Enhancement of Multiple Sensor Images using Joint Image Fusion and Blind Restoration

Nikolaos Mitianoudis, Tania Stathaki

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

Abstract

Image fusion systems aim at transferring "interesting" information from the input sensor images to the fused image. The common assumption for most fusion approaches is the existence of a high-quality reference image signal for all image parts in all input sensor images. In the case that there are common degraded areas in at least one of the input images, the fusion algorithms can not improve the information provided there, but simply convey a combination of this degraded information to the output. In this study, the authors propose a combined spatial-domain method of fusion and restoration in order to identify these common degraded areas in the fused image and use a regularised restoration approach to enhance the content in these areas. The proposed approach was tested on both multi-focus and multi-modal image sets and produced interesting results.

Key words: Spatial-domain Image Fusion, Image Restoration.
PACS:

1 Introduction

Data fusion is defined as the process of combining data from sensors and related information from several databases, so that the performance of the system can be improved, while the accuracy of the results can be also increased. Essentially, fusion is a procedure of incorporating essential information from several sensors to a composite result that will be more comprehensive and thus more useful for a human operator or other computer vision tasks.

Image fusion can be similarly viewed as the process of combining information in the form of images, obtained from various sources in order to construct an artificial image that contains all "useful" information that exists in the input images. Each image has been acquired using different sensor modalities

or capture techniques, and therefore, it has different features, such as type of degradation, thermal and visual characteristics. The main concept behind all image fusion algorithms is to *detect strong salient features* in the input sensor images and *fuse* these details to the synthetic image. The resulting synthetic image is usually referred to as the *fused image*.

Let $x_1(\mathbf{r}), \ldots, x_T(\mathbf{r})$ represent T images of size $M_1 \times M_2$ capturing the same scene, where $\mathbf{r} = (i, j)$ refers to pixel coordinates (i, j) in the image. Each image has been acquired using different sensors that are placed relatively close and are observing the same scene. Ideally, the images acquired by these sensors should be similar. However, there might exist some miscorrespondence between several points of the observed scene, due to the different sensor viewpoints. *Image registration* is the process of establishing point-by-point correspondence between a number of images, describing the same scene. In this study, the input images are assumed to have negligible registration problems or the transformation matrix between the sensors' viewpoints is known. Thus, the objects in all images can be considered geometrically aligned.

As already mentioned, the process of combining the important features from the original T images to form a single enhanced image $y(\mathbf{r})$ is usually referred to as *image fusion*. Fusion techniques can be divided into *spatial domain* and *transform domain* techniques [5]. 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:

$$y(\mathbf{r}) = g(x_1(\mathbf{r}), \ldots, x_T(\mathbf{r})) \tag{1}$$

Moving to a transform domain enables the use of a framework, where the image's salient features are more clearly depicted than in the spatial domain. Let $\mathcal{T}\{\cdot\}$ represent a transform operator and $g(\cdot)$ the applied fusion rule. Transform-domain fusion techniques can then be outlined, as follows:

$$y(\mathbf{r}) = \mathcal{T}^{-1}\{g(\mathcal{T}\{x_1(\mathbf{r})\}, \ldots, \mathcal{T}\{x_T(\mathbf{r})\})\} \tag{2}$$

Several transformations were proposed to be used for image fusion, including the *Dual-Tree Wavelet Transform* [5,7,12], *Pyramid Decomposition* [14] and image-trained Independent Component Analysis bases [10,9]. All these transformations project the input images onto localised bases, modelling sharp and abrupt transitions (edges) and therefore, describe the image using a more meaningful representation that can be used to detect and emphasize salient features, important for performing the task of image fusion. In essence, these transformations can discriminate between salient information (strong edges

2

and texture) and constant or non-textured background and can also evaluate the quality of the provided salient information. Consequently, one can select the required information from the input images in the transform domain to construct the "fused" image, following the criteria presented earlier on.

In the case of multi-focus image fusion scenarios, an alternative approach has been proposed in the spatial domain, exploiting current error estimation methods to identify high-quality edge information [6]. One can perform error minimization between the fused and input images, using various proposed error norms in the spatial domain in order to perform fusion. The possible benefit of a spatial-domain approach is the reduction in computational complexity, which is present in a transform-domain method due to the forward and inverse transformation step.

In addition, following a spatial-domain fusion framework, one can also benefit from current available spatial-domain image enhancement techniques to incorporate a possible restoration step to enhance areas that exhibit distorted information in all input images. Current fusion approaches can not enhance areas that appear degraded in any sense in all input images. There is a necessity for some pure information to exist for all parts of the image in the various input images, so that the fusion algorithm can produce a high quality output. In this work, we propose to reformulate and extend Jones and Vorontsov's [6] spatial-domain approach to fuse the non-degraded common parts of the sensor images. A novel approach is used to identify the areas of common degradation in all input sensor images. A double-regularised image restoration approach using robust functionals is applied on the estimated common degraded area to enhance the common degraded area in the "fused" image. The overall fusion result is superior to any traditional fusion approach since the proposed approach goes beyond the concept of transferring useful information to a thorough fusion-enhancement approach.

2 Robust Error Estimation Theory

Let the image $y(\mathbf{r})$ be a recovered version from a degraded observed image $x(\mathbf{r})$, where $\mathbf{r} = (i, j)$ are pixel coordinates (i, j). To estimate the recovered image $y(\mathbf{r})$, one can minimise an error functional $E(y)$ that expresses the difference between the original image and the estimated one, in terms of y. The error functional can be defined by:

$$E(y) = \int_{\Omega} \rho\left(\mathbf{r}, y(\mathbf{r}), |\nabla y(\mathbf{r})|\right) d\mathbf{r} \tag{3}$$

where Ω is the image support, $\nabla y(\mathbf{r})$ is the image gradient. The function $\rho(\cdot)$ is termed the *error norm* and is defined according to the application, i.e. the type of degradation or the desired task. For example, a least square error norm can be appropriate to remove additive Gaussian noise from a degraded image. The extremum of the previous equation can be estimated, using the *Euler - Lagrange* equation. The Euler-Lagrange equation is an equation satisfied by a function f of a parameter t which extremises the functional:

$$E(f) = \int F\left(t, f(t), f'(t)\right) dt \tag{4}$$

where F is a given function with continuous first partial derivatives. The Euler-Lagrange equation is described by the following ordinary differential equation, i.e. a relation that contains functions of only one independent variable, and one or more of its derivatives with respect to that variable, the solution t of which extremises the above functional [21].

$$\frac{\partial}{\partial f(t)} F\left(t, f(t), f'(t)\right) - \frac{d}{dt} \frac{\partial}{\partial f'(t)} F\left(t, f(t), f'(t)\right) = 0 \tag{5}$$

Applying the above rule to derive the extremum of (3), the following Euler-Lagrange equation is derived:

$$\frac{\partial \rho}{\partial y} - \nabla\left(\frac{\partial \rho}{\partial \nabla y}\right) = 0 \tag{6}$$

Since $\rho(\cdot)$ is a function of $|\nabla y|$ and not ∇y, we perform the substitution

$$\partial \nabla y = \partial |\nabla y| / \text{sgn}(\nabla y) = |\nabla y| \partial |\nabla y| / \nabla y \tag{7}$$

where $\text{sgn}(y) = y/|y|$. Consequently, the Euler-Lagrange equation is given by:

$$\frac{\partial \rho}{\partial y} - \nabla\left(\frac{1}{|\nabla y|} \frac{\partial \rho}{\partial |\nabla y|} \nabla y(\mathbf{r})\right) = 0 \tag{8}$$

To obtain a closed-form solution $y(\mathbf{r})$ from (8) is not straightforward. Hence, one can use numerical optimisation methods to estimate y. Gradient-descent optimisation can be applied to estimate $y(\mathbf{r})$ iteratively using the following update rule:

$$y(\mathbf{r}, t) \leftarrow y(\mathbf{r}, t-1) - \eta \frac{\partial y(\mathbf{r}, t)}{\partial t} \tag{9}$$

4

where t is the time evolution parameter, η is the optimisation step size and

$$\frac{\partial y(\mathbf{r}, t)}{\partial t} = -\frac{\partial \rho}{\partial y} + \nabla\Big(\frac{1}{|\nabla y|}\frac{\partial \rho}{\partial |\nabla y|}\nabla y(\mathbf{r}, t)\Big) \tag{10}$$

Starting with the initial condition $y(\mathbf{r}, 0) = x(\mathbf{r})$, the iteration of (10) continues until the minimisation criterion is satisfied, i.e. $|\partial y(\mathbf{r}, t)/\partial t| < \epsilon$, where ϵ is a small constant ($\epsilon \sim 0.0001$). In practice, only a finite number of iterations are performed to achieve visually satisfactory results [6]. The choice of the error norm $\rho(\cdot)$ in the Lagrange-Euler equation is the next topic of discussion.

2.1 Isotropic diffusion

As mentioned previously, one candidate error norm $\rho(\cdot)$ is the least-squares error norm. This norm is given by:

$$\rho(\mathbf{r}, |\nabla y(\mathbf{r})|) = \frac{1}{2}|\nabla y(\mathbf{r})|^2 \tag{11}$$

The above error norm smooths Gaussian noise and depends only on the image gradient $\nabla y(\mathbf{r})$, but not explicitly on the image $y(\mathbf{r})$ itself. If the least-squares error norm is substituted in the time evolution equation (10), we get the following update:

$$\frac{\partial y(\mathbf{r}, t)}{\partial t} = \nabla^2 y(\mathbf{r}, t) \tag{12}$$

which is the *isotropic diffusion* equation having the following analytic solution [2]:

$$y(\mathbf{r}, t) = G(\mathbf{r}, t) * x(\mathbf{r}) \tag{13}$$

where $*$ denotes the convolution of a Gaussian function $G(\mathbf{r}, t)$ of standard deviation t with $x(\mathbf{r})$, the initial data. The solution specifies that the time evolution in (12) is a convolution process performing Gaussian smoothing. However, as the time evolution iteration progresses, the function $y(\mathbf{r}, t)$ becomes the product of the convolution of the input image with a Gaussian of constantly increasing variance, which will finally produce a constant value. In addition, it has been shown that isotropic diffusion may not only smooth edges, but also causes drifts of the actual edges in the image edge, because of the Gaussian filtering (smoothing) [2,13]. These are two disadvantages that need to be seriously considered when using isotropic diffusion.

