LOCAL IMAGE FUSION USING DISPERSION MINIMISATION

Qi Li and Tania Stathaki

Communication and Signal Processing Group
Department of Electrical and Electronics Engineering
Imperial College London, UK

ABSTRACT

On the spatial domain, *image fusion* can be approached by estimating a set of fusion weights, that optimally measure the contribution of each pixel in source images to the fused one. Combining fusion weights with source images yields the fused result with improved visual perception. This paper aims to find these weights by minimising a Constant-Modulus (CM) cost function that describes the dispersion of the fused image. In order to accelerate convergence rate and avoid spurious solutions, we also introduce optimal learning rates while updating fusion weights. Experimental results reveal that our scheme provides comparable performance on fusing multi-focus images to multi-scale wavelet methods, such as Shift-Invariant Discrete Wavelet Transform (SI-DWT).

1. INTRODUCTION

Let $f_1, \ldots, f_k, \ldots, f_K$ denote K images capturing the same true f scene from different sensors. The task of combining them and form one single perceptually enhanced scene $\hat{f}$ is defined as *image fusion*. Image fusion methods have been explored in a variety of fields, such as surveillance, remote sensing, and medical applications and others.

A comprehensive survey of existing fusion techniques can be found in [1]. In general, fusion techniques can be divided into *transform* and *spatial* domain. In a transform domain, fusion schemes are performed onto the transformation coefficients using various pixel-based or region-based rules. The composed image will be recovered back to the spatial domain by an inverse transform. Popular choices of transforms include the pyramid transform [2], the Mallat algorithm [3], the a-trous algorithm [4], the Shift Invariant Discrete Wavelet algorithm (SI-DWT) [1] and others. As far as spatial domain is concerned, fusion is simply performed on the image itself by combining source images using appropriate weights that measure the contribution of pixels in each image. Spatial domain fusion has the advantage of easy implementation and low computational complexity [1]. While, a transform domain fusion scheme can produce improved fusion results than methods in the other category, as a good transform highlights the important features, i.e. edges information of images [1]. In this study, we propose a framework for an optimisation-based fusion scheme on the spatial domain, where fusion results arising from the proposed method will be comparable to those generated in the selective transform domains.

On the spatial domain, the key issue is to determine a set of optimal fusion weights that can effectively capture the useful information from the source images. Unlike the methods proposed in [1], such as global PCA and Local Measurement-based Methods, which measure the weights in a direct computational fashion, the main contribution of this paper is to employ optimisation-based method to estimate fusion weights. This is achieved by minimising the Constant-Modulus (CM) cost function [5], which approximates the dispersion of fused image from the original scene. The process of minimisation is performed alternatively on the estimated fusion weights and the original dispersion value of original scene, along with the introduction of optimal learning rates at each iteration.

The rest of paper is organized as follows: section 2 describes the mathematical model for image fusion. A detailed explanation of the optimisation is then given in section 3. Finally, section 4 supports the approach by presenting experimental results on benchmark images and section 5 concludes the paper.

2. PROBLEM FORMULATION

Assume K images of size $M \times N$ describing the same true scene f are captured by different sensors and registered to each other. After column-stacking we denote each image using a vector $\underline{f_k} = [f_k(1,1), \ldots, f_k(m,n), \ldots, f_k(M,N)]^T$ $(k = 1 \ldots K)$, where the index in $f_k(m,n)$ labels the pixel of the kth image at mth row $(m = 1, \ldots, M)$ and nth column $(n = 1, \ldots, N)$. The superscript T denotes vector transpose. The aim of image fusion is to reconstruct the fused image arranged in a vector $\underline{\hat{f}} = [\hat{f}(1,1), \ldots, \hat{f}(m,n), \ldots, \hat{f}(M,N)]^T$, which demonstrates an enhanced image perception over any individual image f_k.

