Image and Vision Computing 24 (2006) 1278–1287

Enhancing effective depth-of-field by image fusion using mathematical morphology

Ishita De [b], Bhabatosh Chanda [a,*], Buddhajyoti Chattopadhyay [a]

[a] *Electronics and Communication Sciences Unit, Indian Statistical Institute, Kolkata 700 108, India*
[b] *Department of Computer Science, Barrackpore Rastraguru Surendranath College, Kolkata 700 120, India*

Received 5 March 2004; received in revised form 1 February 2006; accepted 5 April 2006

Abstract

Reduced depth-of-field (DOF) poses a problem in the light optical imaging system, since the objects present outside this zone appear blurry in the recorded image. The effective DOF of the sensor may be enhanced considerably without compromising the quality of the image by fusing images captured with different focused regions. This paper presents an image fusion technique suitable for combining multifocus images of a scene. The method employs morphological filters to select sharply focused regions from various images and then combines them together to reconstruct the image in which all the regions are properly focused. A performance measure based on image gradients is used to compare the results obtained by the proposed method with those obtained by other image fusion techniques.

Keywords: Depth-of-field; Multifocus images; Registration; Correlation; Image fusion; Morphological filters; Multiscale top-hat transformation; Image gradient

1. Introduction

A scene to be photographed usually includes objects at varying distances from the camera. The sharpness distribution of an image of such a scene is affected by various factors. The object focused by the camera and the objects at the same distance from the camera as the focused object appear to be the sharpest in the image. The sharpness of the objects in front of and behind the focused distance decreases gradually in the image. This sharpness-loss is not significant within a certain range of object distances. This range is called the depth-of-field or depth-of-focus (DOF) of the camera. It depends on various factors such as, the amount of sharpness-loss regarded as acceptable, the aperture used (decreasing the aperture will increase the DOF), the distance of the focused object (nearer the object, shorter the DOF) and the focal length of the lens (longer the focal length, shorter the DOF). An extreme case of decreasing the aperture for maximizing the DOF happens in the case of pinhole camera and it has an infinite DOF. Unfortunately, the optical power in the image plane is reduced considerably due to the infinite DOF. Therefore cameras with finite DOF are preferred. But the problem is that they cannot generate the images of all objects at various distances (from the camera) with equal clarity. One way to enhance the effective DOF is to acquire several images of a scene with focus on different parts of it and then fusing them in such a way that all regions of the scene are in focus. This process is known as multifocus image fusion. Before fusion, the constituent images must be registered or brought into a common coordinate system so that the corresponding objects are spatially matched properly. Both the reliability of redundant information and the quality of complementary information present in the constituent images are improved in an image that is registered and fused. Therefore it gives a better view for human and/or machine perception. A fused data can also render itself more successfully for any subsequent processing like object recognition, feature extraction, segmentation, etc.

Work is being done in this field by a number of researchers. The problem of reduced DOF in light optical microscopes due to increment in magnification and aperture has been discussed in [1,2]. One method to enhance the effective DOF by image fusion using controllable cameras is described in [3]. But this method depends on controlled camera motion and do not work for arbitrary set of images. There are several other techniques for multifocus image fusion. Simple techniques in which the fusion operation is performed directly on the source images (e.g. weighted average method) often have serious side effects like reduction in the contrast of the fused image. There are

probabilistic techniques [4], which involve huge computation using floating-point arithmetic and thus require a lot of time and memory-space. Fusion can be done using image gradients with majority filtering [5]. But this method has the drawback that the defocused zone of one image is enhanced at the expense of focused zone of others.

An image often contains physically relevant features at many different scales or resolutions. Multiresolution and multiscale approaches provide a means to exploit this fact. This is one of the reasons why these techniques are so popular. Various methods employing multiresolution techniques with pyramid or wavelet transform have been suggested in [6–9]. A survey on these works may be found in [8]. Burt and Kolezynski [6] suggested a method in which the images are first decomposed into a gradient pyramid. Activity measure of each pixel is computed then by finding out the variance of a 3×3 or 5×5 window centered on that pixel. Depending on this measure, either the larger value or the average value is chosen. Finally, the reconstruction is done. Li et al. [7] used similar method except the fact that wavelet transforms are used for decomposition and consistency verification is done along with window-based activity measure and maximum selection. Their method reduces the artifacts such as blocking effects, which are common in image fusion using multiresolution. Methods described in [6,7] are complex and time-consuming. Moreover, it was not mentioned whether the method could be applied to more than two multifocus images. In the method suggested by Yang et al. [9], an impulse function is defined at first to describe the quality of an object in a multifocus image. Then sharply focused regions are extracted by analyzing the wavelet coefficients of two primary images and two blurred images. To fuse two images, this method compares the wavelet coefficients of four images and thus involves double computation. In general, there are certain problems in multiresolution techniques of image fusion. These techniques belong to any of the following categories, pixel-based, window-based or region-based [8]. Pixel-based methods concentrate on a single pixel and window-based methods concentrate on a $k\times k$ window where k is very small. It is known that, even a very small error in registration results in mismatch of all the pixels in consideration. Pixel-based methods would not be able to tackle the situation and produce erroneous results. So they are not robust against misregistration. Window-based and region-based multiresolution techniques are better in this respect but they require a sequence of complex and time-consuming processes and hence take a large computation time.

Methods for image fusion using multiscale morphology are described in [10,11]. In [10], Matsopoulos et al. described a method for fusion of two multimodal images. In their method, morphological filters with structuring elements of varying size are used to construct a morphological pyramid. Such pyramids are constructed for all the input images. Next, morphological difference pyramids are constructed for each of the above pyramids. Third, an intermediate pyramid is constructed by combining information at each level from the above difference pyramids. Finally, reconstruction is done by using appropriate morphological operations on the intermediate pyramid to produce the fused image. This method can be used for multifocus image as well, but it was not mentioned whether the method could be applied to more than two input images. Since, in the third step the difference pyramids are combined by choosing the maximum at each pixel, this method is sensitive to the problem of misregistration as mentioned before. Mukhopadhyay and Chanda used similar method in [11] except that they have used morphological towers instead of morphological pyramids. But their method involves processing and storing of scaled data at various levels, which are of the same size as that of the original images. This results in a huge amount of memory and time requirement.