Image fusion aims at transferring salient features to the fused image. In this work and in most fusion systems, saliency is interpreted as edge information and therefore, image fusion aims at highlighting edges in the fused image. An additional desired property can be to smooth out any possible Gaussian noise. In order to achieve the above tasks using an error estimation framework, the objective is to create an error norm that will enhance edges in an image and simultaneously smooth possible noise. The following error norm, combining *isotropic smoothing* with *edge enhancement*, was proposed in [6]:

$$\rho(\mathbf{r}, y(\mathbf{r}, t), |\nabla y(\mathbf{r}, t)|) = \frac{\alpha}{2}|\nabla y(\mathbf{r}, t)|^2 + \frac{\beta}{2}J_x(\mathbf{r})(y(\mathbf{r}, t) - x(\mathbf{r}))^2 \qquad (14)$$

where α, β are constants that define the level of smoothing and edge enhancement respectively that is performed by the cost function, t is the time evolution and J_x is commonly termed the *anisotropic gain function*, which is a Gaussian smoothed edge map. One possible choice for implementing a Gaussian smoothed edge map is the following :

$$J_x(\mathbf{r}) = \kappa \int |\nabla x(\mathbf{q})|^2 G(\mathbf{r} - \mathbf{q}, \sigma)d^2\mathbf{q} \qquad (15)$$

where $G(\cdot)$ is a Gaussian function of zero-mean and standard deviation σ and κ is a constant. Another choice can be a smoothed Laplacian edge map. The anisotropic gain function has significantly higher values around edges or where sharp features are dominant compared to blurred or smooth regions.

Substituting the above error norm into the gradient descent update of (10) yields the following time evolution equation with anisotropic gain:

$$\frac{\partial y(\mathbf{r}, t)}{\partial t} = \alpha \nabla^2 y(\mathbf{r}, t) - \beta J_x(\mathbf{r})(y(\mathbf{r}, t) - x(\mathbf{r})) \qquad (16)$$

The above equation essentially smoothes noise while enhancing edges. The parameters α and β control the effects of each term. The parameter α controls the amount of noise smoothing in the image and β controls the anisotropic gain, i.e. the preservation and enhancement of the edges. For noiseless images, an evident choice is $\alpha = 0$ and $\beta = 1$. In this case, for short time intervals, the anisotropic gain function J_x induces significant changes dominantly around regions of sharp contrast, resulting in edge enhancement.

There is always a possibility that in some regions of interest, the anisotropic gain function is not high enough and therefore the above update rule can

potentially degrade the quality of information that is already integrated into the input image and consequently in the enhanced image. To prevent such erasing effects, however small might be, John and Vorontsov [6] introduced the following modified anisotropic gain function:

$$J(\mathbf{r}, t) = J_x(\mathbf{r}) - J_y(\mathbf{r}, t) \tag{17}$$

The general update formula to estimate $f(\mathbf{r})$ becomes then:

$$\frac{\partial y(\mathbf{r}, t)}{\partial t} = \alpha \nabla^2 y(\mathbf{r}, t) - \Theta(J(\mathbf{r}, t)) J(\mathbf{r}, t)(y(\mathbf{r}, t) - x(\mathbf{r})) \tag{18}$$

where

$$\Theta(J) = \begin{cases} 1 \, , J \geq 0 \\ 0 \, , J < 0 \end{cases} \tag{19}$$

The new term $\Theta(J)J$ allows only high quality information, interpreted in terms of edge presence, to transfer to the enhanced image. In the opposite case that $J_x(\mathbf{r}) < J_y(\mathbf{r})$, the information in the enhanced image has better edge representation than the original degraded image for several $\mathbf{r}$ and therefore, no processing is necessary. In the case of a single input image, the above concept might not seem practical. In the following section, the proposed concept is employed in a multiframe input scenario, where the aim is to transfer only high quality information to the enhanced image $y(\mathbf{r})$. In this case, this positive edge injection mechanism is absolutely vital to ensure information enhancement.

3 Fusion with Error Estimation theory

In this section, the authors propose a novel spatial-domain fusion algorithm, based on the basic formulation of John and Vorontsov. In [6], a sequential approach to image fusion based on Error Estimation theory was proposed. Assuming that we have a number of T input frames $x_n(\mathbf{r})$ to be fused, one can easily perform selective image fusion, by iterating the update rule (18) for the estimation of $y(\mathbf{r})$ using each of input images x_n consecutively for a number of K iterations. In a succession of intervals of K iterations, the synthetic frame finally integrates high-quality edge areas from the entire set of input frames.

The possibility of data fusion occurring in regions where the anisotropic gain function is not high enough, can potentially degrade quality information already integrated into the synthetic frame. To prevent such erasing effects, as

mentioned in the previous section, a differential anisotropic gain function can be introduced to transfer only high quality information to the fused image $y(\mathbf{r})$. The proposed approach by John and Vorontsov can be applied mainly in the case of a video stream, where the quality of the observed image is enhanced, based on previous and forthcoming frames. However, this framework is not efficient in the case of fusion applications, where the input frames are simultaneously available for processing and fusion. In this case, a reformulation of the above procedure is needed and is described in full in the following section.

3.1   A novel fusion formulation based on error estimation theory

Assume there are T images $x_n(\mathbf{r})$ that capture the same observed scene. The input images are assumed to be registered and each image contains exactly the same scene. This assumption is valid, since in most real-life applications, the input sensors are arranged in a close-distance array and similar zoom level in order to minimise the need for registration or the viewpoint transformation matrix is known. Different parts of the images are blurred using different amounts and types of blur. The objective is to combine the useful parts of input information to form a composite ("fused") image.

The described setup can model a possible out-of-focus scenario of image capture. We have all witnessed the case, where we want to take a photograph of an object in a scene and the camera focuses on a background point/object by mistake. As a result, the foreground object appears blurred in the final image, whereas the background texture is properly captured. In a second attempt to photograph the object correctly, the foreground object appears properly and the background appears blurred. Ideally, we would like to combine the two images into a new one, where everything would appear in full detail. This is an example of a real-life application for the fusion of out-of-focus images. The same scenario can also appear in military surveillance and general surveillance applications, where one would like to enhance the surveillance output, by combining multiple camera inputs at different focal length.

The fused image $y(\mathbf{r}, t)$ can be constructed as a linear combination of the T input registered images $x_n(\mathbf{r})$. The fusion problem is usually solved by finding the weights $w_n(\mathbf{r}, t)$ that transfer all the useful information from the input images x_n to the fused image y [10,9].

$$y(\mathbf{r}, t) = w_1(\mathbf{r}, t)x_1(\mathbf{r}) + \ldots + w_T(\mathbf{r}, t)x_T(\mathbf{r}) \tag{20}$$

where $w_n(\mathbf{r}, t)$ denotes the nth weight of the image x_n at position $\mathbf{r}$. To estimate these weights, we can perform error minimisation using the previously

mentioned approach of Isotropic Diffusion with edge enhancement. The problem is now to estimate the weights w_n simultaneously, so as to achieve edge preservation. This cannot be accomplished directly by the scheme proposed by Jones and Vorontsov.

In other words, we need to estimate the derivative $\partial w_n / \partial t$ simultaneously, for all $n = 1, \ldots, T$. We can associate $\partial w_n / \partial t$ with $\partial y / \partial t$ that has already been derived before.

$$\frac{\partial y}{\partial t} = \frac{\partial y}{\partial w_n} \frac{\partial w_n}{\partial t} = x_n \frac{\partial w_n}{\partial t} \tag{21}$$

Therefore, we can use the previous update rule to estimate the contribution of each image to the fused one:

$$\frac{\partial w_n(\mathbf{r}, t)}{\partial t} = \frac{1}{x_n(\mathbf{r})} \frac{\partial y(\mathbf{r}, t)}{\partial t} \tag{22}$$

The fusion weight $w_n(\mathbf{r}, t)$ of each input image can then be estimated using sequential minimisation with the following update rule $\forall n = 1, \ldots, T$:

$$w_n(\mathbf{r}, t + 1) \leftarrow w_n(\mathbf{r}, t) - \eta \frac{\partial w_n(\mathbf{r}, t)}{\partial t} \tag{23}$$

where

$$\frac{\partial w_n(\mathbf{r}, t)}{\partial t} = -\frac{1}{x_n(\mathbf{r})} \Theta(J_n(\mathbf{r}, t)) J_n(\mathbf{r}, t)(y(\mathbf{r}, t) - x_n(\mathbf{r})) \tag{24}$$

and $J_n(\mathbf{r}, t) = J_{x_n}(\mathbf{r}) - J_y(\mathbf{r}, t)$. To avoid possible numerical instabilities, for those $\mathbf{r}$ that $x_n(\mathbf{r}) = 0$, a small constant is added to these elements so as to become nonzero. All weights are initialised to $w_n(\mathbf{r}, t) = 1/T$, which represents the "mean" fusion rule. As this scheme progresses over time, the weights are adapting and tend to emphasise more the useful details that exist in each image and suppress the information that is not very accurate. In addition, all the fusion weights are estimated simultaneously using this scheme. Therefore, after a couple of iterations the majority of the useful information is extracted from the input images and transferred to the composite image.

3.2   Fusion experiments of out-of-focus and multimodal image sets using error estimation theory

In this section, we perform several fusion experiments of both out-of-focus and multimodal images to evaluate the performance of the proposed approach.