To examine a local fusion scheme, we assign each pixel $f_k(m,n)$ with a distinct weight $w_k(m,n)$, measuring the relevance of pixel $f_k(m,n)$ in image $\underline{f}_k$. It is convenient to gather all the weights and intensity values at location (m,n)

This work is sponsored by Ministry of Defence UK.

together and denote them by single vectors, as $\underline{w}(m,n) = [w_1(m,n),\ldots,w_k(m,n),\ldots,w_K(m,n)]^T$ and $\underline{f}(m,n) = [f_1(m,n),\ldots,f_k(m,n),\ldots,f_K(m,n)]^T$. Consequently each pixel $\hat{f}(m,n)$ in the fused image is obtained by weighted sum of pixels $f_k(m,n)$ at the same location from source images $\hat{f}(m,n) = \sum_{k=1}^{K} w_k(m,n) f_k(m,n) = \underline{w}(m,n)^T \underline{f}(m,n)$

In the literature, popular choices of local fusion weights are directly based on local measurements: such as variance, contrast, sharpness and entropy.

3. LOCAL FUSION USING DISPERSION MINIMISATION

However, in this paper, We propose a new optimisation approach for image fusion, instead of calculating the local fusion weights in a direct fashion, stated previously. Specifically, the estimation of a set of optimal fusion weights are obtained through minimising an cost function, which describes certain image properties.

If the ground truth image is known as a priori, it is natural to construct the cost function by means of the criterion of Least Mean Sqaure Error (LMSE) and ultimately find the fused image. However, LMSE is unsuitalbe to the case of image fusion due to the unavailablity of the ground truth image. We alternatively attempt to minimise the *dispersion* of $\hat{f}(m,n)$ in a similar fashion to [6], The concept of *dispersion* was thoroughly used for blind equalisation of communication signals over dispersive channels [5], and it describes a process where transmitted signals are degraded in the physical medium [7]. The aim of dispersion minimisation is to remove as much dispersion of the output as possible by minimizing the Constant-Modulus (CM) cost function.

Following the similar idea of dispersion minimisation used in *Image Restoration* [6], we believe that the actual non-distorted representation of the observed scene, as well as the fused one should be less dispersed, in other words, distored than the distorted or different sensor images. Therefore, a fusion process can be described as a process of minimising the dispersion of the fused image $\hat{f}$, described by the two dimensional CM cost function J [6].

$$\begin{aligned} J &= \mathcal{E}[\left(\hat{f}(m,n)^2 - D\right)^2] \\ &= \mathcal{E}[\left((\underline{w}(m,n)^T \underline{f}(m,n))^2 - D\right)^2] \end{aligned} \tag{1}$$

where the constant D illustrates the true dispersion value of the original image f. $\mathcal{E}(\cdot)$ denotes the expectation (i.e. average) of an image along the axis m and n. As in [6], D is defined in such a way that when $\hat{f}(m,n)$ is ideally equal to $f(m,n)$, the cost function J reaches the minimum. The mathematical expression of D is described as the division of 4-th order central moment and variance of the original image f (assuming the image mean has already been deducted).

$$D = \frac{\mathcal{E}[f(m,n)^4]}{\mathcal{E}[f(m,n)^2]} \tag{2}$$

However, in the context of image fusion, the true dispersion value D of original image is not accessible. Unlike the work in [6], our problem of minimisation in (1) thus involves two unknown variables $\underline{w}(m,n)$ and D, and can be reformulated in (3).

$$\mathrm{Min}_{\underline{w}(m,n),D} J(\underline{w}(m,n), D) = \mathcal{E}[\left((\underline{w}(m,n)^T \underline{f}(m,n))^2 - D\right)^2] \tag{3}$$

For the normalisation and nonnegativity purpose, equation (3) is subject to the constraints (4):

$$D > 0 \qquad \underline{w}(m,n) \geq 0 \qquad \underline{e}^T \underline{w}(m,n) = 1 \tag{4}$$

where $\underline{e} = [1\,1\ldots1]^T$.

