# Pixel-clarity-based multifocus image fusion

**Zhenhua Li (李振华), Zhongliang Jing (敬忠良), and Shaoyuan Sun (孙韶媛)**

*Institute of Aerospace Information and Control, School of Electrical and Information Engineering, Shanghai Jiaotong University, Shanghai 200030*

Received October 9, 2003

Due to the limited depth-of-field of optical lenses, it is difficult to get an image with all objects in focus. One way to overcome this problem is to take several images with different focus points and combine them into a single composite which contains all the regions full focused. This paper describes a pixel-clarity-based multifocus image fusion algorithm. The characteristic of this approach is that the pixels of the fused image are selected from the clearest pixels in the input images according to pixel clarity criteria. For each pixel in the source images, the pixel clarity is calculated. The fusion procedure is performed by a selection mode according to the magnitude of pixel clarity. Consistency verification is performed on the selected pixels. Experiments show that the proposed algorithm works well in multifocus image fusion.

*OCIS code:* 100.0100.

Image fusion is widely used in many fields such as remote sensing, medical imaging, machine vision and military application. The purpose of image fusion is to produce a single image from a set of input images. The fused image should have more complete information that is more useful for human or machine perception. The system reliability and capability can be improved by fusing the redundant information and complementary information in the source images.

Multifocus image fusion is one of the important research topics of image fusion. Suffering from the problem of limited depth of field, optical lenses usually cannot obtain good focus for all objects in a picture. One possible solution is applying image fusion. By taking several pictures with different focus points and fusing them together to form a single image, we can get a picture with full focus. This could be very useful, for example, in digital camera design or in industrial inspection applications. This paper presents a pixel-visibility-based multifocus image fusion algorithm.

In Refs. [1 − 3], multiresolution decomposition is used in multifocus image fusion. Multiresolution decomposition can decompose the signal into several components, each of them that capture information presents at a given scale. A region-based multiresolution image fusion is used to fuse multifocus images in Refs. [1,2]. Images are fused by combining their wavelet transforms guided by the important features such as edges and regions of interest. In Refs. [3,4], an area-based maximum selection rule and a consistency verification step are used for feature selection in combining the source wavelet transform coefficients.

A feature of human-vision-system-based multifocus image fusion algorithm is proposed in Ref. [5], which can achieve equivalent or superior performance to the algorithm based on wavelet decomposition. Registered input images are divided into several blocks firstly. Then contrast error of every block is calculated considering contrast masking sensitivity. Finally, the contrast errors of corresponding blocks are compared in order to decide which should be used to composite the fused image. Reference [6] describes a multifocus image fusion algorithm using artificial neural networks based on the use of image blocks. The algorithm first decomposes the source images into blocks. From each image block, three features (visibility, spatial information and edge feature) are extracted to reflect its clarity. Given two of these blocks (one from each source image), a neural network is trained to determine which one is clearer according to the features extracted from each block. Fusion then proceeds by selecting the clearer block in constructing the final image.

The two algorithms above both take the feature of human vision system into consideration and achieve high performance. However, the two algorithms all need to divide the source images into several blocks. It is a difficult problem to choose the optimal size of decomposed block for all kinds of multifocus images automatically. When dealing with blocks containing clear and blurry areas simultaneously, simple selection or weighted averaging of corresponding blocks in the source images cannot generate the best results. In this paper, we present a pixel-clarity-based multifocus image fusion algorithm. The fused image is combined by selecting all the clearest pixels in the source images.

Multifocus images have different focus points. Each of the multifocus images may contain some clear areas differently. By selecting all the clearest areas from the source images, we can composite a fused image with all objects full focused. Our fusion algorithm is based on pixel clarity defined in this paper. The fusion procedure is performed by a selection mode according to the magnitude of clarity. Only the clearest pixels in the source images are selected to combine the fused image. This fusion scheme is accordant with the characteristic of multifocus images.