9

	Petrovic	Piella
TopoICA	0.6151	0.9130
Fusion with EE	0.6469	0.9167

Table 1
Performance evaluation of the Diffusion approach and the TopoICA-based fusion approach using Petrovic [18] and Piella's [15] metrics.

Most test images were taken from the Image Fusion server [3]. The numerical evaluation in most experiments was performed using the indexes proposed by Piella [15] and Petrovic [18].

In the first experiment, the system is tested with an out-of-focus example, the "Disk" dataset. The ICA-based fusion algorithm, proposed in [10], was employed as a benchmark to the new proposed algorithm. We used 40 TopoICA 8×8 bases, trained from 10000 patches that were randomly selected from natural images. Then, the "Weighted Combination" rule was selected to perform fusion of the input images. On the other hand, for the spatial-domain fusion scheme, the parameters were set to $\alpha = 0$ (no visible noise), $\beta = 0.8$ and the learning parameter was set to $\eta = 0.08$. The Gaussian smoothed edge map of (15) was calculated by extracting an edge map using the Sobel mask, which was subsequently smoothed by a Gaussian 5×5 kernel of standard deviation $\sigma = 1$. The fusion results of the two methods are depicted in Figure 1. We notice that the proposed approach produces sharper edges compared to the ICA-Based method. The difference is more visible around the edges of the tilted books in the bookcase and the eye on the cover of the book that is in front of the bookcase. In Figure 2, the convergence rate of the estimation of one of the fusion weights is shown. The proposed algorithm demonstrates almost linear convergence, which is expected for a gradient algorithm.

In Table 1, the performance of the proposed method is compared with the ICA-based method, in terms of the Petrovic and Piella method. The metrics give slightly higher performance to the proposed methodology. However, we can observe an improvement in the visual representation of edges using the proposed method in the particular application of fusion of out-of-focus images.

The estimated fusion weights $w_1(\mathbf{r}), w_2(\mathbf{r})$ are depicted in Figure 3. It is clear that the weights w_1, w_2 highlight the position of high-quality information in the input images. The cost function that is optimised in this case aims at highlighting edges in the "fused" image. This is essentially what is estimated by the weight maps $w_1(\mathbf{r}), w_2(\mathbf{r})$. This information can be used to identify common areas of inaccurate information in the input images. A restoration algorithm could be applied to these areas and enhance the final information that is conveyed to the "fused" image.

The next step is to apply the proposed algorithm to a multimodal scenario.

(a) Input Image 1   (b) Input Image 2

(c) TopoICA Fusion   (d) Proposed Scheme
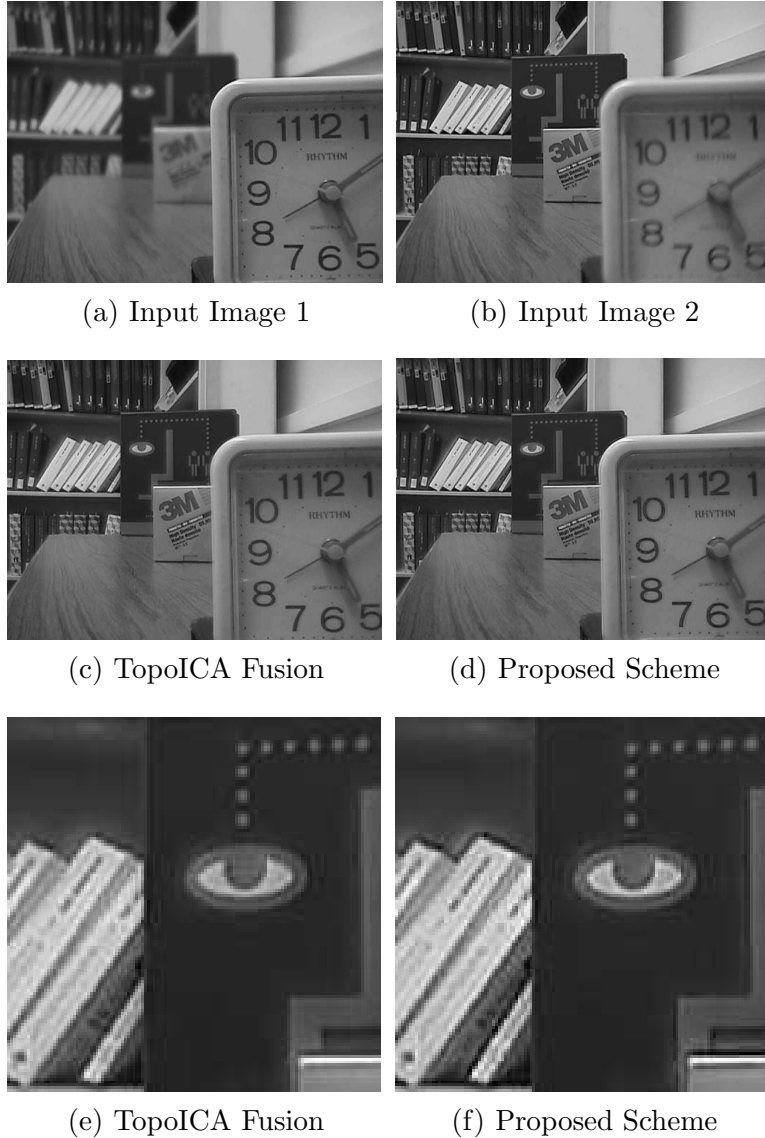
(e) TopoICA Fusion   (f) Proposed Scheme

Fig. 1. An out-of-focus fusion example using the "Disk" dataset available by the Image Fusion server [3]. We compare the TopoICA-based fusion approach and the proposed Diffusion scheme.

We will use an image pair from the "Dune" dataset of surveillance images from TNO Human Factors, provided by L. Toet [16] in the ImageFusion Server [3]. We applied the TopoICA-based approach [10] using the "maxabs" fusion rule and the proposed algorithm on the dataset, using the same settings as in the previous example. In Figure 4, we plot the fused results of the two methods and in Table 2, we plot their numerical evaluation using Petrovic and Piella's indexes.

According to the performance evaluation indexes, the ICA-based approach performs considerably better than the proposed approach. The same trend is also observed in the metrics. However, the proposed approach performs
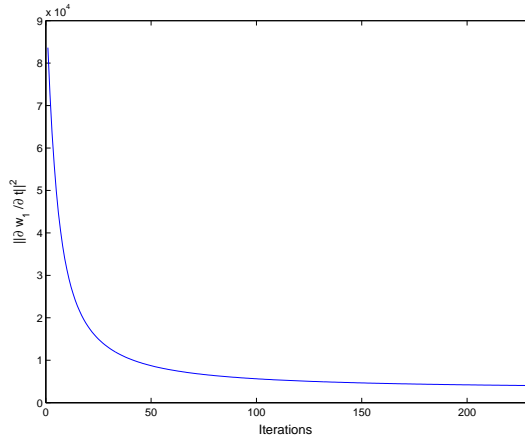
Fig. 2. Convergence of the estimated fusion weight w_1 using the proposed fusion algorithm in terms of $||\partial w_1/\partial t||^2$.
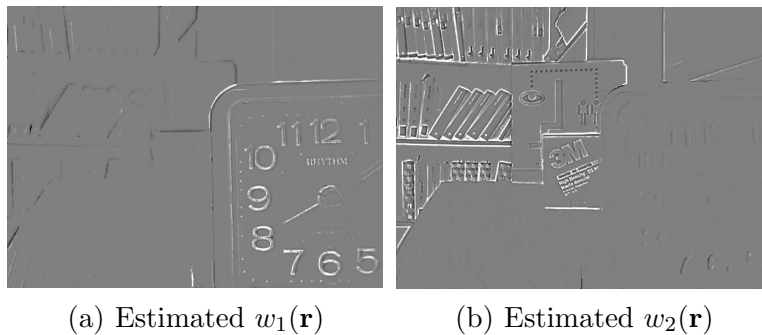

(a) Estimated $w_1(\mathbf{r})$ (b) Estimated $w_2(\mathbf{r})$

Fig. 3. The weights w_1, w_2 highlight the position of high quality information in the input images.

	Petrovic	Piella
TopoICA	0.4921	0.7540
Fusion with EE	0.4842	0.6764

Table 2
Performance evaluation in the case of a multimodal example from the Toet database. The TopoICA-based approach is compared with the proposed fusion approach.

differently to a common fusion approach. It aims at highlighting the edges of the input images to the fused image, due to the edge enhancement term in the cost function. This is can be observed directly in Figure 4(d). All edges and texture areas are highly enhanced in the fused image together with the outline of the important target, i.e. the hidden man in the middle of the picture. Consequently, one should also consult the human operators of modern fusion systems, apart from proposed fusion metrics [15,18], in order to evaluate efficiently the performance of these algorithms. Perhaps the outlined fusion result is more appealing to human operators and the human vision system in general and therefore may be also be examined as a preferred solution.

12

(a) Input Image 1   (b) Input Image 2

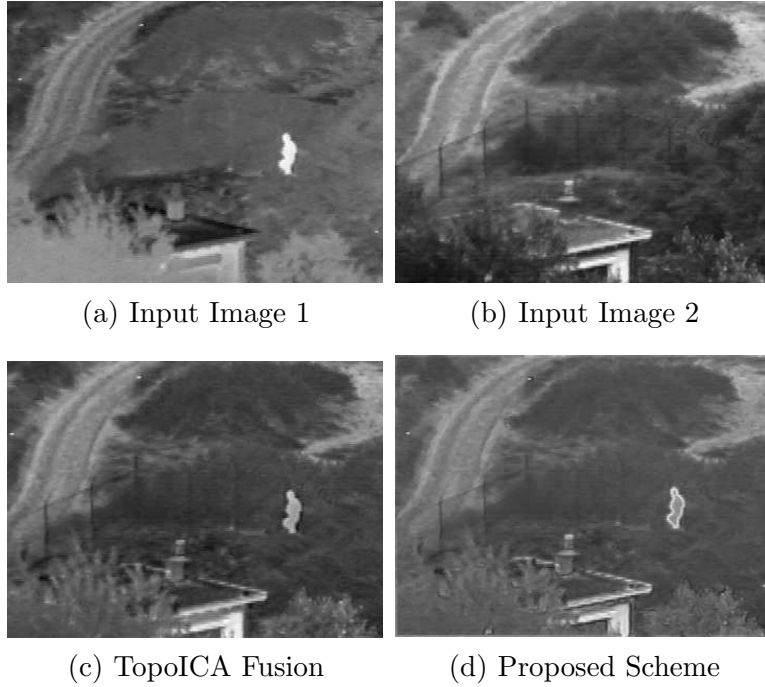(c) TopoICA Fusion   (d) Proposed Scheme

Fig. 4. Comparison of a multimodal fusion example using the TopoICA method and the Diffusion approach. Even though the metrics demonstrate worse performance, the diffusion approach highlights edges giving a sharper fused image.