Moreover, we can approximate the expectation operator with an instantaneous operator, and the formulation of (3) is reduced to $J_I(\underline{w}^l(m,n), \mathcal{D})$, as follows:

$$\mathrm{Min}_{\underline{w}(m,n),D} J_I(\underline{w}(m,n), D) = \frac{1}{4}\left((\underline{w}(m,n)^T \underline{f}(m,n))^2 - D\right)^2 \tag{5}$$

Consequently, in terms of two unknown variables involved in J_I, we employ the stochastic descent procedures of J_I alternatively with respect to $\underline{w}(m,n)$ and D and impose the constraints (4) on the solution after each iteration to improve the stability of the algorithm. It is easy to view that when the convergence of the minimisation occurs, $\hat{f}$, estimated in (**??**) is closest to f and least dispersed.

To solve the alternating minimisation of J_I, we involve the first partial derivative of J_I with respect to $\underline{w}(m,n)$ and D, given as follows:

$$\frac{\partial J_I}{\partial \underline{w}(m,n)} = \mathcal{A} \cdot \underline{f}(m,n) \tag{6}$$

where $\mathcal{A} = \left(\left((\underline{w}(m,n)^T \underline{f}(m,n))^2 - D\right) \underline{w}(m,n)^T \underline{f}(m,n)\right)$ and

$$\frac{\partial J_I}{\partial D} = -\frac{1}{2}\left(\left(\underline{w}(m,n)^T \underline{f}(m,n)\right)^2 - D\right) \tag{7}$$

As a result, the procedures of alternative minimisation can be depicted in Figure 1, and summarised as follows:

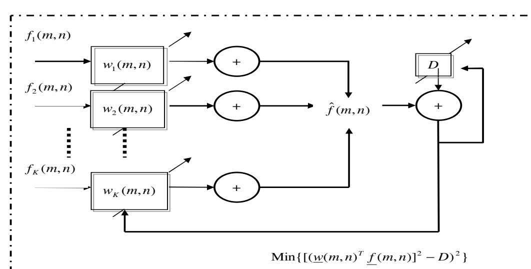


Fig. 1. Structure of the optimisation process.

- Initialise $\underline{w}(m,n) = \underline{e}/K$, which implies an equal weight assigned to each source image, and D is equal to the average of the dispersion values of K source images.

- At nth iteration, update $\underline{w}(m,n)$ with fixed D, using the following updating equation involving (6)

$$\underline{w}(m,n)^{+} \Leftarrow \underline{w}(m,n) - \mu \frac{\partial J_I}{\partial \underline{w}(m,n)} \tag{8}$$

$$\underline{w}(m,n)^{+} \Leftarrow |\underline{w}(m,n)| / (e^T |\underline{w}(m,n)|) \tag{9}$$

- At $(n+1)$th iteration, update D with fixed $\underline{w}(m,n)$, using the following updating equation involving (7)

$$D^{+} \Leftarrow D - \eta \frac{\partial J_I}{\partial D} \tag{10}$$

$$D^{+} \Leftarrow |D| \tag{11}$$

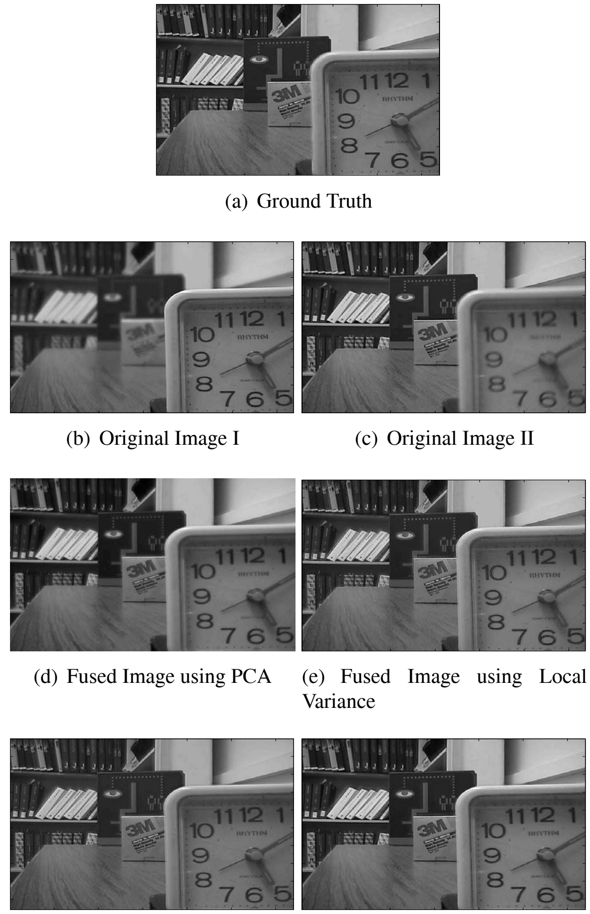
- Stop when converged.