In Refs. [5,6], the clarity of an image block is represented by spatial frequency ($SF$), visibility ($VI$), and edge feature ($EG$). The spatial frequency, visibility, and edge feature of an image block are defined in Refs. [5 − 7]. By experiments, we found that the clarity of a pixel can also be measured by the spatial frequency, visibility, and entropy ($EN$) of the small window centered on this pixel.

To simplify the description of the fusion scheme, we make an assumption that there are just two source images, $A$ and $B$, and the fused image is $F$. We note that

the fusion scheme described in this paper can also be extended to cases with more than two source images.

We use $f(x,y)$ to denote the intensity of pixel $(x,y)$, where $x$ and $y$ are its coordinates. The visibility of pixel $(x,y)$ is defined as

$$VI(x,y) = \frac{1}{m \cdot n} \sum_{|i| \le m, |j| \le n} \phi(m_{x,y}) \cdot \varpi(i,j) \cdot \frac{|f(x+i, y+j) - m_{x,y}|}{m_{x,y}}, \quad (1)$$

where $i, j$ are integers, and the subscripts $m$ and $n$ are positive integers that define a small window with the size $(2m+1) \times (2n+1)$ centered on the current pixel ($m$, $n$ are typically small, i.e. $m = n = 8$), $m_{x,y} = \frac{1}{m \cdot n} \sum_{|i| \le m, |j| \le n} f(x+i, y+j)$, $\phi(m_{x,y}) = \left(\frac{1}{m_{x,y}}\right)^{\alpha}$, $\varpi(i,j)$ is the coefficient of the Gaussian template $\varpi$, $\sum_{|i| \le m} \sum_{|j| \le n} \varpi(i,j) = 1$, $\alpha$ is a visual constant ranging from 0.6 to 0.7.

The spatial frequency of pixel $(x,y)$ is defined as

$$SF(x,y) = \sqrt{[RF(x,y)]^2 + [CF(x,y)]^2}, \quad (2)$$

where $RF(x,y)$ and $CF(x,y)$ are the row frequency

$$RF(x,y) = \sqrt{\frac{1}{m \cdot n} \sum_{|i| \le m, |j| \le n} [f(x+i, y+j) - f(x+i, y+j-1)]^2}, \quad (3)$$

and column frequency

$$SF(x,y) = \sqrt{\frac{1}{m \cdot n} \sum_{|i| \le m, |j| \le n} [f(x+i, y+j) - f(x+i-1, y+j)]^2}, \quad (4)$$

respectively, where $i$, $j$ are the same meaning as Eq. (1).

$EN(x,y)$ is the entropy of pixel $(x,y)$ which is based on the entropy of the window centered on this pixel with the size $(2m+1) \times (2n+1)$. It is defined as

$$EN(x,y) = -\sum_{l} [h(l) \log_2 h(l)], \quad (5)$$

where $l$ is an arbitrary pixel intensity in the window and $h$ is its normalized histogram in the window.

After the calculation of $VI_A(x,y)$, $SF_A(x,y)$, $EN_A(x,y)$ of image $A$, and $VI_B(x,y)$, $SF_B(x,y)$, $EN_B(x,y)$ of image $B$, the normalization is defined as

$$\begin{cases} VI_A(x,y) = \frac{VI_A(x,y)}{VI} \\ VI_B(x,y) = \frac{VI_B(x,y)}{VI} \end{cases}, \quad (6)$$

where $VI = VI_A(x,y) + VI_B(x,y)$,

$$\begin{cases} SF_A(x,y) = \frac{SF_A(x,y)}{SF} \\ SF_B(x,y) = \frac{SF_B(x,y)}{SF} \end{cases}, \quad (7)$$

where $SF = SF_A(x,y) + SF_B(x,y)$, and

$$\begin{cases} EN_A(x,y) = \frac{EN_A(x,y)}{EN} \\ EN_B(x,y) = \frac{EN_B(x,y)}{EN} \end{cases}, \quad (8)$$

where $EN = EN_A(x,y) + EN_B(x,y)$.

The clarity of pixel $(x,y)$ is the combination of normalized $VI(x,y)$, $SF(x,y)$ and $EN(x,y)$ as

$$CL(x,y) = w_1 \cdot VI(x,y) + w_2 \cdot SF(x,y) + w_3 \cdot EN(x,y), \quad (9)$$

where $w_1$, $w_2$ and $w_3$ are the weights.