We propose a method for enhancing the effective depth-of-field of a sensor by multifocus image fusion using morphological filters. Our method does not employ analysis and synthesis operators an employed in methods employing multiresolution with pyramid or wavelet transform [6–11]. In our method, at first the constituent images are brought into a common coordinate system by registration. Then the focused regions in each image are identified by employing morphological filters, namely, opening and closing. Finally, these regions are fused into a single image by simply copying them into the resultant image. It is a region-based method and thus is robust to error in registration. It is known that morphological operators treat an image as an ensemble of sets and hence have the capability of handling objects of different shapes. Since, the best-focused regions are selected and copied from one image only, a slight error in registration will have no effect in fusion except in the border regions. In the pixel-based methods, all the pixels will be affected by misregistration. Manual cut-and-paste of focused regions from multifocus images is considered to be the best and it is often used for comparison purposes [7]. Our method is a close approximation to this and produces good results. A performance measure based on image gradients is used to compare the results obtained by our method with those obtained by other image fusion techniques. The paper is organized as follows. Section 2 describes the proposed method in detail. Section 2.1 presents the method employed for registration, Sections 2.2, 2.3 and 2.4 present the methods for multiscale top-hat transformation, detection of focused regions and image reconstruction, respectively. Experimental results and discussion including performance analysis are given in Section 3. Finally, concluding remarks are placed in Section 4.

2. Proposed method

The basic objective of our work is to find the focused region from every image and then simply copying the pixels of focused regions from input images to the fused image. Gray-scale morphology is used here to identify the focused portions of a particular image. An object common to all the images may have different position and scale in different images. Fusion of such images is meaningful only when such common objects are made to have identical geometrical configuration with respect to size, location and orientation in all the images. So the first and very important step towards fusion, which may be interpreted as preprocessing step, is registration. Then, the

focused regions from all images are detected. Finally, the fused image is reconstructed by combining the focused regions from all the images.

2.1. Registration

Integration or fusion of multifocused images is possible only if the images are registered or positioned with respect to a common coordinate system [12–14]. Establishing a correspondence requires matching of identical shapes in the related image pairs. There are various techniques for registering images. They can be categorized in two types, area-based techniques and feature-based techniques [13]. Area-based techniques are preferably applied when local shapes or structures vary over the images due to blurring and noise and the distinctive information is provided by graylevels/colors rather than by local shapes and structures. Our objective in this work is to enhance the effective depth-of-focus of a camera and we assume that input images are acquired one by one in identical environmental conditions in respect to sensor, light, view-point and the object-content in the scene. However, the input images may vary in terms of intensity levels and focusing. Hence, an area-based correlation method is employed for registration of the input images. This method does not require any manual intervention for feature selection.

The correspondence between the source and destination images is determined based on a measure of similarity of the intensity maps of the graylevels. Though all the images of the collection are equally authentic with their co-ordinate system, to register the images one of them is arbitrarily chosen as the destination image and the rests are treated as the source images. The model for similarity based matching between the destination image g_d and a source image g_s could be given by

$$g_d(x_\mathrm{d},y_\mathrm{d}) = g_s(x_\mathrm{s},y_\mathrm{s}) + m + \eta(x_\mathrm{s},y_\mathrm{s})$$

Mapping from the source coordinate $(x_\mathrm{s}, y_\mathrm{s})$ to the destination coordinate $(x_\mathrm{d}, y_\mathrm{d})$ is handled by an affined transformation T, i.e.

$$(x_\mathrm{d},y_\mathrm{d}) = T((x_\mathrm{s},y_\mathrm{s}))$$

while m is a bias term modeling the possible difference in the mean intensity level. The parameter η is assumed to be the signal independent zero-mean Gaussian noise that is associated with every image acquisition process. The correlation between g_d and g_s may be calculated as:

$$C(g_d,g_s) = \frac{\sum_{(x,y)}[g_d(x,y)-\bar{g}_d][g_s(x,y)-\bar{g}_s]}{\sqrt{\sum_{(x,y)}[g_d(x,y)-\bar{g}_d]^2}\sqrt{\sum_{(x,y)}[g_s(x,y)-\bar{g}_s]^2}} \tag{1}$$

To register the image g_s with g_d, we need to find out the transformation T such that the correlation $C(g_d, g_s)$ is maximum. Since, g_d and g_s are multifocus images of a scene acquired with preconditions as mentioned before, T is essentially an affine transformation which takes care of probable translation, rotation and scaling between g_d and g_s.

To register the images in a reasonable time, we have used the motion estimation techniques common in video compensation [15,16]. The destination or reference image g_d is divided into n number of blocks of size, say, $b_x \times b_y$. Suppose top-left corner of jth block has the co-ordinate $(x_{\mathrm{d}_j},y_{\mathrm{d}_j})$. Then, we try to find out the best match for this block through exhaustive search within the window $[x_{\mathrm{d}_j}-t_x, x_{\mathrm{d}_j}+b_x+t_x] \times [y_{\mathrm{d}_j}-t_y, y_{\mathrm{d}_j}+b_y+t_y]$ in g_s by calculating the above correlation. Let $(x_{\mathrm{s}_j},y_{\mathrm{s}_j})$ be the top-left corner of the corresponding best-matching block of g_s. The process is repeated for n blocks and thus, n pairs of corresponding points $(x_{\mathrm{d}_j},y_{\mathrm{d}_j})$ and $(x_{\mathrm{s}_j},y_{\mathrm{s}_j})$, $j=1, 2, \ldots, n$ are obtained. We rank these pairs according to their correlation values calculated and about one-fourth of the pairs with higher ranks are chosen. The parameters of the affine transformation T from g_s to g_d are estimated by using the least square method by using these best-matching pairs of points. Finally, the registered image is obtained from g_s by applying T along with bilinear interpolation. The time required for registration is negligible. An example of a set of three multifocus colored images in RGB format and the corresponding registered graylevel images are shown in Fig. 1. The graylevel intensity image was computed as $(R+G+B)/3$ before registration. The performance of registration is discussed in Section 3.