where μ and η represent the constant learning rates., which can be determined empirically by the users.

However, the dispersion minimisation-based fusion scheme is slow to converge due to the non-convexity of the cost function J_I and constant learning rates. To overcome the convergence problem, we alternatively introduce the *line search* method to find the optimal learning rates for $\underline{w}(m,n)$ at each iteration, as given in

$$\begin{aligned} \mu_{opt} &= \min J_I(\mu) \\ &= \min[J_I((\underline{w}(m,n) - \mu \frac{J_I}{\underline{w}(m,n)}), D)] \end{aligned} \tag{12}$$

It is noteworthy that for given $\underline{w}(m,n)$ and D, J_I in (5) is a rational function of learning rate μ. Therefore, the optimal value of μ can be obtained by minimising the rational function $J_I(\mu)$ with respect to μ.



(a) Ground Truth

(b) Original Image I

(c) Original Image II

(d) Fused Image using PCA

(e) Fused Image using Local Variance

(f) Fused Image using SI-DWT

(g) Fused Image using Dispersion Minimisation scheme

Fig. 2. Fused images using PCA, Local Variance, SI-DWT and Dispersion Minimisation-based fusion scheme

4. FUSION MEASUREMENTS

4.1. Visual Analysis

In this part, we conduct the visual assessment of four fusion algorithms, including global Principal Component Analysis (PCA), local variance based fusion [1] and Shift Invariant Discrete Wavelet Transform (SI-DWT), and our dispersion minimiastion based fusion scheme, all performed on a set of multi-focus images. Figure 2 (a) illustrated the ground truth image with correct focus globally and (b)-(c) a pair of out-of-focus images, with correct foci in different parts. A maximum absolute selection rule is applied onto wavelet coefficients. Figure 2 (d)-(g) illustrate the fusion results arising from the tested methods. Compared to the original images with different foci, Figure 2 (f)-(g) are closest to the ground truth Figure 2 (a), and both results effectively relieve the edge distortion (see Figure 2 (e)), caused by Local Variance fusion, as well as the spectral distortion (see Figure 2 (d)), caused by Global PCA fusion. The underlying reason is that both methods more accurately capture local features for fusion process. The difference lies in that the wavelet method is developed on a multi-scale basis, while our scheme is on an optimisation basis. Overall, the results produced by using our method is the closet to the original image.

4.2. Quantitative Measurements

Moreover, we compare the fusion performances of all four algorithms in terms of quantitative measurements, which consist of the following three objective quantitative assessments. The numerical results are given in Table 1. Edge preservation Index (EI) and Image quality index (Q_0) examine the similarity of the fused result to the ground truth image, known as a priori, while Mean Gradient (MG) can examine the fused result without the ground truth image as reference.

