)^2}\right\} \qquad (16)$$

Performing gradient ascent on the cost function that is based on Piella index, the following updates are given for the weights $\underline{q}$:

$$\underline{q}^+ \leftarrow \underline{q} + \eta \mathcal{E}\left\{\underline{m}_x \sum_{i=1}^{T} A_{\sigma_{x_i}} m_{x_i} \frac{m_{x_i}^2 - m_f^2}{(m_{x_i}^2 + m_f^2)^2}\right\} \qquad (17)$$

where η denotes the learning rate. The above rule is iterated until convergence and consequently the estimated weights are employed to reconstruct the means of the fused image. To avoid extreme situations or deformations in the weights q_i during the adaptation, the following restriction is imposed during each step:

$$\underline{q}^+ \leftarrow |\underline{q}| / ([\ 1 \quad 1 \quad \ldots \quad 1\ ] |\underline{q}|) \qquad (18)$$

This restriction ensures that the weights remain positive during the adaptation and their summation is unity.

The existence of a single solution in the optimization of the above function can not be guaranteed. The nominator of the cost function's first derivative is a polynomial of degree $4T-2$, that may have more than a single root in the solution space in general and i.e. multiple optima. There is no sufficient and required condition that can guarantee the existence of a single maximum for this cost function. Nevertheless, no single case of multiple optima was encountered in this work or in Cvejic et al [4], [5].

IV. COLOR IMAGE FUSION USING THE ICA-BASED FRAMEWORK

An additional problem is the creation of a "color" fused image. Let us assume that one of the input sensors is a visual sensor. A visual sensor will provide a color input image or in other terms a number of channels representing the color information. The most common representation in Europe is the three channel RGB (Red-Green-Blue) representation. If the traditional fusion methodology is applied on this problem, a single channel "fused" image will be produced featuring only intensity changes in grayscale. However, most users and operators will demand a color rather than a grayscale representation of the "fused" image. Therefore, the problem of creating a 3-channel representation of the "fused" image from T channels available by the input sensors can be rather demanding.

A first thought would be to treat each of the visual color channels independently and fuse with the input channels from the other sensors independently to create a three channel representation of the "fused" image. Although this technique seems rational and may produce satisfactory results in several cases, it does not utilize the dependencies between the color channels that might be beneficial for the fusion framework [9]. Another proposed approach [9] is to move to another color space, such as the YUV color space that describes a color image using one luminance and two chrominance channels [9] or the HSV color space that describes a color image using Hue, Saturation and Intensity (luminance) channels. The two chrominance channels as well as the hue-saturation channels convey color information solely, whereas the Intensity channel describes the image details more accurately. Therefore, the proposed strategy is to fuse the intensity channel with the other input sensor channels and create the intensity channel for the "fused" image. The chrominance/hue-saturation channels can be used to provide color information for the "fused" image. This scheme features reduced computational complexity as one visual channel is fused instead of the original three. In addition, as all these color transformations are linear mappings from the RGB space, one can use Principal Component Analysis to define the principal channel in terms of maximum variance. This channel is fused with the other input sensors and the resulting image is mapped back to the RGB space, using the estimated PCA matrix. The above techniques are producing satisfactory results in the case of color out-of-focus input images, since all input images have the same chrominance channels. In the case of multi-modal or multi-exposure images, these methods may not be sufficient and then one can use more complicated color channel combination and fusion schemes in order to achieve an enhance "fused" image. The discussed schemes may offer enhanced performance for selected applications but may not in several other applications thus their effectiveness should be investigated beforehand.

V. EXPERIMENTS

In this section, the performance of the proposed adaptive ICA-based scheme is evaluated using a variety of datasets

that were employed by the Image Fusion community or were provided by our partners in the Applied Multi-Dimensional Fusion (AMDF) cluster project : QinetiQ and General Dynamics UK. We used the typical training procedure for the ICA framework, training 60 8×8 ICA bases for each dataset. For performance comparison, the Dual-Tree Wavelet Transform (DT-WT) method using the "max-abs" rule will also be employed[2]. The Piella Index that is calculated in this section will constantly represent the second version of the Piella Index, as explained earlier.

