MULTISCALE PHASE CONGRUENCY IMAGE FUSION SCHEME FOR INFRARED IMAGE AND VISIBLE IMAGE

Shuigen Wang[1], Dongshun Cui[1], M. R. Anjum[1], Juan Ma[2] and Baojun Zhao[1]

[1]School of Information and Electronics, Beijing Institute of Technology, Beijing, China.

[2]Institute of Microelectronics of Chinese Academy of Sciences, Beijing, China.

ABSTRACT: In this paper, a novel multiscale image fusion methodology for infrared image and visible image is proposed. The phase congruency (PC), which is a dimensionless measurement of the local structures of an image, is used as the feature for multiscale pyramid decomposition. Following the existing multiscale image fusion schemes, the source images are firstly decomposed by multiscale PC pyramid transform. The decomposed PC images of each scale are then combined into one image by a perceptual fusion operator. Finally, the fused image is reconstructed by inverse PC pyramid transform. The proposed method is compared with the state-of-the-art image fusion methods by merging parallel visible and infrared images. The experimental results demonstrate the better fusion performance of the proposed method.

Keywords: Phase congruency, image fusion, infrared image, visible image

I. INTRODUCTION

In the actual applications, for an image scene, one scene sensor may not capture all the useful objects in the image. Multi sensors are thus adopted to achieve multisource images. Image fusion is used for integrating different input source images into a single image, which has the ability to employ all the features in all the source images which can be used for human observation [1]. Up to date, infrared and visible images are the mostly used multisource images.

Numerous algorithms for multi-resolution image fusion [2-5] have been developed in recently few years. In [2], Laplacian pyramid image fusion is presented as a typical multiscale image fusion metric. Wavelet pyramid fusion is demonstrated to perform better than Laplacian pyramid fusion method in [3]. It is widely believed that the human visual system (HVS) is more sensitive to local luminance contrast than pixel values or luminance. Then the RoLP (ratio of low-pass) pyramid, namely contrast pyramid, is put forward by 'Toet' and shown to have better image fusion performance compared with other existing methods in [4, 5]. However, they suffer from using simple fusion operators with some simple logic/weighted combinations. Therefore, a more elaborate fusion algorithm is in demand to obtain better fusion visual effects.

According to a plethora of psychophysical and physiological evidences, it is found that visually discernable features coincide with those points, where the Fourier waves at different frequencies have congruent phases [6-8], *i.e.*, at points of high phase congruency (PC), we can extract highly informative features [9]. Therefore, PC is used as the feature for pyramid decomposition. Besides, a new novel image fusion operator by

Computing a standard parameter based on the segmented pyramid images is presented to fuse the PC pyramid maps for each scale. The presented extensive experimental results show that the proposed method has better fusion results than the existing state-of-the-art ones.

II. PC EXTRACTION OF IMAGES

Compared with extracting image features directly with obvious changes in intensity, the PC method extracts image features from image pixels which have the maximal Fourier transform components of image phase. According to the introduction of PC in [7], it exist several different kinds of implementations to calculate the accurate PC map for an input image. In our paper, the approach which was proposed by Kovesi [10] is adopted for extracting PC maps of images.

To extract PC features of 2-D grayscale images, we start from the computation of 1-D input$g(x)$. Assuming that M_n^e and M_n^o are the two even and odd filters respectively on the n layer, which compose a pair of quadrature filters. The signal responses for each filter can finally constitute a vector including all the responses of the position x on the n image scale: $[e_n(x), o_n(x)] = [g(x) * M_n^e, g(x) * M_n^o]$. The amplitude of the response is $A_n(x) = \sqrt{e_n(x)^2 + o_n(x)^2}$. Besides, we let $F(x) = \sum_n e_n(x)$ and $H(x) = \sum_n o_n(x)$. Therefore, the PC map of 1-D signal can be obtained as follows:

$$PC(x) = \frac{E(x)}{\varepsilon + \sum_n A_n(x)}$$

Where ε is an empirical positive value to avoid denominator zero and $E(x) = \sqrt{F^2(x) + H^2(x)}$.

In order to compute the two filters, *i.e.*, M_n^e and M_n^o, there exist two well-known filters, namely, Gabor and log-Gabor filters. The log-Gabor filters is employed in our method for following reasons: 1) one cannot construct Gabor filters of arbitrarily bandwidth and still maintain a reasonably small DC component in the even-symmetric filter, while log-Gabor filters, by definition, have no DC component [9]; 2) the transfer function of the log-Gabor filter has an extended tail at the high-frequency end, which makes it more capable to encode natural images than ordinary Gabor filters [9,10]. Therefore, the log-Gabor filter is used and its transform function is$G(\omega) = \exp\left(-(\log(\omega/\omega_0))^2/2\sigma_r^2\right)$, where ω_0 is the center frequency of the filter, and σ_r is the variable controlling the bandwidth of filter.