The proposed fusion scheme is shown in Fig. 1. It can be divided into three steps:

1) Calculate the pixel clarity of each pixel for the source images: $CL_A(x,y)$ and $CL_B(x,y)$.

2) Select the pixels with higher magnitude of pixel clarity

$$f_F(x,y) = \begin{cases} f_A(x,y), & \text{if } CL_A(x,y) \ge CL_B(x,y) \\ f_B(x,y), & \text{if } CL_A(x,y) < CL_B(x,y) \end{cases}. \quad (10)$$

3) Perform consistency verification on the selected pixel $f_F(x,y)$: If the center pixel comes from image $A$ while the majority of surrounding pixels come from image $B$, the center pixel is then changed to come from image $B$, vice versa. The size of consistency verification window is denoted as $(2M+1) \times (2N+1)$.

The examples are given here to illustrate the fusion process described above. Figure 2 illustrates the fusion results of two real multifocus images. Two synthesized multifocus images were fused in Fig. 3. The synthesized images were used to simulate multifocus images. According to the diagonal line, the reference image (a) is divided into two parts: the left triangle and the right triangle. The simulated multifocus images (b) and (c) are obtained by performing Gaussian blurring (radius = 2.0 pixels) on the left and right triangles, respectively.

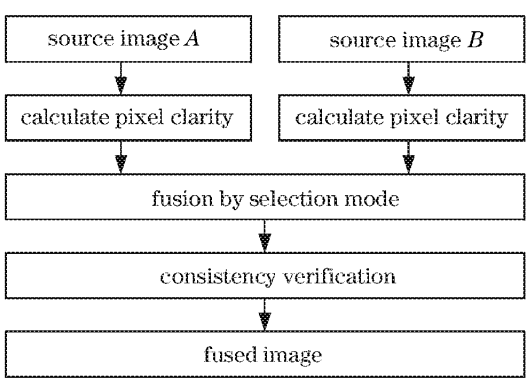


Fig. 1. Pixel-clarity-based multiforcus image fusion scheme.