2.2. Multiscale top-hat transformation

After registration, our next step is multiscale opening and closing operation on the images. The shape of the structuring element B plays a crucial role in extracting features or objects of given shape from the image. A morphological operation with a scalable structuring element can extract features based on shape and size simultaneously [17,18]. Multiscale opening and closing are defined, respectively, as

$$(g \circ nB)(x,y) = ((g \ominus nB) \oplus nB)(x,y) \tag{2}$$

$$(g \bullet nB)(x,y) = ((g \oplus nB) \ominus nB)(x,y) \tag{3}$$

where B is a set representing structuring element of a definite shape while n is an integer representing the scale factor. If B is convex, in discrete domain, nB is obtained as:

$$nB = B \underbrace{\oplus B \oplus B \oplus \cdots \oplus B}_{n-1 \text{ times}} \tag{4}$$

By convention $nB=\{(0, 0)\}$ when $n=0$ and B is taken here to be a disk of unit size. Then the top-hat transformation for opening and closing filters are defined, respectively, as:

$$d_o^{(n)}(x,y) = (g \circ (n-1)B)(x,y) - (g \circ nB)(x,y) \tag{5}$$

$$d_c^{(n)}(x,y) = (g \bullet nB)(x,y) - (g \bullet (n-1)B)(x,y) \tag{6}$$

Thus, $d_o^{(n)}(x,y)$ contains all the bright features that have size greater than or equal to $n-1$ but less than n. Similarly, $d_c^{(n)}(x,y)$ contains all the dark features within the same range of size. Hence, the feature image defined as

$$D^{(n)}(x,y) = \max\{d_o^{(n)}(x,y), d_c^{(n)}(x,y)\} \tag{7}$$

Fig. 1. A set of multifocus colored images and the registered graylevel images. (a) Near-focused image, (b) middle-focused image, (c) far-focused image, (d)–(f) corresponding registered graylevel images.

contains all the image features having size within the range $[n-1, n)$. Hence, image features are thus sieved out based on their size and stored in corresponding $D^{(n)}$.

2.3. Detection of focused regions

It is evident from the previous discussion that if a particular feature (bright or dark) of an image is sharply focused it is sieved out in relatively lower scale. Suppose g_j $(j=1, 2, \ldots, k)$ denote a set of registered multifocus images of a single scene and $D_j^{(n)}$ s are their feature images at scale n. Suppose $F_j^{(n)}$ marks the focused regions of g_j at scale n. $F_j^{(n)}(x,y)=1$ means that the point (x, y) in image g_j is in focus and that is detected at the nth iteration. Now, if the pixel (x, y) is sharply focused in the image g_j, then at lower scale, $D_j^{(n)}(x,y)$ should be greater than $D_i^{(n)}(x,y)$ for all $i \neq j$. Thus, the focused regions of each image g_j can be identified and marked. The detection of focused regions at various scales $F_j^{(n)}$ can be algorithmically presented as

Step-1 $F_j^{(0)}(x,y)=0$ for all j
Step-2 $n=1$
Step-3 Calculate $D_j^{(n)}(x,y)$ for all images g_j
Step-4 $F_j^{(n)}(x,y)=1$, if $D_j^{(n)}(x,y) > \max_{i,i\neq j}\{D_i^{(n)}(x,y)\}$
Step-5 $F^{(n)}(x,y)=F_1^{(n)}(x,y) \vee F_2^{(n)}(x,y) \vee \cdots \vee F_k^{(n)}(x,y)$
Step-6 If all pixels of $F^{(n)}(x, y)$ are not equal to 1, increase n by 1 and go to Step-3

Hence, the focused regions or, more specifically, the focused pixels in the image g_j are marked by 1 in $F_j^{(n)}$. In practice, we terminate this algorithm when at least $p\%$ pixels of $F^{(n)}$ become 1 or no further change occurs in $F^{(n)}$. Rest unresolved pixels where $F^{(n)}(x, y)=0$ usually belong to smooth regions and are taken care of at the subsequent stage. Focused regions $F_j^{(n)}$ detected at various scales from the near-focused image in Fig. 1(d) are shown in Fig. 2.

2.4. Reconstruction

Image of focused region $F_j^{(n)}$ for jth input image may appear to contain spurious white spots in sharply focused region (shown here as black colored) and black spots in the out-of-focus region (shown here as white colored). This phenomenon can be observed in Fig. 2. It is evident that an object or

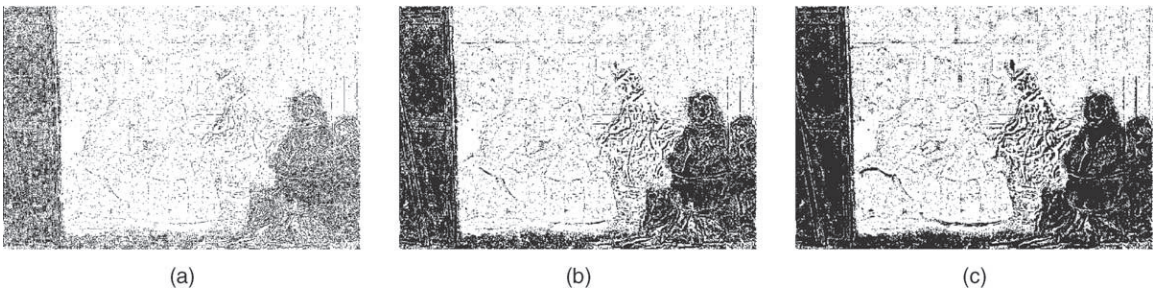

Fig. 2. Binary images corresponding to focused regions detected at various scales for the near-focused image in Fig. 1(d). Focused regions obtained by using (a) 2×2 SE, (b) 4×4 SE, (c) 8×8 SE.