4 Joint Image Fusion and Restoration

The basic Image Fusion concept assumes that there is some useful information for all parts of the observed scene at least in one of the input sensors. However, this assumption might not always be true. This means that there might be parts of the observed scene where there is only degraded information available. The current fusion algorithms will fuse all high quality information from the input sensors and for the common degraded areas will form a blurry mixture of the input images, as there is no high quality information available.

In the following section, the problem of identifying the areas of common degraded information in all input images is addressed. A mechanism is established for identifying common degraded areas in an image. Once this part is identified, an image restoration approach can be applied as a second step in order to enhance these parts for the final composite "fused" image.

4.1 Identifying common degraded areas in the sensor images

The first task will be to identify the areas of degraded information in the input sensor images. An identification approach, based on local image statistics, will

13

be pursued to trace the degraded areas.

The "fused" image will be employed, as it emerges from the fusion algorithm. As mentioned earlier, the fusion algorithm will attempt to merge the areas of high detail to the fused image, whereas for the areas of degraded information, i.e. areas of weak edges or texture in all input images, will not impose any preference to any of the input images and therefore the estimated fusion weights will remain approximately equal to the initial weights $w_i = 1/T$. Consequently, the areas of out-of-focus distortion will be described by areas of low edge information in the fused image. Equivalently, some areas of very low texture or constant background also need to be excluded, since there is no benefit in restoring them. These areas can be traced, by evaluating the local standard deviation of an edge information metric in small local neighbourhoods around each pixel. The following algorithm for extracting common degraded areas is described in the following steps:

(1) Extract an edge map of the fused image f, using the Laplacian kernel, i.e. $\nabla^2 f(\mathbf{r}, t)$.
(2) Find the local standard deviations $V_L(\mathbf{r}, t)$ for each pixel of the Laplacian edge map $\nabla^2 f(\mathbf{r}, t)$, using 5×5 local neighbourhoods.
(3) Reduce the dynamic range by calculating $\ln(V_L(\mathbf{r}, t))$.
(4) Estimate $V_{sL}(\mathbf{r}, t)$, by smoothing $\ln(V_L(\mathbf{r}, t))$ using a 15×15 median filter.
(5) Create the common degraded area map $A(\mathbf{r})$ by thresholding $V_{sL}(\mathbf{r}, t)$. The mask $A(\mathbf{r})$ is set to 1, for those $\mathbf{r}$ that $q \min_r(V_{sL}(\mathbf{r}, t)) < V_{sL}(\mathbf{r}, t) < p\,\mathrm{mean}_r(V_{sL}(\mathbf{r}, t))$, otherwise is set to zero.

Essentially, we create an edge map, as described by the Laplacian kernel. The Laplacian kernel was chosen because it was already estimated during the fusion stage of the framework. The next step is to find the local activity in 5×5 neighbourhoods around each pixel in the edge map. A metric of local activity is given by the local standard deviation. A pixel of high local activity should be part of an "interesting" detail in the image (edge, strong texture etc), whereas a point of low local activity might be a constant background or weak texture pixel. We can devise a heuristic thresholding scheme in order to identify these areas of weak local activity, i.e. possible degraded areas in all input images for fusion. The next step is to reduce the dynamic range of these measurements, using a logarithmic nonlinear mapping, such as $\ln(\cdot)$. To smooth out isolated pixels and connect similar areas, we perform median filtering of the log-variance map. Consequently, the common degraded area map is created by thresholding the values of the log-variance map with a heuristic threshold set to $q \min_r(V_{sL}(\mathbf{r}, t)) < V_{sL}(\mathbf{r}, t) < p\,\mathrm{mean}_r(V_{sL}(\mathbf{r}, t))$, where p, q are constants. The aim is to avoid high quality edge/texture and constant background information. The level of detail along with the level of constant background differ for different images. In order to identify the common degraded area with accuracy, the parameters p, q need to be defined manually for

(a) Input Image 1   (b) Input Image 2

(c) Fusion Scheme   (d) $V_{sL}(\mathbf{r})$ (e) Degraded area map
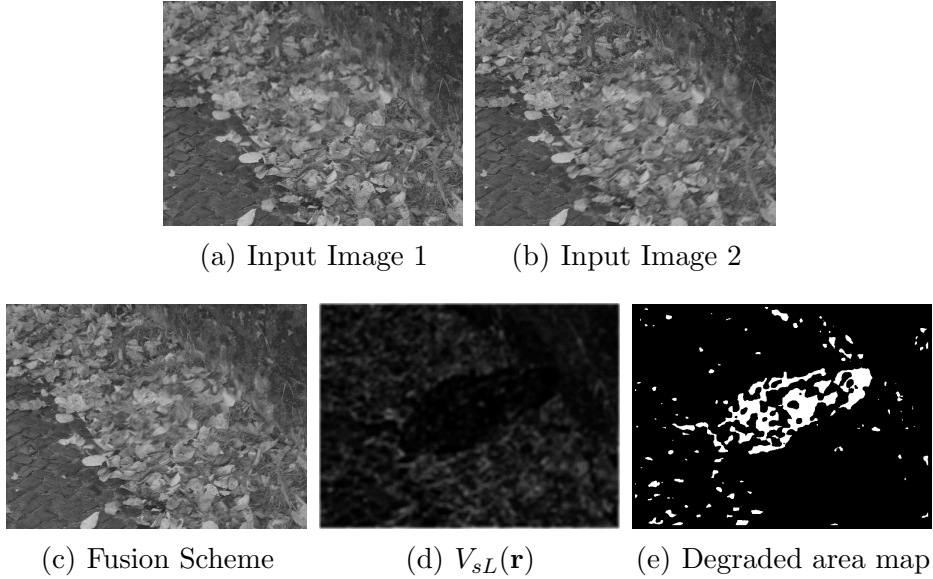
Fig. 5. If there exist blurry parts in all input images, common Image Fusion algorithms cannot enhance these parts, but will simply transfer the degraded information to the fused image. However, this area of degraded information is still identifiable

each image. The parameter q defines the level of background information that needs to be removed. In a highly active image, q is usually set to 1, however, other values have to be considered for images with large constant background areas. The parameter p is the upper bound threshold to discriminate between strong edges and weak edges, possibly belonging to a common degraded area. Setting p around the mean edge activity, we can find a proper threshold for the proposed system. Values that were found to work well in experiments were $q \in [0.98, 1]$ and $p \in [1, 1.1]$. Some examples of common degraded area identification using the above technique are shown in Figures 5, 6.

4.2 Image restoration

A number of different approaches for tackling the image restoration problem have been proposed in the literature, based on various principles. For an overview of image restoration methods, one can always possibly refer to Kundur and Hatzinakos [8] and Andrews and Hunt [1]. In this study, the double-weighted regularised image restoration approach in the spatial domain is pursued, that was initially proposed by You and Kaveh [19], with additional robust functionals to improve the performance in the case of outliers. The restoration problem is described by the following model:

$$y(\mathbf{r}) = h(\mathbf{r}) * f(\mathbf{r}) + d(\mathbf{r}) \tag{25}$$

(a) Input Image 1   (b) Input Image 2

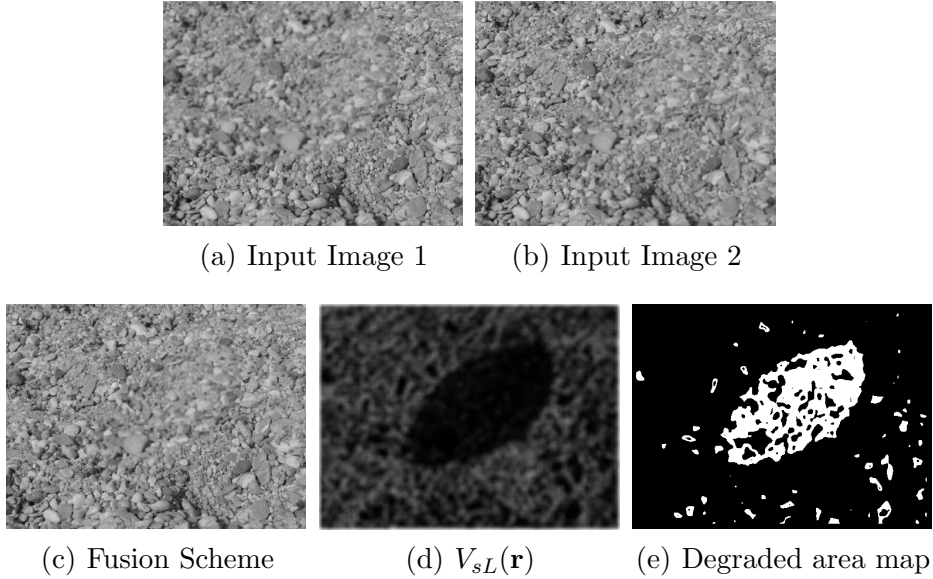(c) Fusion Scheme   (d) $V_{sL}(\mathbf{r})$ (e) Degraded area map

Fig. 6. Another example of degraded area identification in "fused" images.

where $*$ denotes 2D convolution, $h(\mathbf{r})$ the degradation kernel, $f(\mathbf{r})$ the estimated image and $d(\mathbf{r})$ possible additive noise.