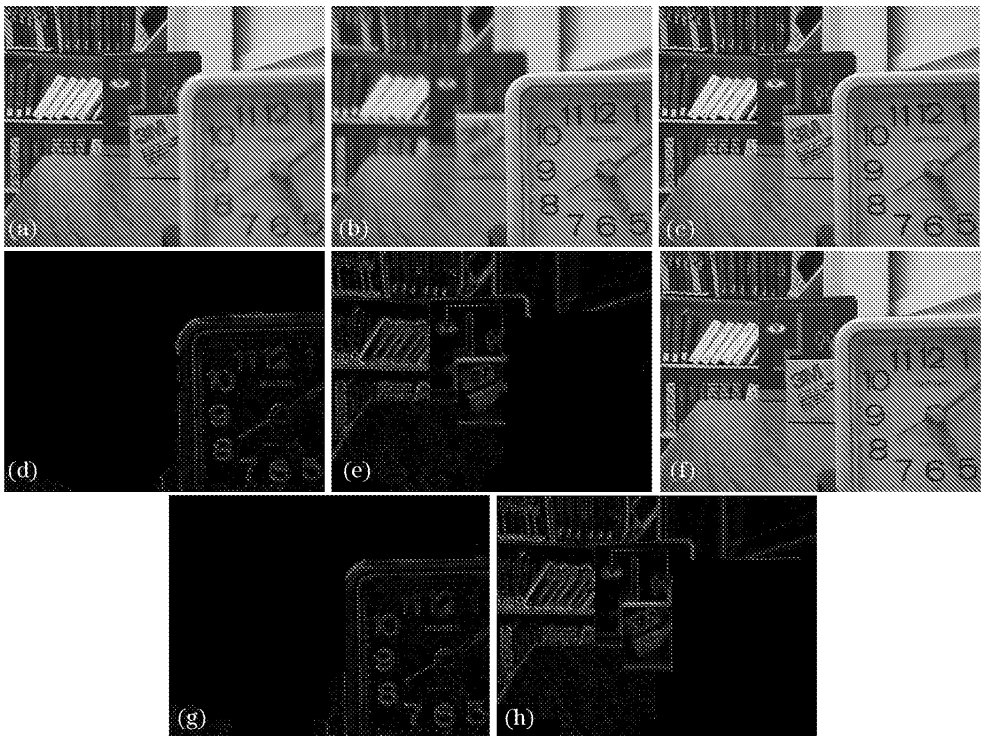


Fig. 2. Fusion of real multifocus images. (a) Focus on the left; (b) focus on the right; (c) fused image using our algorithm; (d) difference between the fused image and source image ($A$) using our algorithm; (e) difference between the fused image and source image ($B$) using our algorithm; (f) fused image using the algorithm of Ref. [6]; (g) difference between the fused image and source image ($A$) using the algorithm of Ref. [6]; (h) difference between the fused image and source image ($B$) using the algorithm of Ref. [6].

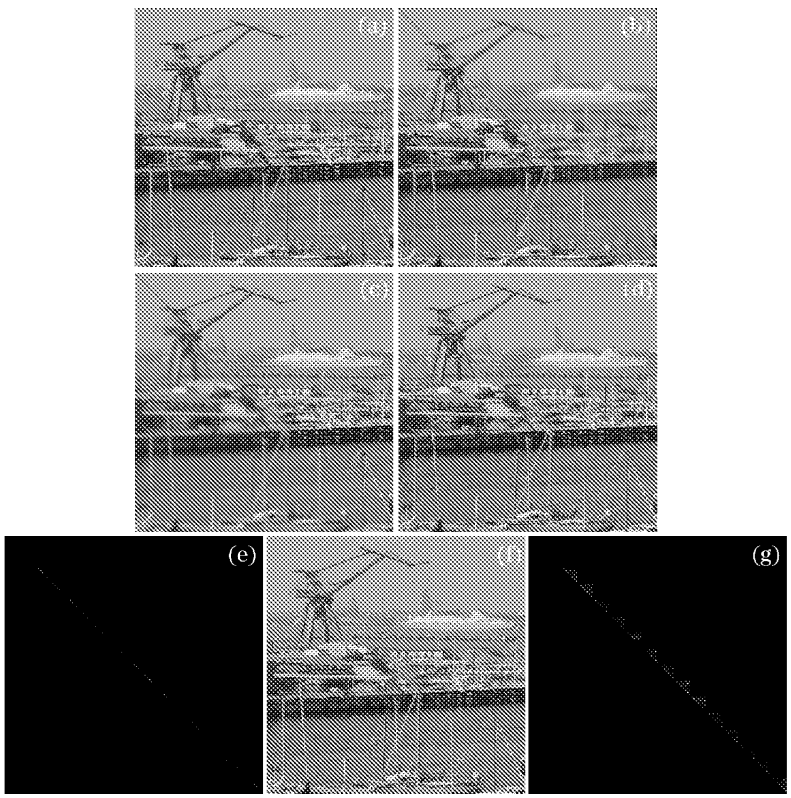


Fig. 3. Fusion of synthesized multifocus images. (a) The reference image; (b) right triangle blurred; (c) left triangle blurred; (d) fused image using our algorithm; (e) difference between the fused and reference images using our algorithm; (f) fused image using the algorithm of Ref. [6]; (g) difference between the fused and reference images using the algorithm of Ref. [6].

Objective evaluation methods are employed to evaluate the performance of each image fusion algorithm:

1) Entropy ($H$)

$$H = -\sum_{i=0}^{L}[h(i)\log_2 h(i)], \tag{11}$$

where $h$ is the normalized histogram of the fused image $F$ to be evaluated, $L$ is the maximum value for a pixel in the image, in our tests, $L$ is equal to 255. The entropy is used to measure the overall information in the fused image. The larger the value is, the better fusion results we get.