We then apply the 1-D computing algorithm over several orientations and adopt some laws to fuse the resultant 1-D PC maps to obtain the PC feature maps of 2-D images in grayscale. And the log-Gabor filters used in the 1-D PC computing introduced above should be extended to 2-D ones by applying some spreading function across the filter perpendicular to its orientation [9]. In this paper, Gaussian is utilized as the spread function, since the function phases can stay invariant functioned by Gaussian. With the Gaussian function, the transfer function for 2-D log-Gabor function thus can be represented as:

$$G_2(\omega,\theta_j) = \exp\left(-\frac{\left(\log\left(\frac{\omega}{\omega_0}\right)\right)^2}{2\sigma_r^2}\right)\cdot\exp\left(-\frac{(\theta-\theta_j)^2}{2\sigma_\theta^2}\right)$$

Where $\theta_j = j\pi/J$, $j=\{0,1,\dots,J-1\}$ is defined as the angles of various orientations, J is the total number of orientations, and σ_θ is the variable controlling the bandwidth of the filter.

As for the 2-D images, by modulating ω_o and θ_j and convolving G_2 with them, a series of responses can be obtained for each pixel x as $[e_{n,\theta_j}(\mathrm{x}), o_{n,\theta_j}(\mathrm{x})]$. The amplitude of the response on the n image scale and orientation θ_j is $A_{n,\theta_j}(\mathrm{x}) = \sqrt{e_{n,\theta_j}(\mathrm{x})^2 + e_{n,\theta_j}(\mathrm{x})^2}$, and the energy for the orientation θ_j is $E_{\theta_j}(\mathrm{x}) = \sqrt{F_{\theta_j}(\mathrm{x})^2 + H_{\theta_j}(\mathrm{x})^2}$, where $F_{\theta_j}(\mathrm{x}) = \sum_n e_{n,\theta_j}(\mathrm{x})$ and $H_{\theta_j}(\mathrm{x}) = \sum_n o_{n,\theta_j}(\mathrm{x})$. The PC maps of 2-D images at position x is defined as:

$$PC_{2D}(\mathrm{x}) = \frac{\sum_j E_{\theta_j}(\mathrm{x})}{\varepsilon + \sum_n\sum_j A_{n,\theta_j}(x)}$$

III. THE PROPOSED ALGORITHM

In this section, the detailed PC pyramid image

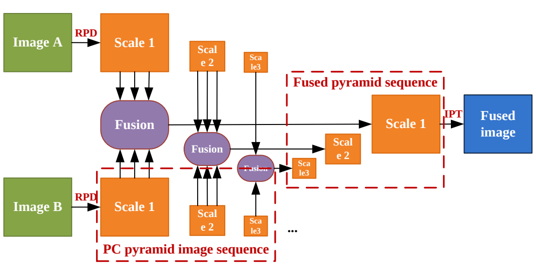

Figure 1. Schematic diagram of the proposed PC pyramid image fusion method. The RPD is the ratio of low-pass (RoLP) decomposition and IPT represents the inverse pyramid transform.

fusion algorithm is to be described. As shown in the Fig.1, the basic schematic diagram of the PC pyramid image fusion is the same as those of traditional multiscale fusion schemes which contains pyramid decomposition, image fusion and reconstruction.

A. Image Decomposition

Assume the initial image is represented as G_0 and it is also the bottom level image of the Gaussian pyramid decomposition. Following this way, the l level of the Gaussian pyramid is G_l, which is then obtained from the convolution results of the $l-1$ level image and the generating kernel ω as follows:

$$G_l(i,j) = \sum_{m,n=-2}^{2}\omega(m,n)G_{l-1}(2i+m,2j+n)$$

where $0<l<N$, N is the number of all the pyramid images. The $\omega(\mathrm{m},\mathrm{n})$ is represented as:

$$\omega(\mathrm{m},\mathrm{n}) = \frac{1}{400}\begin{bmatrix}1 & 5 & 8 & 5 & 1\\ 5 & 25 & 40 & 25 & 5\\ 8 & 40 & 64 & 40 & 8\\ 5 & 25 & 40 & 25 & 5\\ 1 & 5 & 8 & 5 & 1\end{bmatrix}$$

Since each pyramid image is reduced in half in the spatial domain, each image can thus be written with an array whose dimension is the half of the previous one. Therefore, the array G_l is expanded into array G_{l-1}^* by inserting some new values inside the primary pixel values, and the array G_{l-1}^* has the same size as G_{l-1} like

$$G_{l-1}^*(\mathrm{i},\mathrm{j}) = 4\sum_{m,n=-2}^{2}\omega(m,n)G_l\left(\frac{i+m}{2},\frac{j+n}{2}\right)$$

Then for each scale of pyramid images as G_l, we compute the PC maps of it as PC_l. Therefore, the corresponding RoLP pyramid image C_l is the ratio result of two successive pyramid images in the image sequence and computed as:

$$\begin{cases} C_l = PC_l/PC_l^* & for\quad 0\le l\le N-1\\ C_N = PC_N & for\qquad\qquad l=N\end{cases}$$

where C_l represents the l layer image of the RoLP pyramid decomposition; PC_l is the l layer image of PC pyramid decomposition.