4.2.1 Double weighted regularised image restoration

Conventional double weighted regularization for blind image restoration [1] estimates the original image by minimizing the cost function $Q(h(\mathbf{r}), f(\mathbf{r}))$ of the following quadratic form:

$$Q(h(\mathbf{r}), f(\mathbf{r})) = \underbrace{\frac{1}{2}||A_1(\mathbf{r})\left(y(\mathbf{r}) - h(\mathbf{r}) * f(\mathbf{r})\right)||^2}_{\text{residual}}$$

$$+ \underbrace{\frac{\lambda}{2}||A_2(\mathbf{r})\left(C_f * f(\mathbf{r})\right)||^2}_{\text{image regularisation}}$$

$$+ \underbrace{\frac{\gamma}{2}||A_3(\mathbf{r})\left(C_h * h(\mathbf{r})\right)||^2}_{\text{blur regularisation}} \qquad (26)$$

where $||\cdot||$ represents the $\mathcal{L}_2$-norm. The above cost function has three distinct terms. The residual term, the first term on the right-hand side of (26), represents the accuracy of the restoration process. This term is similar to a second-order error-norm (least-squares estimation), as described in a previous paragraph. The second term, called the regularising term, imposes a smoothness constraint on the recovered image and the third term acts similarly to the estimated blur. Additional constraints must be imposed, including the *non-*

16

negativity and *finite-support constraint* for both the blurring kernel and the image. Besides, the blurring kernel must always preserve the energy, i.e. all the coefficients should sum to 1. The regularization operators C_f and C_h are high-pass Laplacian operators applied on the image and the PSF respectively. The functions A_1, A_2 and A_3 represent spatial weights for each optimisation term. The parameters λ and γ control the trade-off between the residual term and the corresponding regularising terms for the image and the blurring kernel.

One can derive the same cost function through a Bayesian framework of estimating $f(\mathbf{r})$ and $h(\mathbf{r})$. To illustrate this connection, we assume that the blurring kernel $h(\mathbf{r})$ is known and the aim is to recover $f(\mathbf{r})$. A Maximum-A-Posteriori (MAP) estimate of $f(\mathbf{r})$ is given by performing $\max_f \log p(y, f|\mathbf{r}) = \max_f \log p(y|f, \mathbf{r}) p(f|\mathbf{r})$, where $\mathbf{r}$ denotes the observed samples. Assuming Gaussian noise for $d(\mathbf{r})$, we have that $p(y|f, \mathbf{r}) \propto \exp\left(-0.5a||y(\mathbf{r}) - h(\mathbf{r}) * f(\mathbf{r})||^2\right)$. Assuming smoothness for the image profile, one can employ the image prior $p(f|\mathbf{r}) \propto \exp\left(-0.5b||C_f * f(\mathbf{r})||^2\right)$, which has been widely used by the engineering community [11] in setting constraints on first or second differences, i.e. restricting the rate of changes in an image (a, b are constants that can determine the shape of the prior). Using the proposed models, one can derive a MAP estimate by optimising a function that is the same as the first two terms of (26), illustrating the connection between the two approaches.

To estimate $f(\mathbf{r})$ and $h(\mathbf{r})$, the above cost function needs to be minimised. Since each term of the cost function is quadratic, it can simply be optimized by applying alternating Gradient Descent optimisation [1]. This implies that the estimates for the image and the PSF can be estimated alternatively, using the gradients of the cost function with respect to $f(\mathbf{r})$ and $h(\mathbf{r})$. More specifically, the double iterative scheme can be expressed, as follows:

- At each iteration, update:

$$f(t+1) = f(t) - \eta_1 \frac{\partial Q(h(t), f(t))}{\partial f(t)} \tag{27}$$

$$h(t+1) = h(t) - \eta_2 \frac{\partial Q(h(t), f(t+1))}{\partial h(t)} \tag{28}$$

- Stop, if f and h converge.

The terms η_1 and η_2 are the step size parameters that control the convergence rates for the image and *Point Spread Function* (PSF) (blurring kernel) respectively. After setting the initial estimate of the image as the degraded image, and the PSF as a random mask, the cost function is differentiated with respect to the image first, while the PSF is kept constant, and vice versa. The required derivatives of the cost function are presented below:

$$\frac{\partial Q(h,f)}{\partial f} = -A_1(\mathbf{r})h(-\mathbf{r}) * (y(\mathbf{r}) - h(\mathbf{r}) * f(\mathbf{r}))$$
$$+ \lambda \left(A_2(\mathbf{r})C_f^T * (C_f * f(\mathbf{r})) \right) \tag{29}$$

$$\frac{\partial Q(h,f)}{\partial h} = -A_1(\mathbf{r})f(-\mathbf{r}) * (y(\mathbf{r}) - h(\mathbf{r}) * f(\mathbf{r}))$$
$$+ \gamma \left(A_3(\mathbf{r})C_h^T * (C_h * h(\mathbf{r})) \right) \tag{30}$$

where the superscript T denotes the transpose operation. Substituting (29) and (30) into (27) and (28) yields the final form of the algorithm (27) and (28), where the corresponding functions are iterated until convergence.

4.2.2   Robust functionals to the restoration cost function

There exist several criticisms regarding the conventional double regularisation restoration approach. One is the non-robustness of the least squares estimators employed in the traditional residual term, once the assumption of Gaussian noise does not hold [20]. Moreover, the quadratic regularising term penalises sharp gray-level transitions, due to the linearity of the derivative of the quadratic function. This implies that sudden changes in the image are filtered, and thus, the image edges are blurred. To alleviate this problem, we can introduce *robust functionals* in the cost function, in order to rectify some of the problems of this estimator. Therefore, the original cost function becomes:

$$Q(h(\mathbf{r}), f(\mathbf{r})) = \frac{1}{2}||A_1(\mathbf{r})\rho_n(y(\mathbf{r}) - h(\mathbf{r}) * f(\mathbf{r}))||^2$$
$$+ \frac{\lambda}{2}||A_2(\mathbf{r})\rho_f(C_f * f(\mathbf{r}))||^2$$
$$+ \frac{\gamma}{2}||A_3(\mathbf{r})\rho_d(C_h * h(\mathbf{r}))||^2 \tag{31}$$

Three distinct robust kernels $\rho_n(\cdot)$, $\rho_f(\cdot)$ and $\rho_d(\cdot)$ are introduced in the new cost function and are referred to as the robust residual and regularizing terms respectively. The partial derivatives of the cost function take the following form:

$$\frac{\partial Q(h,f)}{\partial f} = -A_1(\mathbf{r})h(-\mathbf{r}) * \rho_n'(y(\mathbf{r}) - h(\mathbf{r}) * f(\mathbf{r}))$$
$$+ \lambda \left(A_2(\mathbf{r})C_f^T * \rho_f'(C_f * f(\mathbf{r})) \right) \tag{32}$$

$$\frac{\partial Q(h, f)}{\partial h} = -A_1(\mathbf{r})f(-\mathbf{r}) * \rho_n' \left(y(\mathbf{r}) - h(\mathbf{r}) * f(\mathbf{r}) \right)$$
$$+ \gamma \left(A_3(\mathbf{r})C_h^T * \rho_d' \left(C_h * h(\mathbf{r}) \right) \right) \tag{33}$$

Robust estimation is usually presented in terms of the influence function $l(r) = \partial \rho / \partial r$. The influence function characterises the bias of a particular measurement on the solution. Traditional least squares kernels fail to eliminate the effect of outliers, with linearly increasing and non-bounded influence functions. On the other hand, they also tend to over-smooth the image's details, since such edge discontinuities lead to large values of smoothness error. Thus, two different kernel types are investigated, in order to increase the robustness and reject outliers in the context of the blind estimation.

To suppress the effect of extreme noisy samples ("outliers") that might be present in the observations, the derivative of an ideal robust residual term should increase less rapidly than a quadratic term in the case of outliers. One candidate function can be the following:

$$\rho_n'(x) = \frac{1}{1 + (\frac{x}{\theta})^{2\upsilon}} \tag{34}$$

Obviously, the specific function associated with the residual term assists in suppressing the effect of large noise values in the estimation process, by setting the corresponding influence function to small values. Optimal values for the θ and υ parameters have been investigated in [4]. These parameters determine the "shape" of the influence function and as a consequence the filtering of outliers.

In order to find a trade-off between noise elimination and preservation of high-frequency details, the influence functional for the image regularising term must approximate the quadratic structure at small to moderate values and alternatively deviate from the quadratic structure at high values, so that the sharp changes will not be greatly penalised. One possible formulation of the image regularising term is expressed by the absolute entropy function shown below, which reduces the relative penalty ratio between large and small signal deviations, compared with the quadratic function [20]. Hence, the absolute entropy function produces sharper boundaries than the quadratic one, and therefore can be employed for blind restoration.

$$\rho_f(x) = (|x| + e^{-1})\ln(|x| + e^{-1}) \tag{35}$$

$$\rho_f'(x) = \frac{1}{2}\mathrm{sgn}(x)\left(\ln(|x| + e^{-1}) + 1\right) \tag{36}$$

For simplicity, the robust functional for the stabilising term of the Point Spread

Function (PSF) is kept the same as the image regularising term ($\rho'_d(x) = \rho'_f(x)$). The actual PSF size can still be estimated at a satisfactory level. The PSF support is initially set to a large enough value. The boundaries of the assumed PSF support are trimmed at each iteration in a fashion which is described later, until it reduces to a PSF support that approximates the true support. [19].