2) Spatial frequency ($SF$)

Spatial frequency is used to measure the overall activity level of an image. The larger the value is, the better fusion results we get. The spatial frequency of a fused image $F$ is defined as

$$SF = \sqrt{RF^2 + CF^2}, \tag{12}$$

where $RF$ is the row frequency

$$RF = \sqrt{\frac{1}{N_1 \cdot N_2}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}(f(i,j+1) - f(i,j))^2}, \tag{13}$$

and $CF$ is the column frequency

$$CF = \sqrt{\frac{1}{N_1 \cdot N_2}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}(f(i+1,j) - f(i,j))^2}, \tag{14}$$

$N_1$, $N_2$ are the height and width of fused image $F$, and $f(i,j)$ is the intensity of pixel $(i,j)$.

3) The root mean squared error ($RMSE$) between the reference image $R$ and the fused image $F$

$$RMSE = \sqrt{\frac{1}{N_1 \cdot N_2}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}(r(i,j) - f(i,j))^2}, \tag{15}$$

where $N_1$, $N_2$ are the height and width of reference image and $r(i,j)$ is the intensity of pixel $(i,j)$ in the reference image. The root mean squared error is used to measure the difference between the reference image and the fused image. The less the value is, the better fusion results we get. It is used for the evaluation of synthesized multifocus image fusion.

In the test of our algorithm, we use $m = n = 8$, $M = N = 8$, and the weights $w_1$, $w_2$, $w_3$ are 0.4, 0.4, 0.2 respectively.

In the test of algorithm of Ref. [6], the PNN network is used and $\sigma = 0.09$. We use an image block size of $32 \times 32$.

The objective evaluation of our algorithm and the algorithm of Ref. [6] are shown in Tables 1 and 2. From the tables, we can see the performance of our fusion algorithm is superior to the algorithm of Ref. [6].

**Table 1. Objective Evaluation of Real Multifocus Image Fusion**

| | Our Algorithm | Algorithm of Ref. [6] |
|---|---|---|
| $H$ | 7.2791 | 7.2746 |
| $SF$ | 15.6208 | 15.5596 |

**Table 2. Objective Evaluation of Synthesized Multifocus Image Fusion**

| | Our Algorithm | Algorithm of Ref. [6] |
|---|---|---|
| $H$ | 6.7573 | 6.7567 |
| $SF$ | 29.0234 | 28.6138 |
| $RMSE$ | 0.8351 | 3.0342 |

A pixel-clarity-based multifocus image fusion scheme is proposed in this paper. The input multifocus images are fused by a selection mode according to the magnitude of pixel clarity. Consistency verification is performed on the selected pixels. By experiments, we find the clearest pixels in source images are selected into the fused image with few pixels selected wrong.

In future work, the features, which reflect the clarity of pixels more exactly, will be developed. When dealing with the pixels on the edge between blur areas and clear areas, the selection mode works not very well. Other methods should be employed to deal with such pixels.

This work was jointly supported by the National 863 Project (No. 2001AA135091), the National Natural Science Foundation of China (No. 60375008), Shanghai Key Scientific Project (No. 02DZ15001), China Ph.D Discipline Special Foundation (No. 20020248029) and China Aviation Science Foundation (No. 02D57003). Z. Li's e-mail address is randy_lee@sjtu.edu.cn.

## References

1. Z. Zhang and R. S. Blum, Proceedings of IEEE **87,** 1315 (1999).
2. Z. Zhang and R. S. Blum, in *Proceedings of 30th Conference on CISS* (March 1997).
3. H. Li, B. S. Manjunath, and S. K. Mitra, Graphical Models and Image Processing **57,** 235 (1995).
4. S. T. Li, Y. N. Wang, and L. Z. Gong, Systems Engineering and Electronics (in Chinese) **24,** 45 (2002).
5. S. T. Li, Y. N. Wang, and C. F. Zhang, Acta Electron. Sin. **29,** 1699 (2001).
6. S. T. Li, T. K. James, and Y. N. Wang, Pattern Recogn. Lett. **23,** 985 (2002).
7. J. W. Huang, Y. Q. Shi, and X. H. Dai, Image Graphics **4,** 400 (1999).