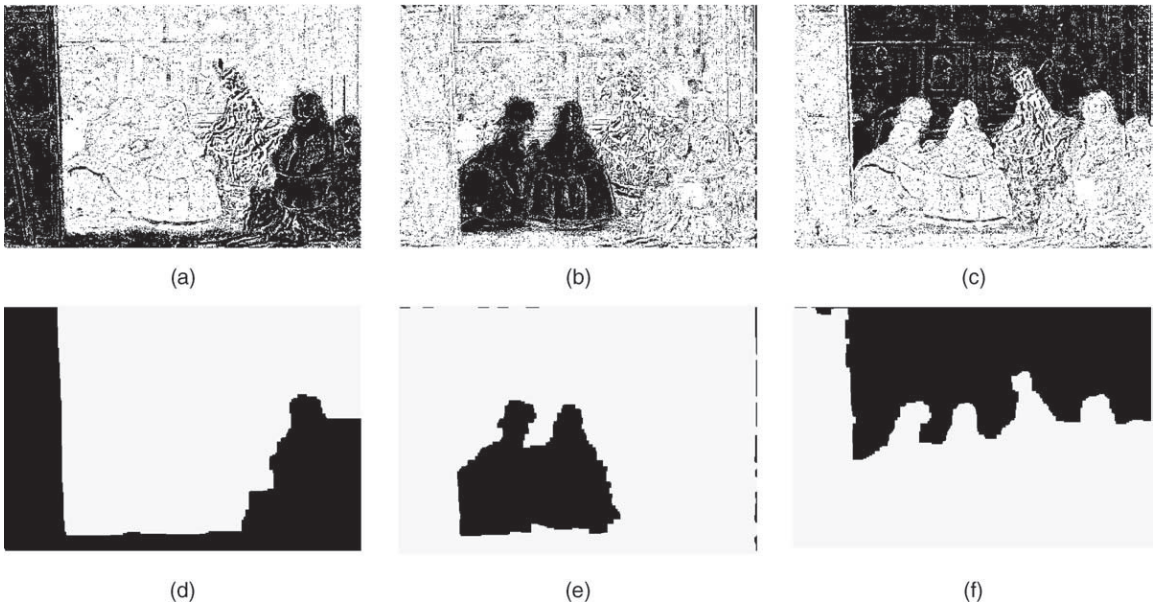

Fig. 3. Detected focused regions and the corresponding largest connected regions obtained from the registered multifocus images in Fig. 1.

focused region must be wider than the dimension of these spurious spots. So these spots are due to noise or due to the unresolved pixels that remain unlabeled after the previous processing. It is well known that opening and closing filter, respectively, can remove additive and subtractive noise efficiently. Hence, an alternating sequential filter formed by concatenating opening and closing with a small disk structuring element is applied on each binary image $F_j(x, y)$ to obtain $R_j(x, y)$ containing solid black blob(s). So the large connected regions in R_j mark the focused regions in g_j. Focused regions detected at the third iteration $F_j^{(3)}$ from the registered multifocus images in Fig. 1 and the corresponding R_j's is shown in Fig. 3. Now the image where all regions are properly focused may be reconstructed by putting together the pixels of g_j's corresponding to marked (black) regions of R_j's. However, it should be mentioned here that opening and closing with a disk structuring element trims some sharp convex portions from the blob and appends some sharp concave portions to the blob, respectively. In addition to this, some unresolved pixels might still be present. As a result R_j's are neither disjoint nor exhaustive. That means neither $R_i \wedge R_j$, $i \neq j$ produces a blank (or white) image nor $\vee_j R_j$ produces a filled (or black) image. Hence, the resultant fused image $\tilde{g}(x,y)$ is generated by non-linear superposition of $g_j(x, y)$ depending on $R_j(x, y)$ as follows.

$$\tilde{g}(x,y) = \begin{cases} g_j(x,y), \text{ if } R_j(x,y) = 1 \text{ and } R_i(x,y) = 0 \text{ for all } i \neq j \\ \text{avg}\{g_j(x,y) | R_j(x,y) = 0 \text{ for all } j\} \\ \text{avg}\{g_j(x,y) | R_j(x,y) = 1 \text{ for more than one } j\} \end{cases} \tag{8}$$

The function avg$(\cdot)$ stands for pixel-wise average from a set of images. Reconstructed graylevel and colored images in which all the focused regions are copied from the original multifocus images given in Fig. 1 are shown in Fig. 4.

3. Experimental results and discussions

The proposed method for image fusion has been implemented in C language on Unix environment and has been tested on a number of multifocus images. For color images in RGB format, at first the graylevel intensity image are computed as $(R+G+B)/3$. Then, registration and fusion methods were applied to get fused graylevel images. It should be noted that to get fused colored images after detecting the focused regions, during reconstruction instead of intensity

Fig. 4. Reconstructed graylevel and colored images where all regions are in sharp focus.

Table 1
RMSE of the images before and after registration

Figure	Before registration	After registration
Fig. 1(a) and (b)	63.607082	21.256035
Fig. 1(b) and (c)	62.827560	20.671530
Fig. 1(c) and (a)	22.625168	21.979143

values $g_j(x, y)$ in the Eq. (8) the RGB-color vectors $(g_{j_R}(x,y),g_{j_G}(x,y),g_{j_B}(x,y))$ are used.