4.3 Combining image fusion and restoration

In this section, we propose an algorithm that can combine all the previous methodologies and essentially perform fusion of all the parts that contain valid information in at least one of the input images and restoration of those image parts that are found to be degraded in all input images.

The proposed methodology consists of splitting the procedure in several individual parts:

(1) The first step is to use the proposed fusion update algorithm of section 3.1 to estimate the fused image $y(\mathbf{r})$. In this step, all useful information from the input images has been transferred to the fused image and the next step is to identify and restore the areas where only low quality information is available. In other words, this step ensures that all high quality information from the input images has been transferred to the fused image. The result of this step is the fused image $y(\mathbf{r})$.

(2) The second step is to estimate the common degraded area, using the previous methodology based on the Laplacian edge map of the fused image $y(\mathbf{r})$. More specifically, this step aims at identifying possible corrupted areas in all input images that need enhancement in order to highlight more image details that were not previously available. This will produce the common degraded area mask $A(\mathbf{r})$.

(3) The third step is to estimate the blur $h(\mathbf{r}, t)$ and the enhanced image $f(\mathbf{r}, t)$, using the estimated mask of the Common Degraded area as $A(\mathbf{r})$ and the produced fused image $y(\mathbf{r})$. This step is essentially enhancing only the common degraded area and not the parts of the image that have been identified to contain high quality information. The restoration is performed as described in the previous section, however, the updates for $f(\mathbf{r}, t)$ and $h(\mathbf{r}, t)$ are influenced only by the common degraded area. More specifically, the update for the enhanced image of (27) becomes

$$f(\mathbf{r}, t+1) = f(\mathbf{r}, t) - \eta_1 A(\mathbf{r}) \frac{\partial Q(h(\mathbf{r}, t), f(\mathbf{r}, t))}{\partial f(\mathbf{r}, t)} \tag{37}$$

In a similar manner the update for the Point Spread Function (PSF) needs to be influenced only by the common degraded area, i.e. in (33)

$f(\mathbf{r})$ is always substituted by $A(\mathbf{r})f(\mathbf{r})$.

4.4   Examples of joint image fusion and restoration

In this section, three synthetic examples are constructed to test the performance of the joint fusion and restoration approach. The proposed joint approach is compared to the performance of the Error-Estimation based fusion and the previously proposed ICA-based Image fusion approach. Three natural images are employed and two blurred sets were created from each of these images. These image sets are created so that: i) a different type/amount of blur is used in the individual images, ii) there is an area that is blurred in both input images, iii) there is an area that is not blurred in any of the input images. We have to note that in this case, the ground truth image needs to be available, to evaluate these experiments efficiently. The enhanced images will be compared with the ground truth image, in terms of Peak Signal-to-Noise Ratio (PSNR) and Image Quality Index Q_0, as proposed by Wang and Bovik [17]. In these experiments, the fusion indexes proposed by Petrovic and Xydeas [18] and Piella [15], cannot be used since they measure the amount of information that has been transferred from the input images to the fused image. Since the proposed fusion-restoration approach aims at enhancing the areas that have low quality information in the input images, it makes no sense to use any evaluation approach that employs the input images as a comparison standard. The images used in this experimental section can be downloaded [1] or requested by email from the authors.

There were several parameters that were manually set in the proposed fusion-restoration approach. For the Fusion part, we set $\alpha = 0$ (noise free examples 1-2) or $\alpha = 0.08$ (noisy example 3), $\beta = 0.8$, the learning parameter was set to $\eta = 0.08$. The Gaussian smoothed edge map of (15) was again calculated by extracting an edge map using the Sobel mask, which was subsequently smoothed by a Gaussian 5×5 kernel of standard deviation $\sigma = 1$. For the common degraded area identification step, a separate set of values for p, q will be given for each experiment. For the restoration step, we followed the basic guidelines proposed by You and Kaveh [19]. Hence, the regularisation matrices C_f, C_h were set, as follows:

$$C_f = \begin{bmatrix} 0 & -0.25 & 0 \\ -0.25 & 1 & -0.25 \\ 0 & -0.25 & 0 \end{bmatrix}, C_h = \begin{bmatrix} 2 & -1 \\ -1 & 0 \end{bmatrix} \tag{38}$$

---

[1]   http://www.commsp.ee.ic.ac.uk/~nikolao/Fusion_Restoration.zip

Some parameters were fixed to $\lambda = 0.1$, $\gamma = 10$, $\eta_1 = 0.25$, $\eta_2 = 0.00001$. The functions $A_1(\mathbf{r})$ and $A_3(\mathbf{r})$ were fixed to 1, whereas $A_2(\mathbf{r})$ was adaptively estimated for each iteration step, to emphasize regularisation on low-detail areas according to local variance (as described in [19]). For the robust functionals, we set $v = 2$ and $\theta \in [1.5, 3]$ was set accordingly for each case. The estimate kernel $h(\mathbf{r})$ was always initialised to $1/L^2$, where $L \times L$ is its size. All elements of the kernel were forced to be positive along the adaptation and sum to 1, so that the kernel does not perform any energy change. This is achieved by performing the mapping $h(\mathbf{r}) \leftarrow |h(\mathbf{r})|/\sum_{\mathbf{r}} |h(\mathbf{r})|$. The size L was usually set in advance, according to the experiment. If we need to estimate the size of the kernel automatically, we can assume initially a "large" size of kernel L. There is a mechanism to reduce the effective size of the kernel along the adaptation. The variance (energy) of a smaller $(L-1) \times (L-1)$ kernel is always compared to the variance (energy) of the $L \times L$ kernel. In the case that the smaller kernel captures more than 85% of the total kernel variance, its size becomes the new estimated kernel size in the next step of the adaptation. For the ICA-based method, the settings described in Section 3.2 were used.

In Figure 7, the first example with the "leaves" dataset is depicted. The two artificially created blurred input images are depicted in Figures 7 (a), (b). In Figure 7(a), Gaussian blur is applied on the upper left part of the image and in Figure 7(b) motion blur is applied on the bottom right part of the image. The amount of blur is randomly chosen. It is obvious that the two input images contain several areas of common degradation in the image centre and several areas that were not degraded at the bottom left and the top right of the image. In Figure 7(c), the result of the fusion approach using Isotropic Diffusion is depicted. As expected, the fusion algorithm manages to transfer all high quality information to the fused image, however, one area in the centre of the image still remains blurred since there is no high quality reference in any of the input images. Therefore, the output remains blurred in the fused image in the common degraded area. The common degraded area can be identified by the algorithm as depicted in the previously illustrated Figure 5 (e), using $p = 1.07$ and $q = 1$. In Figure 7 (d), we can see the final enhanced image, after the restoration process has been applied on the common degraded area for $L = 5$. An overall enhancement to the whole image quality can be witnessed with a significant edge enhancement compared to the original fused image. In Figure 7 (e), (f), a focus on the common degraded area in the fused and the fused/restored image can verify the above conclusions. In Figure 8, we plot the convergence of the restoration part of the common degraded area, in terms of the update for the restored image $f(\mathbf{r})$ and the update for the estimated blurring kernel $h(\mathbf{r})$. In addition, the estimated kernel is also depicted in Figure 8. The estimated kernel follows our intuition of a motion blur kernel around 20°, blurred by a Gaussian kernel. In Table 3, the performance of the TopoICA-based fusion scheme, the fusion scheme based on Error Estimation and the fusion+restoration scheme are evaluated in terms of Peak Signal-to-
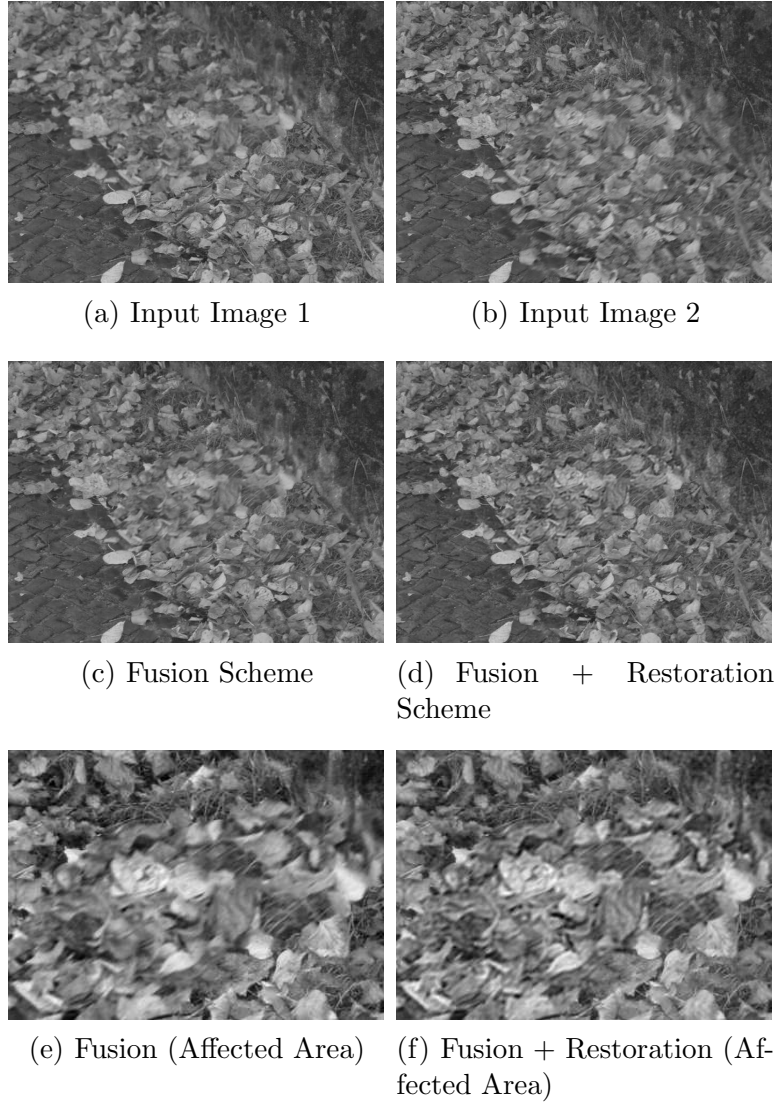
(a) Input Image 1

(b) Input Image 2

(c) Fusion Scheme

(d) Fusion + Restoration Scheme

(e) Fusion (Affected Area)

(f) Fusion + Restoration (Affected Area)

Fig. 7. Overall fusion improvement using the proposed fusion approach enhanced with restoration. Experiments with the "leaves" dataset.