B. Pyramid Image Fusion Process

As it is widely known that the intensity contrast is considered to be more important in mining and representing the informative image features. And this theory has strong biological grounds with numerous psychological evidences and has been employed in several image applications. Therefore, a standard parameter $d(B_k)$ is used as the fusion variable:

$$d(B_k) = \frac{1}{m'\times n'}\sum_{(i,j)\in B_k}\frac{|G_l(i,j)-\mu_k|}{\mu_k}$$

Where μ_k is the mean pixel value of image block B_k, and $m'\times n'$ is the size of B_k.

Assuming that there are two inputs, an infrared (IR) image and a visible (VIS) image, the two fused pyramid images of the two source images for each level are segmented into a set of smaller image blocks. We compare the values of standard parameters of the fused image blocks to obtain the final fused block image as

$$BC_{Fk} = \begin{cases} BC_{VISk} & if \quad d(BC_{VISk}) \geq d(BC_{IRk}) \\ BC_{IRk} & otherwise \end{cases}$$

where $d(BC_{VISk})$ and $d(BC_{IRk})$ are standard parameters of image blocks BC_{VISk} and BC_{IRk} related to pyramid image pair C_{VISl} and C_{IRl}. From the equation above, we can know that the final fused image is determined by the bigger one between $d(BC_{VISk})$ and $d(BC_{IRk})$. Finally, a set of fused images $C_{F0}, C_{F1}, \ldots, C_{FN}$ are generated.

C. Fused image reconstruction

The final step is reconstructing the final fused image obtained from the fused RoLP images. The method is defined as follows:

$$\begin{cases} F_N = C_{FN} & for \quad l = N \\ F_l = C_{Fl} F_l^* & for \ 0 \leq l \leq N-1 \end{cases}$$

Therefore, we can exactly get the reconstructed composite image F_0.

IV. Experimental Results

In this section, experimental results are presented to demonstrate the better performance of the proposed method compared with other three image fusion methods, i.e., Laplacian pyramid [2], wavelet pyramid [3], and Toet pyramid [4]. The results are demonstrated in the Fig.2.

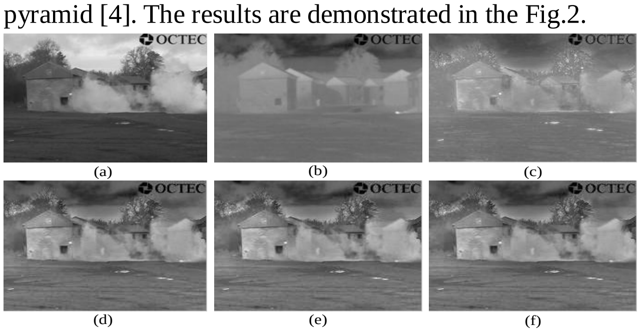

Figure 2. Experimental results of four methods.

Fig. 2(a) is the source visible image and Fig. 2(b) is the source infrared image. Fig. 2(c) to Fig. 2(f) are the fused results of Laplacian pyramid, wavelet pyramid, Toet pyramid and the proposed PC pyramid, respectively. From the Fig. 2, we can see that the proposed PC pyramid method has the best fusion performance while Laplacian pyramid method performs the worst.

V. CONCLUSION

In this paper, a new novel PC pyramid image fusion method was proposed to fuse the visible and infrared images. PC was adopted as feature for pyramid decomposition and a new perceptual fusion operator was presented. The experimental results have shown that the proposed method has good fusion performance with respect to the existing state-of-the-art fusion methods.

VI. REFERENCES

[1] Varshney P.K., Electron. Commun.1997.
[2] Burt P.J., Adelson E.H., Proc. SPIE. 1985.
[3]Burt P.J., Kolczynski R.J., *Enhaced image capture through fusion*, in *IEEE* 4th ICCV, **vol. 4** , pp. 173-182, 1993.
[4] Toet A., Ruyven L.J., Valeton J.M., Opt. Eng. 28 , 789, 1989.
[5] Toet A., Opt. Eng. 31 (1992), 1026.
[6] M. C. Morrone and D. C. Burr, “Feature detection in human vision: Aphasedependent energy model,” *Proc. R. Soc. Lond. B*, 1988.
[7] M. C. Morrone, J. Ross, D. C. Burr, and R. Owens, “Mach bands arephase dependent,” *Nature*, **vol. 324(6049)**, pp.250–253,1986.
[8] M. C. Morrone and R. A. Owens, “Feature detection from local energy,”*Pattern Recognit.Lett.*, **vol. 6(5),** pp. 303–313, Dec. 1987.
[9] L.Zhang, L.Zhang, X.Mou and D.Zhang, “FSIM:A feature similarity index for image quality assessment,” *IEEE Transactions on Image Processing.*, 2011.
[10] D. J. Field, “Relations between the statistics of natural images and theresponse properties of cortical cells,” *J. Opt. Soc. Amer.*, **vol. 4(12)**, pp. 2379–2394, 1987.