Root-mean-square-error (RMSE) is a common measure for estimating the performance of registration [19]. We have calculated the RMSE of the multifocus images before and after registration. The results for Fig. 1 are given in Table 1. We see that the error has reduced in each case after registration. The inherent difference in the intensity levels in each pair of images is there, since they are the images with focus on different regions of the scene. Therefore, the error cannot be zero.

We have compared our results with those obtained by three other fusion methods, namely, image gradient method with majority filtering as described in [5], fusion by using Haar wavelet [7] and fusion by weighted averaging. For our method, we have taken p to be 95, i.e. at least 95% of the focused regions considering all the multifocus images are detected.

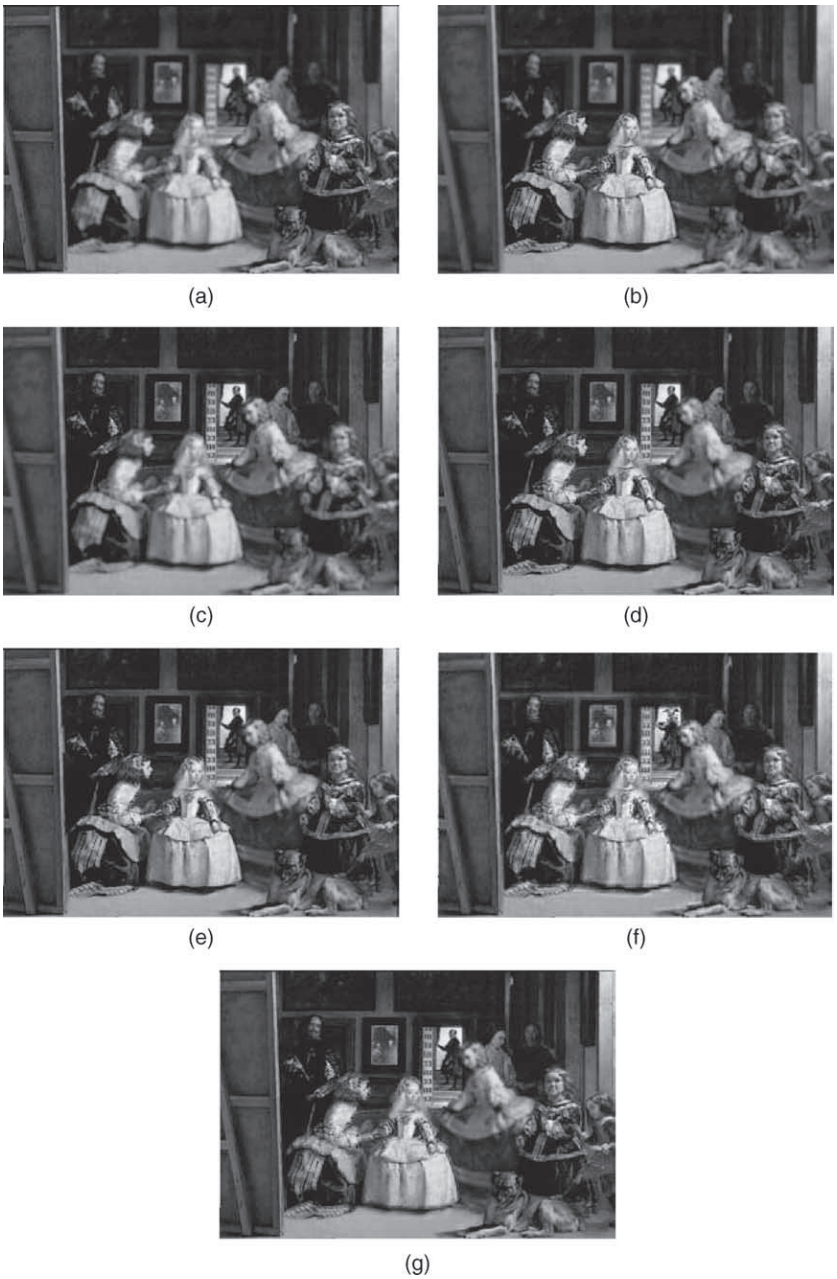

Fig. 5. Original multifocus images and the fused images by various algorithms. (a) Near focused image, (b) middle focused image, (c) far focused image, (d) fused image with proposed algorithm, (e) fused image with gradient method with majority filtering, (f) fused image with Haar wavelet method, (g) fused image with weighted averaging.