Noise Ratio (PSNR) and the Image Quality Index Q_0, proposed by Wang and Bovik [17]. The visible edge enhancement in the common degraded area, provided by the extra restoration step is also confirmed by the two metrics.

Similar conclusions follow the next example with the "pebbles" dataset in Figure 9. The two artificially created blurred input images are depicted in Figures 9 (a), (b). In Figure 9(a), Gaussian blur is applied to the upper left part of the image and in Figure 9(b) Gaussian blur of different variance (randomly chosen) is applied to the bottom right part of the image. Again, the two input images contain an area of common degradation in the image centre and several areas that were not degraded in the bottom left and the top right of the image. In Figure 9(c), the result of the fusion approach using Isotropic Diffusion is depicted. As expected, the fusion algorithm manages to transfer all
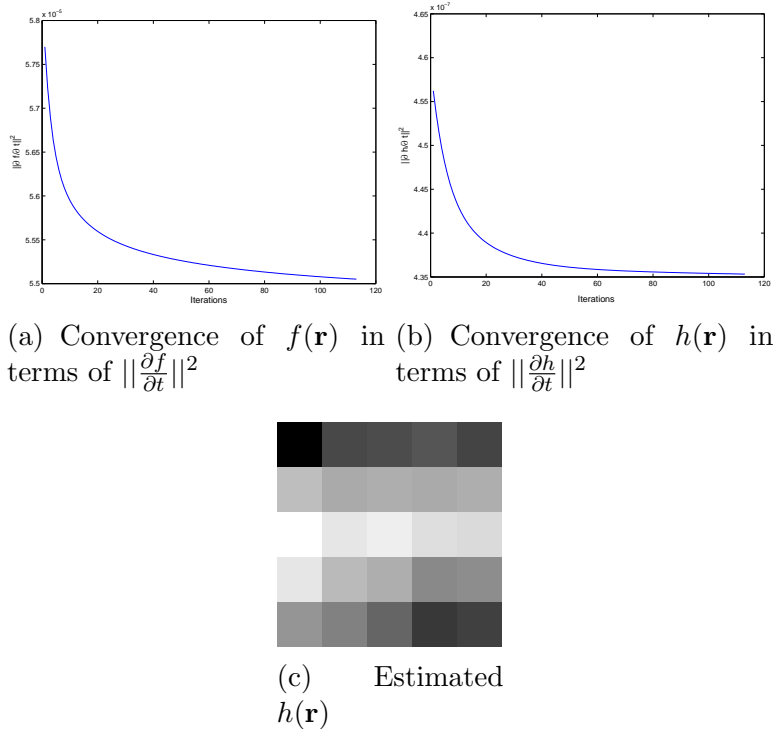
23

(a) Convergence of $f(\mathbf{r})$ in terms of $||\frac{\partial f}{\partial t}||^2$

(b) Convergence of $h(\mathbf{r})$ in terms of $||\frac{\partial h}{\partial t}||^2$

(c) Estimated $h(\mathbf{r})$

Fig. 8. Convergence of the restoration part and the final estimated $h(\mathbf{r})$ for the common degraded area in the "leaves" example. The directivity of the estimated mask indicates the estimation of motion blur.

high quality information to the fusion image except for the area in the centre of the image that still remains blurred. This common degraded area was properly identified by the proposed algorithm, using $p = 1.05$ and $q = 1$, as depicted in Figure 6 (e). In Figure 9 (d), the final enhanced image is depicted after the restoration process that has been applied on the common degraded area for $L = 3$. On the whole, the image quality has been enhanced compared to the original fused image. In Figure 9 (e), (f), a focus on the common degraded area in the fused and the fused/restored image can verify the above conclusions. The visible achieved enhancement of the new method is also supported by the PSNR and Q_0 measurements that are described in Table 3. The two methods based on error estimation also outperformed the ICA-based transform-domain method, as depicted in Table 3.

The third experiment demonstrates the capability of the proposed system to handle noisy cases as well. Two images were artificially created by blurring the upper left and down right respectively of an airplane image (British Airways - BA747) with randomly chosen Gaussian blur kernels. Additive white Gaussian noise of standard deviation 0.03 (input signals normalised to $[0, 1]$) was also added to both images, yielding an average SNR=27dB. As previously, there exists an area in the middle of the image, where the imposed degradations overlap, i.e. there is no ground truth information in any of the input images. The denoising term of the fusion step was activated by selecting $\alpha = 0.08$.

	Fused TopoICA		Fused Error Est.		Fused + Restored	
	PSNR (dB)	Q_0	PSNR (dB)	Q_0	PSNR (dB)	Q_0
Leaves	17.65	0.9727	25.740	0.9853	25.77	0.9864
Pebbles	21.27	0.9697	25.35	0.9713	25.99	0.9755
noisy BA747	17.35	0.9492	24.18	0.9757	24.41	0.9770
Porto	21.33	0.9860	22.94	0.9897	23.37	0.9907
Noisy Porto	19.71	0.9768	20.55	0.9818	20.62	0.9821

Table 3
Performance evaluation of the fused with Isotropic diffusion and the combined fusion - restoration approach in terms of PSNR (dB) and Q_0.

In Figure 10(c), the result of the fusion approach using Isotropic Diffusion is depicted. As previously, the algorithm managed to perform fusion of the areas where valid information is available in the input images, and also suppress the additive Gaussian noise. The common degraded area was identified using $p = 1$ and $q = 0.99$. These images contain large areas of constant background, whereas the two previous images contained a lot of textural detail. In this case, it is essential to avoid these large areas of constant background to be estimated as part of the common degraded area, and therefore, we choose $q = 0.99$ instead of 1 as previously. The restoration step was applied with $L = 3$, offering an overall enhancement in the visual quality and the actual benchmarks, compared to the error-estimation fusion approach and the ICA-based fusion approach. The calculated metrics suggest that there is limited significant improvement, because the enhancement in the relatively small common degraded area is averaged with the rest of the image. However, one can observe that there is obvious visual enhancement in the final enhanced image, especially in the common degraded area.

Another final example demonstrates the capability of the proposed system to handle noisy cases and more complicated scenes. Two images were artificially created by blurring the foreground object (statue in Porto-Portugal) along with some adjacent area and another area surrounding the statue with randomly chosen Gaussian blur kernels. A noiseless and a noisy example were created with additive white Gaussian noise of standard deviation 0.04 and 0.05 respectively(input signals were again normalised to $[0, 1]$). As previously, there exists an area surrounding the statue, where the imposed degradations overlap. In Figure 11, the noiseless example is depicted along with the results
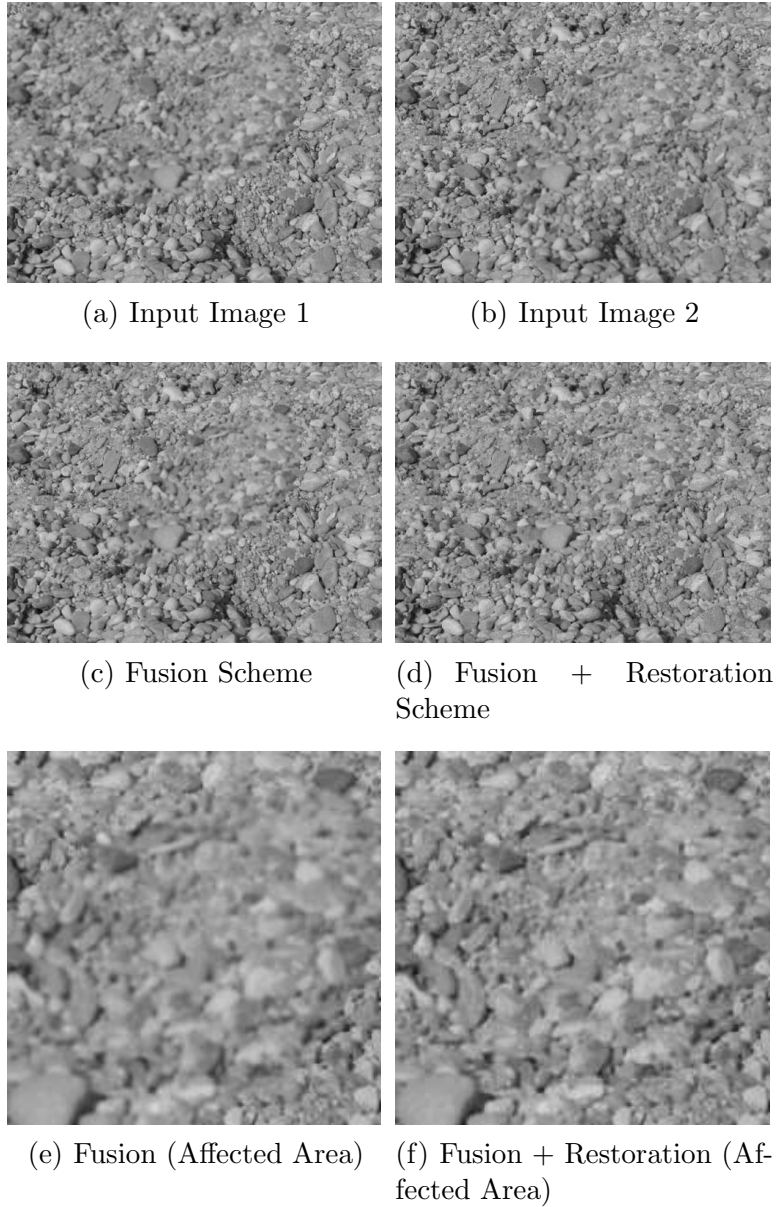
(a) Input Image 1   (b) Input Image 2