The first experiment is to employ two images that can demonstrate the usefulness of Image Fusion and the rectification in contrast to the previous ICA-based framework offered by the proposed approach. The two Octet image sets were employed, as they were available by the ImageFusion Server [11]. These images, captured by Octec Ltd., show men and buildings with (Test Images 2, see Figure 3) and without (Test Images 1, see Figure 2) a smoke screen. They were captured with a Sony Camcorder and a LWIR sensor. We employed the original ICA-based scheme, the proposed ICA-based scheme and the DT-WT framework to perform fusion of the two images. For the ICA-based framework, the Weighted Combination rule was used for the first example and the "max-abs" for the second example. The RGB image was transformed to the YUV representation where the Y-channel (luminance) was employed for fusion with the LWIR image for all fusion approaches. The proposed algorithm converged smoothly after $40 - 50$ iterations to the optimal weights $q = [0.3577\ 0.6423]$ for the first example and $q = [0.3613\ 0.6387]$ for the second example. In Figures 2,3, the fusion results of the three algorithms were depicted. The optimal contrast approach weights the significance of the local intensities of the two images, resulting into a more balanced representation in the fused image. The targets and smoke outlines are more visible and balanced in the visual input, conveying both the existence of targets and smoke in the observed image, without one obstructing the other. The Piella Indexes for the three methods and the two tests are shown in Table I, demonstrating the performance enhancement of the proposed scheme.

The proposed ICA-based Image Fusion approach is consequently tested on the "MX-15" dataset that was kindly provided to AMDF by QinetiQ. The provided videos feature two real-life streams that were captured by an airborne vehicle patrolling a British town. The first stream is a sequence captured by a visual camera and the second stream is a sequence captured by a Medium-Wave Infra-Red (MWIR) camera. Again the objective is to perform and evaluate the performance of image fusion in a real-life application and sequence. In Figure 4, a typical frame from the sequence (frame 100) is depicted. The initial sequence required registration of the two streams that was achieved by employing MATLAB's Image Registration Toolbox and some correspondence points that were manually set in the visual and infrared images. In Figure
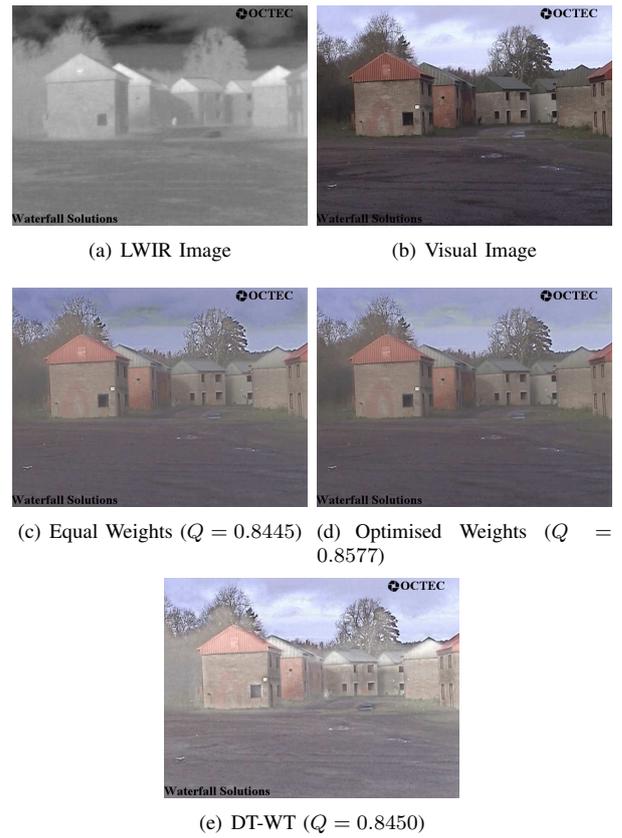

2Code for the Dual-Tree Wavelet Transform available online by the Polytechnic University of Brooklyn, NY at http://taco.poly.edu/WaveletSoftware/

(a) LWIR Image   (b) Visual Image

(c) Equal Weights ($Q = 0.8445$)   (d) Optimised Weights ($Q = 0.8577$)

(e) DT-WT ($Q = 0.8450$)

Figure 2. Optimal Contrast correction for the "Octet 1" dataset. The optimised means output featured enhanced performance compared to the other approaches.