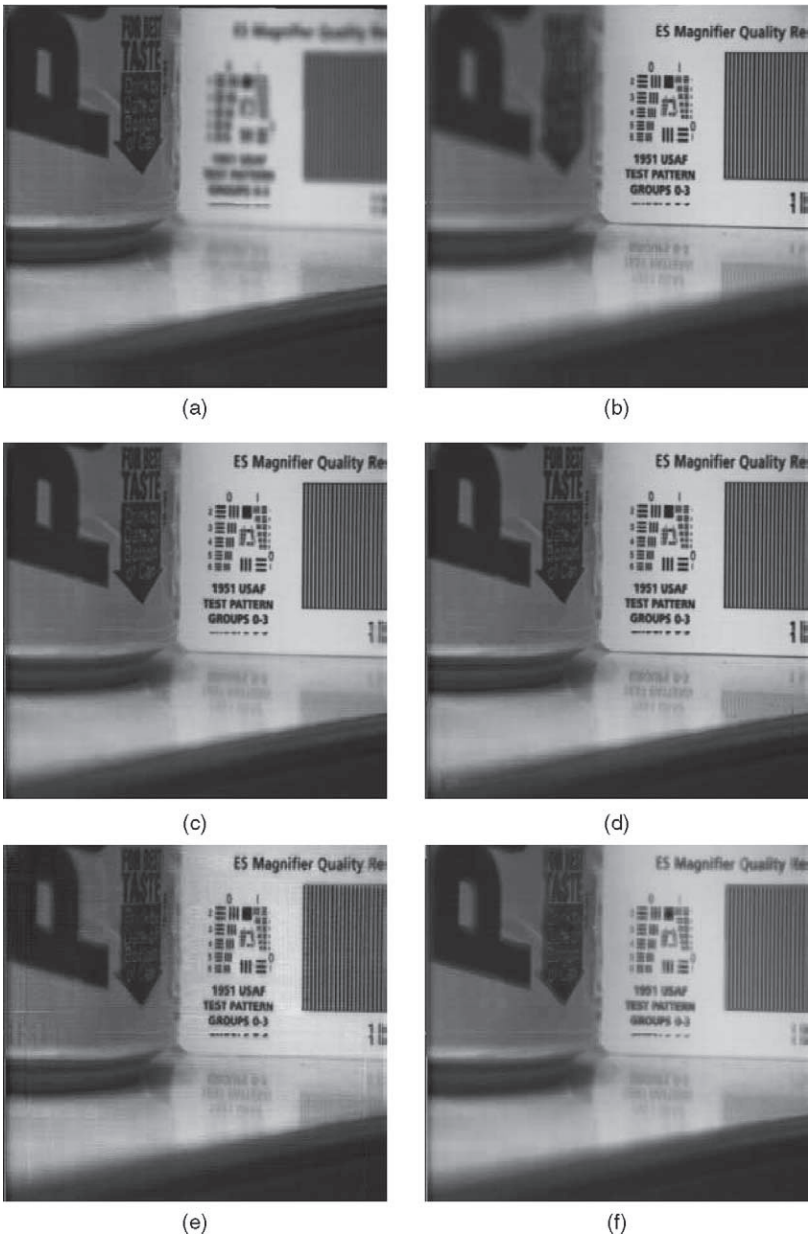

Fig. 6. Original multifocus images and the fused images by various algorithms. (a) Near focused image, (b) far focused image, (c) fused image with proposed algorithm, (d) fused image with gradient method with majority filtering, (e) fused image with Haar wavelet method, (f) fused image with weighted averaging.

This experimentally leads to $n=3$. That means three iterations are sufficient to detect focused regions in all the images (see Section 2.3). Image gradient method with majority filtering were tested by taking the parameters $k=40$, 50, 60 and $\beta=0.5$, 1.0, 1.5 and since, no significant variation was observed we have given the results with $k=40$ and $\beta=1.0$. Fusion by using Haar wavelet [7] has been implemented in the following way. At first, the registered multifocus images are decomposed into scaled images and detail images up to level 3 by using the linear Haar wavelet. Then the decomposed images are combined by choosing the pixel with maximum absolute value for each pixel in the scaled and detail images. Finally, the reconstruction is done from the combined image. In fusion by weighted averaging, the weights are determined by doing the principal component analysis, which employs the variance measure. The experimental results are shown in Figs. 5–8. In each figure, the original multifocus images are given first, followed by the fused images obtained by the proposed method and the other methods mentioned above.

3.1. Performance analysis

Careful manual inspection of Figs. 5–8 reveals that the results obtained by weighted averaging method have lost sharpness (more prominent in Fig. 8(f)) and the results obtained by Haar wavelet method have blocking artifacts in them (more prominent in Fig. 8(e)). The results obtained by the proposed fusion method is better than fusion by Haar