(c) Fusion Scheme   (d) Fusion + Restoration Scheme

(e) Fusion (Affected Area)   (f) Fusion + Restoration (Affected Area)

Fig. 9. Overall fusion improvement using the proposed fusion approach enhanced with restoration. Experiments with the "pebbles" dataset.

of the fusion with error estimation approach, the combined fusion-restoration approach and the Topographic ICA with the "max-abs" rule. The common degraded area was identified using $p = 0.92$ and $q = 1$. In the restoration step, the kernel size was chosen to be $L = 5$. In Figure 12, the corresponding results in the case of additive noise are depicted. The denoising term of the fusion step was activated by selecting $\alpha = 0.08$. As previously, the algorithm managed to perform fusion of the areas where valid information is available in the input images, and also suppress the additive Gaussian noise. The calculated performance indexes in Table 3 verify again the obvious visual enhancement in the final enhanced image, especially in the common degraded area.
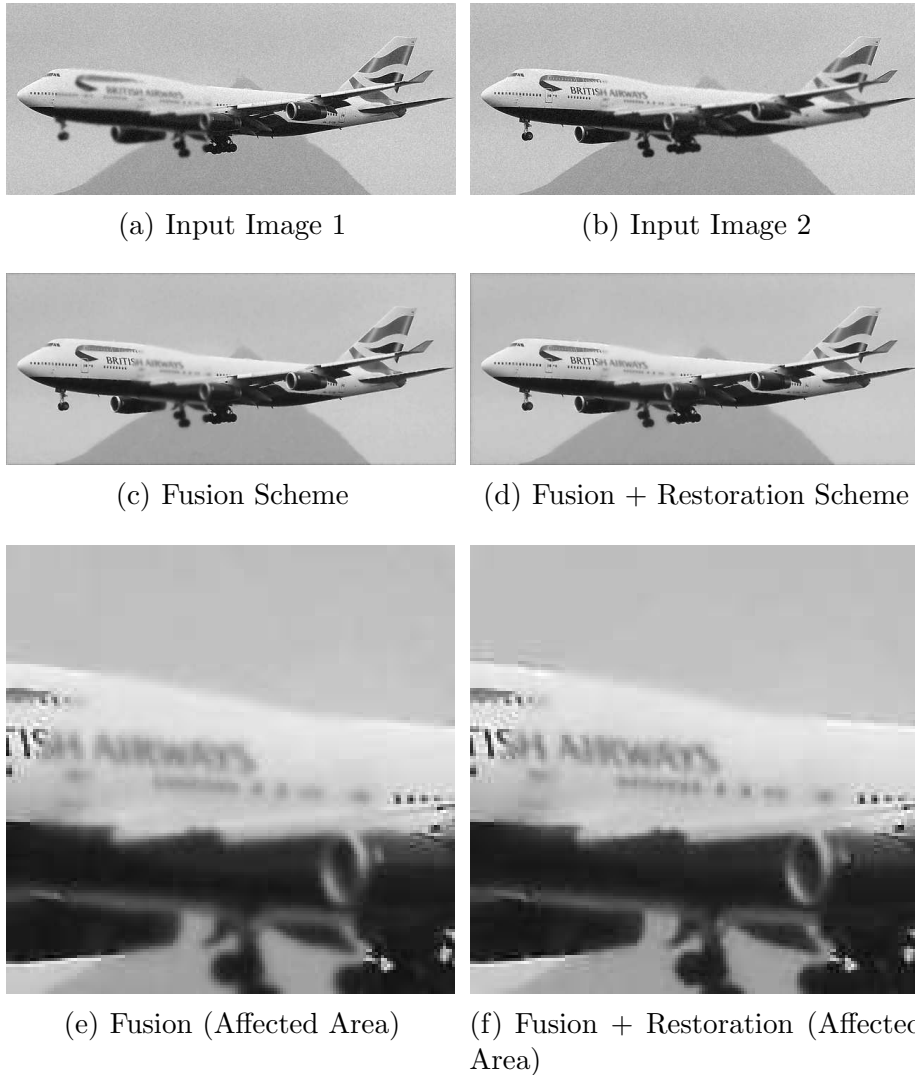
(a) Input Image 1

(b) Input Image 2

(c) Fusion Scheme

(d) Fusion + Restoration Scheme

(e) Fusion (Affected Area)

(f) Fusion + Restoration (Affected Area)

Fig. 10. Overall fusion improvement using the proposed fusion approach enhanced with restoration. Experiments with the "British Airways (BA747)" dataset.

5 Conclusions

The problem of image fusion, i.e. the problem of incorporating useful information from various modality input sensors into a composite image that enhances the visual comprehension and surveillance of the observed scene, was addressed in this study. More specifically, a spatial-domain method was proposed to perform fusion of both multi-focus and multi-modal input image sets. This method is based on error estimation methods that were introduced in the past for image enhancement and restoration and are solely performed in the spatial domain. In the case of multi-focus image sets scenarios the proposed spatial-domain framework seems to match the performance of several current popular transform-domain methods, as for example, the wavelet transform and the trained ICA technique. The proposed methodology exhibits also in-

(a) Input Image 1

(b) Input Image 2

(c) Ground Truth

(d) TopoICA fusion

(e) Fusion Scheme

(f) Fusion + Restoration Scheme

Fig. 11. Overall fusion improvement using the proposed fusion approach enhanced with restoration and comparison with the TopoICA fusion scheme. Experiments with the "Porto" dataset.

(a) Input Image 1   (b) Input Image 2

(c) Ground Truth   (d) TopoICA fusion

(e) Fusion Scheme   (f) Fusion + Restoration Scheme

Fig. 12. Overall fusion improvement using the proposed fusion approach enhanced with restoration and comparison with the TopoICA fusion scheme. Experiments with the "noisy-Porto" dataset.

teresting results in the case of multi-modal image sets, producing outputs with distinctively outlined edges compared to transform-domain methods.

More specifically, a combined method of fusion and restoration was proposed as the next step from current fusion systems. By definition, fusion systems aim only at transferring the "interesting" information from the input sensor images to the fused image, assuming there is proper reference image signal for all parts of the image in at least one of the input sensor images. In the case that there exist common degraded areas in all input images, the fusion algorithms cannot improve the information provided there, but simply convey this degraded information to the output. In this study, we proposed a mechanism of identifying these common degraded areas in the fused image and use a regularised restoration approach to enhance the content in this area. In the particular case of multi-focus images, the proposed approach managed to remove the blur and enhance the edges in the common degraded area, outperforming current transform-based fusion systems.

There are several potential applications of the proposed system. Military targeting or surveillance units can benefit from a combined fusion and restoration platform to improve their targeting and identification performance. Commercial surveillance appliances can also benefit from a multi-camera, multi-focus system that fuses all input information into a composite image with wide and detailed focus. In addition, there are several other applications such as increasing the resolution and quality of pictures taken by commercial digital cameras.

Acknowledgement

References

[1] H.C. Andrews and B.R. Hunt. *Digital Image Restoration*. Prentice-Hall, 1997.

[2] M.J. Black, G. Sapiro, D.H. Marimont, and D. Heeger. Robust anisotropic diffusion. *IEEE Transactions on Image Processing*, 7(3):421–432, 1998.

[3] . The Image fusion server. http://www.imagefusion.org/.

[4] D.B. Gennery. Determination of optical transfer function by inspection of the frequency-domain plot. *Journal of the Optical Society of America*, 63:1571–1577, 1973.

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

[6] S. John and M.A. Vorontsov. Multiframe selective information fusion from robust error estimation theory. *IEEE Transactions on Image Processing*, 14(5):577–584, 2005.

[7] N. Kingsbury. The dual-tree complex wavelet transform: a new technique for shift invariance and directional filters. In *Proc. IEEE Digital Signal Processing Workshop*, Bryce Canyon UT, USA, 1998.

[8] D. Kundur and D. Hatzinakos. Blind image deconvolution. *IEEE Signal Processing Magazine*, 13(3):43–64, 1996.

[9] N. Mitianoudis and T. Stathaki. Adaptive image fusion using ICA bases. In *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, Toulouse, France, May 2006.

[10] N. Mitianoudis and T. Stathaki. Pixel-based and Region-based image fusion schemes using ICA bases. *Information Fusion*, 8(2):131–142, 2007.

[11] R. Molina, A.K. Katsaggelos, and J. Mateos. Bayesian and regularization methods for hyperparameter estimation in image restoration. *IEEE Transactions on Image Processing*, 8(2):231–246, 1999.

[12] S.G. Nikolov, D.R. Bull, C.N. Canagarajah, M. Halliwell, and P.N.T. Wells. Image fusion using a 3-d wavelet transform. In *Proc. 7th International Conference on Image Processing And Its Applications*, pages 235–239, 1999.

[13] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7):629–639, 1990.

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

[15] G. Piella. New quality measures for image fusion. In *7th International Conference on Information Fusion, Stockholm, Sweden*, 2004.

[16] 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*, 4718:118–129, 2002.

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

[18] C. Xydeas and V. Petrovic. Objective pixel-level image fusion performance measure. In *In Sensor Fusion IV: Architectures, Algorithms and Applications , Proc. SPIE, vol. 4051*, pages 88 – 99, Orlando, Florida,, 2000.

[19] Y.L. You and M. Kaveh. A regularisation approach to joint blur identification and image restoration. *IEEE Transactions on Image Processing*, 5(3):416–428, 1996.

[20] M.E. Zervakis and T.M. Kwon. On the application of robust functionals in regularized image restoration. In *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, pages V289–V292, April 1993.

[21] D. Zwillinger. *Handbook of Differential Equations.* Academic Press, Boston, 1997.