- *Edge preservation Index (EI)* is firstly introduced in [8]. The value of EI evaluates the amount of edge information transferred from the ground truth image f to the fused one $\hat{f}$. One could refer to [8] for derivation of the measurement. Here we point out that the value of EI

Table 1. Performance Comparison for four fusion schemes: Clock

Methods	EI	Q_0	MG
PCA	0.5730	0.8446	3.3538
Local variance weight	0.6143	0.8803	4.0528
SI-DWT	0.7210	0.9929	4.8434
Our method	**0.7288**	**0.9950**	**5.4269**

fulfills the property $0 \leq EI \leq 1$. If $EI = 0$, it corresponds to an $\hat{f}$ that has no edge information inherited from f. Whereas, $EI = 1$ corresponds to no loss of edge information of $\hat{f}$, with respect to f.

- We also employ the *Image Quality Index* Q_0 to measure the similarity between the fused and ground truth image. Readers refer to [9] for an in-depth definition of Q_0. As $-1 \leq Q_0 \leq 1$, a closer value Q_0 to 1 indicates a stronger similarity, thus a better fusion performance.
- The third metric to use is the *Mean Gradient (MG)* . A larger value of MG reflects a higher contrast within the detailed variation of a pattern in the image and more clarity of the image. The mathematical definition is given in [10].

It is evident that both SI-DWT and our proposed method delivered a significant quantitative improvement over the other two methods, which is consistent to the results we drew from the visual measurement section. Although there is mild evidence that our method is better than SI-DWT in the visual assessment, the performance enhancement of our method is obvious in all the metrics, compared to SI-DWT.

On the other hand, we examine the performance of convergence between the case of constant learning rate μ and that of optimal μ_{opt}. Figure 3 shows that the introduction of μ_{opt} significantly accelerates convergence and enables a more accurate result.

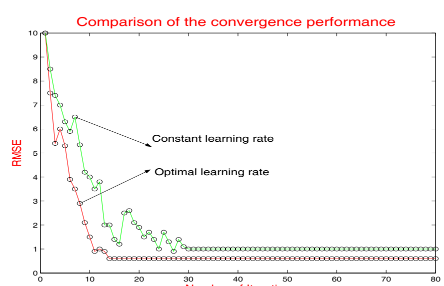


Fig. 3. Comparison of the convergence performances

5. CONCLUSIONS

We have presented a dispersion minimisation-based scheme for image fusion. For the future work, we prepare to explore other types of adaptive filter model for local fusion, such as autoregressive filtering (AR).

6. REFERENCES

[1] O. Rockinger and T. Fechner, "Pixel-level iamge fusion: The case of image sequences," *SPIE Procedding*, vol. 3374, pp. 378–388, 1998.

[2] P. J. Burt, "A gradient pyramid basis for pattern-selective image fusion," *Digest of Society for Information Display Technical Papers*, pp. 467–472, 1992.

[3] J. Zhou, D. L. Civco, and J. A. Silander, "A wavelet transform method to merge Landsat TM and SPOT panchromatic data," *International Journal on Remote Sensing*, vol. 19, no. 4, pp. 743–757, 1998.

[4] B. Aiazzi, L. Alparone, S. Baronti, and A. Garzelli, "Context-driven fusion of high spatial and spectral resolution images based on oversampled multi-resolution analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 10, pp. 2300–2312, 2002.

[5] J. R. Treichler and B. G. Agee, "A new approach to multipath correction of constant modulus signals," *IEEE Transactions on Communications*, vol. 31, no. 2, pp. 459–473, 1983.

[6] C. Vural and W. A. Sethares, "Blind deconvolution of noisy blurred images via dispersion minimization," *Digital Signal Processing, 2002. DSP 2002. 2002 14th International Conference on*, vol. 2, pp. 789–790, 2002.

[7] D. Godard, "Self-recovering equalization and carrier tracking into two dimensional data communication systems," *IEEE Transactions on Communications*, vol. 28, no. 11, pp. 1867–1875, 1980.

[8] C. S. Xydeas and V. Petrovic, "Objective image fusion performance measure," *Electronics Letters*, vol. 36, no. 4, pp. 308–309, 2000.

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

[10] J. Li and Z. J. Liu, "Data fusion for remote sensing imagery based on feature," *China Jounal of Remote Sensing*, vol. 2, no. 2, pp. 103–107, 1998.