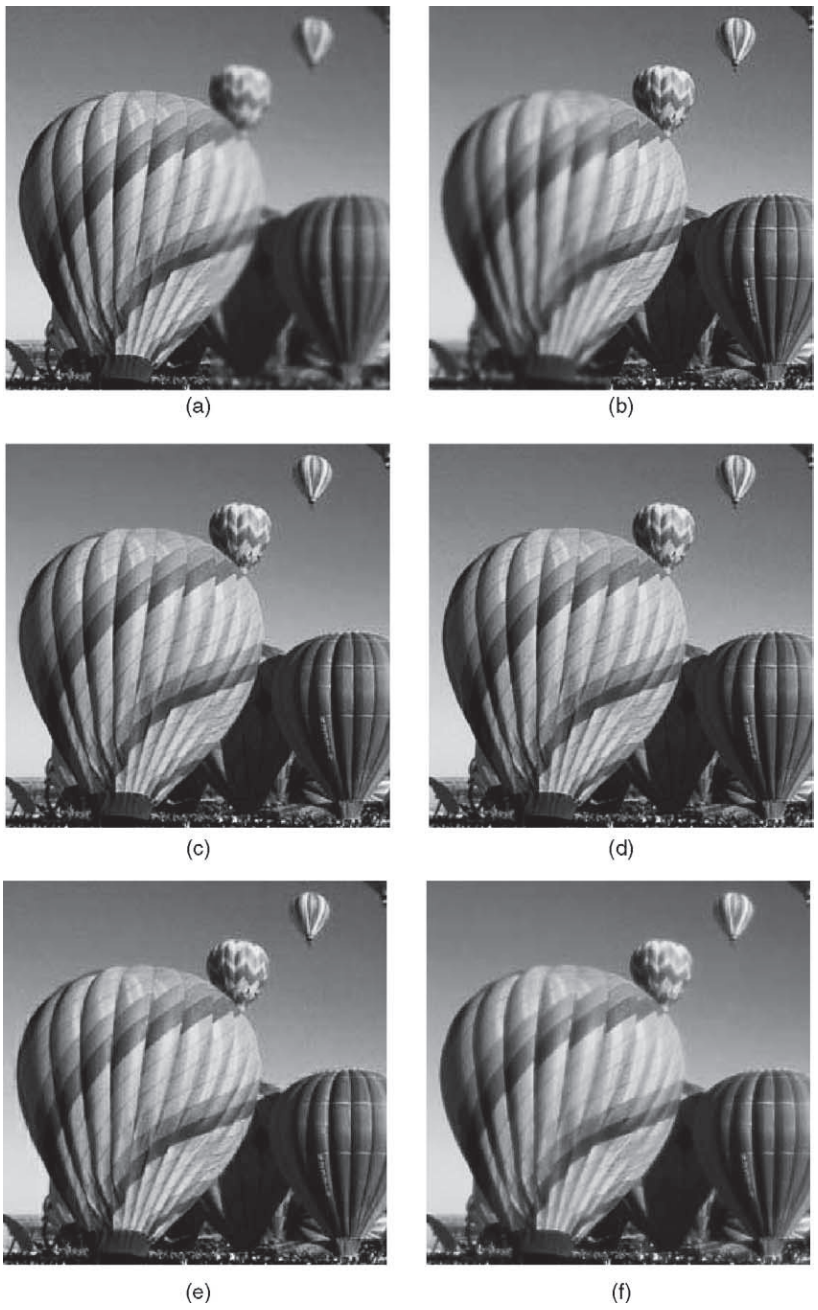

Fig. 7. Original multifocus images and the fused images by various algorithms. (a) Near focused image, (b) far focused image, (c) fused image with proposed algorithm, (d) fused image with gradient method with majority filtering, (e) fused image with Haar wavelet method, (f) fused image with weighted averaging.

wavelet and weighted averaging and is comparable to fusion by image gradient method with majority filtering. However, this is a subjective measure of quality and may not be universally acceptable. Hence, a quantitative measure is also given. Gradient or derivative operators are useful tools to measure the variation of intensity with respect to immediate neighboring pixel of an image [20]. It is observed that a pixel possesses high gradient value when it is sharply focused. An objective criterion based on this knowledge is suggested to measure the quality of the results. The gradient operator suggested by Roberts [21] is used here. For a set of n multifocus images g_i, $i=1, \ldots, n$, the gradient images G_i, $i=1, \ldots, n$ are obtained first. Then, G_i, $i=1, \ldots, n$ are combined into G by taking the maximum gradient value at each position, i.e.

$$G(x,y) = \max\{G_1(x,y),G_2(x,y),\ldots,G_n(x,y)\} \text{ for all } (x,y)$$

Thus, $G(x, y)$ contains highest gradient at each pixel location and approximates the gradient image when all parts of the scene are sharply focused. Suppose $\tilde{G}$ denotes the gradient of the fused or reconstructed image. Then, more similar G and $\tilde{G}$ are, better is the fusion algorithm. Now, following the usual definition of signal-to-noise ratio, we have devised a simple objective measure of the similarity

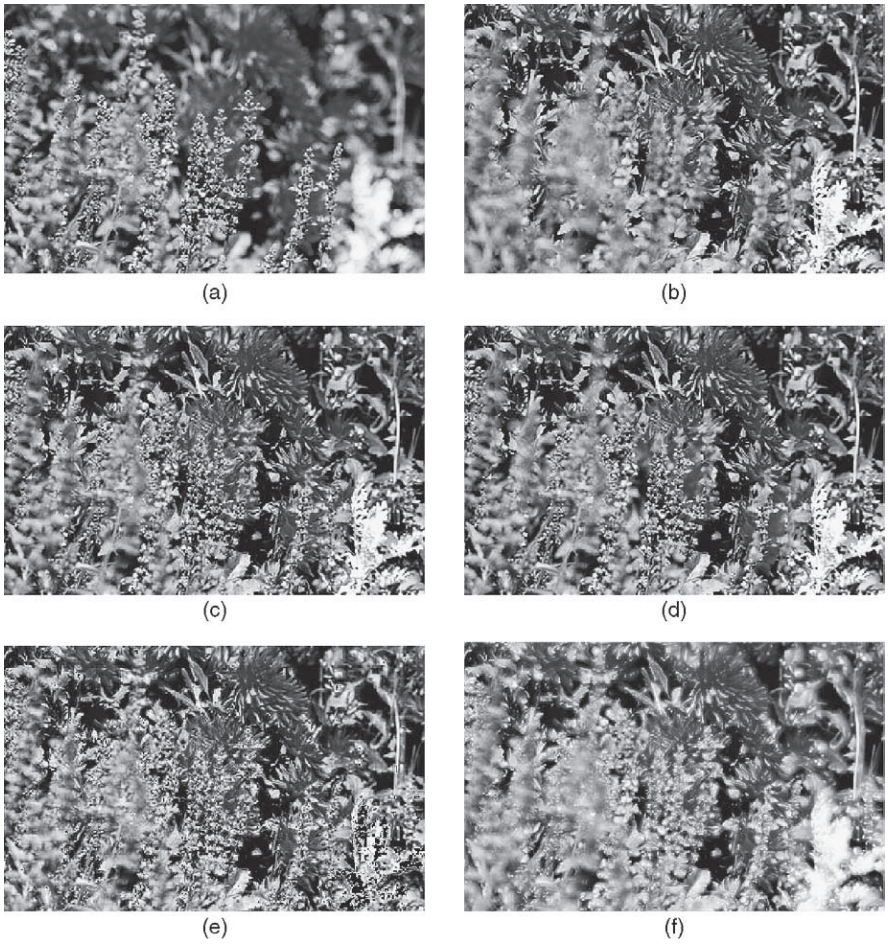

Fig. 8. Original multifocus images and the fused images by various algorithms. (a) Near focused image, (b) far focused image, (c) fused image with proposed algorithm, (d) fused image with gradient method with majority filtering, (e) fused image with Haar wavelet method, (f) fused image with weighted averaging.

between two images as:

$$S(G,\tilde{G}) = 1 - \frac{\sqrt{\sum (G-\tilde{G})^2}}{\sqrt{\sum G^2} + \sqrt{\sum \tilde{G}^2}} \tag{9}$$

Here, $\sqrt{\sum (G-\tilde{G})^2}$ determines the error or dissimilarity between the images and it is normalized by the quantity to make the measure unbiased to any of the images. So for an ideal fused image S approaches the value 1. We have computed similarity values for the results obtained by various fusion techniques. These values are listed in Table 2. They supports the quality observed through manual inspection.