4(a), the registered MWIR input is depicted and in Figure 4(b) the registered visual RGB input. As previously, the RGB image was transformed to the YUV representation where the Y-channel (luminance) was employed for fusion with the MWIR image. The "max-abs" fusion rule was used for all fusion efforts. The ICA-based framework using equal weights for the input images means is depicted in Figure 4(c). The proposed optimal contrast ICA-based framework converged smoothly to the optimal weight $q = [0.3219\ 0.6781]^T$ and produced fused image is depicted in Figure 4(d). The performance of the optimal scheme $Q = 0.7618$ is enhanced compared to that of the original framework $Q = 0.7589$, which verifies the optimal contrast procedure can enhance the fusion performance for image fusion of multimodal images. In Figure 4(e), the corresponding result of the DT-WT based fusion is shown, featuring an inferior performance of $Q = 0.7509$. To investigate the performance of the tested algorithms, the Piella Index was evaluated for all images of the "MX-15" dataset. In Table I, the average scores for the whole dataset of the three tested approaches are shown. The optimal contrast scheme tends to outperform both the original ICA-based scheme and the DT-WT based approach, serving as an effective enhancement of the original ICA framework.

A third example is used to demonstrate the efficiency of our

(a) LWIR sensor

(b) Visual sensor

(c) Equal Weights ($Q = 0.8533$)

(d) Optimised Weights ($Q = 0.8628$)

(e) DT-WT ($Q = 0.8561$)

Figure 3. The Octet 2 example of colour image fusion of a visual and a Long-Wave InfraRed sensor.

(a) InfraRed Image

(b) Visual Image

(c) Equal Weights ($Q = 0.7589$)

(d) Optimised Weights ($Q = 0.7618$)

(e) DT-WT ($Q = 0.7509$)

Figure 4. Optimal Contrast correction for the "MX15" dataset. The optimised means output featured enhanced performance compared to the traditional ICA framework and compared to the Dual-Tree Wavelet Transform (DTWT) scheme.

Table I
AVERAGE FUSION PERFORMANCE MEASUREMENTS USING PIELLA'S INDEX FOR OF THIS EXPERIMENTAL SECTION. THE PROPOSED ICA-BASED SCHEMES ARE COMPARED WITH THE ORIGINAL ICA-BASED FRAMEWORK AND THE DUAL-TREE WAVELET FRAMEWORK.

Method	ICA Equal Weights	ICA Opt. Weights	DT-WT
Octet1	0.8445	0.8577	0.8450
Octet2	0.8533	0.8628	0.8561
MX-15	0.7585	0.7669	0.7569
''Car''	0.6822	0.6857	0.6392

algorithm in the case of more than two sensors. We used some surveillance images from TNO Human Factors, provided by L. Toet [10], obtained from the Image Fusion Server [11]. The images are acquired by three kayaks approaching the viewing location from far away. As a result, their corresponding image size 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. The three input sensor images are depicted in Figures 5 (a), (b), (c). The proposed algorithm was initialised as previously and converged smoothly to the value of $q_{opt} = [0.6111\ 0.1289\ 0.2601]^T$. The Piella index for

the estimated weights is $Q_p = 0.6857$. In Figure 5 (d), (e), we plot the fused image assuming equal weights and optimised weights respectively. The Piella index for the equal weights image is $Q_p = 0.6822$, which implies that the proposed approach has achieved improved performance compared to the original scheme and the DT-WT scheme. The convergence in the three-dimensional case was similar to the one encountered in the previous examples, which implies that the fusion using ICA bases of more than two input sensors is possible and efficient.

VI. CONCLUSIONS

In this paper, the authors proposed an improvement to their previous ICA-based Image Fusion framework. In the original framework, the input sensor images are projected into localised patches of small size. The local mean value of the patches is subtracted and stored in order to reconstruct the local means of the fused image. Originally, an average of the stored means was used to reconstruct the fused image, nonetheless, it was demonstrated that this choice might not be optimal in several multi-modal cases in [4], [5]. In the same work an exhaustive search solution of the optimum performance in terms of the Piella index [6] was proposed for the case of two input sensors only. In this paper, the authors provide a generalised iterative