Table 2
Similarity between maximum gradient and fused gradient images (higher the value better is the fusion algorithm)

Figure	Multiscale morphology	Gradient with MF	Haar wavelet	Weighted average
Fig. 5	0.88823	0.89081	0.82479	0.83949
Fig. 6	0.83967	0.85928	0.81018	0.67054
Fig. 7	0.93648	0.93730	0.90801	0.79241
Fig. 8	0.85838	0.84934	0.82032	0.80632

4. Conclusion

In this work, we have proposed an algorithm for fusing images captured with different focus regions. As a result, the effective depth of field of the camera is enhanced considerably. After registering the multifocus images, we have used morphological filters and multiscale top-hat transformation to detect the focused regions in each of them. Finally, the detected regions from the constituent images are put together to reconstruct the fused image. Since, the best-focused regions are detected and copied from one image only, a slight error in registration will have no effect in fusion except in the borders of the focused regions. Hence, this region-based method is robust to misregistration. This method resembles the manual cut-and-paste method of image fusion, which is often used, for comparison purposes. Thus, the fused image obtained by the method is very similar to the ideal image focused in all the regions. Performance analysis reveals that our method is superior to fusion by Haar wavelet and weighted averaging methods and is comparable to that by image gradient with majority filtering. However, computational cost of the method is higher than that of the methods by Haar wavelet and weighted averaging and is comparable to the image gradient method with majority filtering.

Acknowledgements

We gratefully acknowledge the sources of data we used in our experiment. The multifocus source images of Figs. 5–8 are taken from the following websites http://www.geocities.com/~j1hagan/lessons/depthofield.htm, http://www.ece.lehigh.edu/SPCRL/IF/multifocus.htm, http://www.vision.ece.ucsb.edu/sitaram/research.htm, http://www.photozone.de/4Technique/compose/dof.htm

References

[1] P. Perlman, Basic Microscope Techniques, Chemical Publishing Company, New York, NY, 1971.
[2] S. Bradburn, W.T. Cathey, E.R. Dowski Jr., Applications of extended depth of focus technology to light microscope systems, www.colorado.edu/isl/papers, 1998.
[3] W. Seales, S. Dutta, Everywhere-in-focus image fusion using controllable cameras, Proceedings of SPIE 2905, vol. 11, 1996, pp. 227–234.
[4] I. Bloch, Information combination operators for data fusion: a review with classification, IEEE Transactions on SMC: Part A 26 (1996) 52–67.
[5] H.A. Eltoukhy, S. Kavusi, A computationally efficient algorithm for multi-focus image reconstruction, Proceedings of SPIE Electronic Imaging, June 2003.
[6] P.J. Burt, R.J. Kolezynski, Enhanced image capture through fusion, in: Proceedings of the Fourth International Conference on Computer Vision, (Berlin, Germany), 1993, pp. 173–182.
[7] H. Li, B.S. Manjunath, S.K. Mitra, Multisensor image fusion using the wavelet transform, Graphical Models and Image Processing 57 (3) (1995) 235–245.
[8] Z. Zhang, S. Blum, Image fusion for a digital camera application, Conference Record of the Thirty-Second Asilomar Conference on Signals, Systems and Computers, vol. 1, 1998, p. 603.
[9] X. Yang, W. Yang, J. Pei, Different focus points images fusion based on wavelet decomposition, Proceedings of Third International Conference on Information Fusion, vol. 1, 2000, pp. 3–8.
[10] G.K. Matsopoulos, S. Marshall, J.N.M. Brunt, Multiresolution morphological fusion of mr and ct images of the human brain, IEEE Proceedings Vision, Image and Signal Processing 141 (3) (1994) 137–142.
[11] S. Mukhopadhyay, B. Chanda, Fusion of 2d gray scale images using multiscale morphology, Pattern Recognition 34 (2001) 1939–1949.
[12] L.G. Brown, A survey of image registration, ACM Computing Survey 24 (1992) 325–376.
[13] B. Zitova, J. Flusser, Image registration methods: a survey, Image and Vision Computing, vol. 21, November 2003, pp. 977–1000.
[14] J.P.W. Pluim, J.B.A. Maintz, M.A. Viergever, Mutual-information-based registration of medical images: a survey, IEEE Transactions on Medical Imaging 22 (2003) 986–1004.
[15] M. Ghanbari, The cross-search algorithm for motion estimation, IEEE Transactions on Circuits and Systems for Video Technology 38 (1990) 950–953.
[16] S. Zhu, K.K. Ma, A new diamond search algorithm for fast block-matching motion estimation, IEEE Transactions on Image Processing 9 (2000) 287–290.
[17] M. Chen, P. Yan, A multiscaling approach based on morphological filtering, IEEE Transactions Pattern Analytical Machine Intelligent 11 (1989) 694–700.
[18] P. Maragos, Pattern spectrum and multiscale shape representation, IEEE Transactions on Pattern Analysis and Machine Intelligence 11 (1989) 701–716.
[19] D. Robinson, P. Milanfar, Fundamental performance limits in image registration, IEEE Transactions on Image Processing 13 (9) (2004) 1185–1199.
[20] B. Chanda, D. DuttaMajumdar, Digital Image Processing and Analysis, Prentice Hall of India, New Delhi, 2000.
[21] L.G. Roberts, Machine perception of three dimensional solids in: J.T. Tippet (Ed.), Optical and Electro-optical Information Processing, MIT Press, Cambridge, MA, 1965.