(a) HS IR sensor

(b) micro LW sensor

(c) Visual sensor

(d) Equal Weights
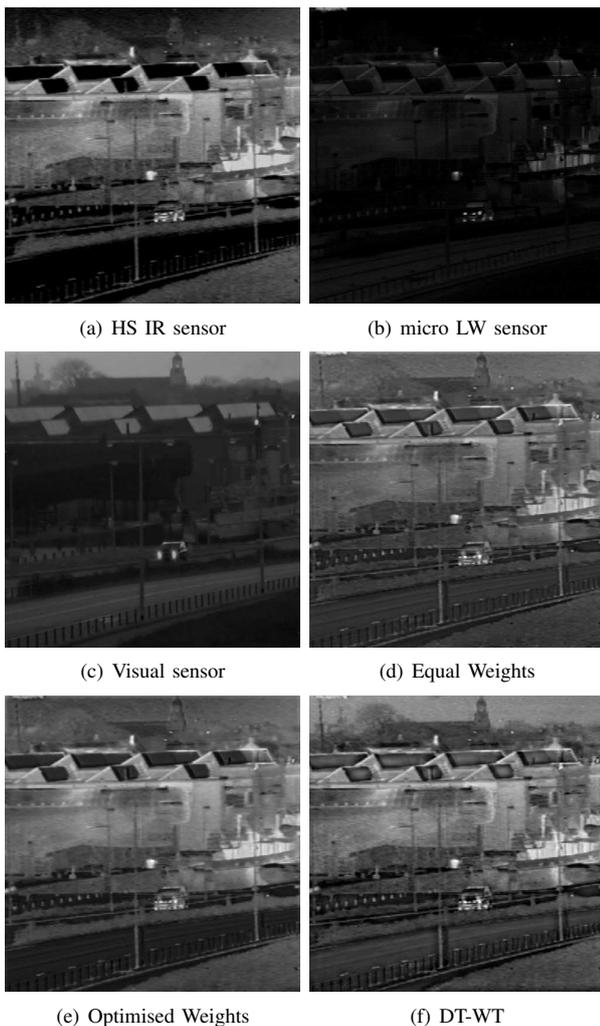

(e) Optimised Weights

(f) DT-WT

Figure 5. Optimal Contrast correction for the TNO Human Factors dataset. The optimal contrast selection scheme chooses to emphasize the contrast of the two Infrared images to the low-contrast night visual image.

solution of the identified problem by analytic optimisation of the Piella index in the general case of T input sensors. The proposed gradient-descent optimisation can identify the optimal value of contrast in maximum 70 iterations, providing an efficient and general solution for the case of T input sensors, removing the need of analytic evaluation of the whole solution space in [4], [5]. The proposed approach enhances the performance of the original ICA-based fusion framework, identifying the actual optimal of the fusion performance index.

ACKNOWLEDGMENT

REFERENCES

[1] A. Mahmood, P.M. Tudor, W. Oxford, R. Hansford, J.D.B. Nelson, N.G. Kingsbury, A. Katartzis, M. Petrou, N. Mitianoudis, T. Stathaki, A. Achim, D. Bull, N. Canagarajah, S. Nikolov, A. Loza, and N. Cvejic, "Applied multi-dimensional fusion," *The Computer Journal*, vol. 50, no. 6, pp. 660–673, 2007.

[2] P. Hill, N. Canagarajah, and D. Bull, "Image fusion using complex wavelets," in *Proc. 13th British Machine Vision Conference*, Cardiff, UK, 2002.

[3] N. Mitianoudis and T. Stathaki, "Pixel-based and Region-based image fusion schemes using ICA bases," *Information Fusion*, vol. 8, no. 2, pp. 131–142, 2007.

[4] N. Cvejic, D. Bull, and N. Canagarajah, "Region-based multimodal image fusion using ICA bases," *IEEE Sensors Journal*, vol. 7, no. 5, pp. 743–751, 2007.

[5] N. Cvejic, J. Lewis, D. Bull, and N. Canagarajah, "Adaptive region-based multimodal image fusion using ica bases," in *Proc. Int. Conf on Information Fusion*, 2006.

[6] G. Piella and H. Heijmans, "A new quality metric for image fusion," in *International Conference on Image Processing (ICIP)*, Barcelona, Spain, 2003, pp. 173–176.

[7] A. Hyvarinen, 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.

[8] Z. Wang and A.C. Bovik, "A universal image quality index," *IEEE Signal Processing Letters*, vol. 9, no. 3, pp. 81–84, 2002.

[9] L. Bogoni, M. Hansen, and P. Burt, "Image enhancement using pattern selective color image fusion," in *Proc. Int. Conf. on Image Analysis and Processing*, pp. 44–49, 1999.

[10] A. Toet, "Detection of dim point targets in cluttered maritime backgrounds through multisensor image fusion," *Targets and Backgrounds VIII: Characterization and Representation, Proceedings of SPIE*, vol. 4718, pp. 118–129, 2002.

[11] The Image fusion server, "http://www.imagefusion.org/".