

Pattern Selective Image Fusion for Multi-focus Image Reconstruction*

Vivek Maik, Jeongho Shin, and Joonki Paik

Image Processing and Intelligent Systems Laboratory,
Department of Image Engineering,
Graduate School of Advanced Imaging Science, Multimedia and Film
Chung-Ang University, Seoul 156-756, Korea
paikj@cau.ac.kr

Abstract. This paper presents a method for fusing multiple images of a static scene and shows how to apply the proposed method to extend depth of field. Pattern selective image fusion provides a mechanism for combining multiple monochromatic images through identifying salient features in the source images and combining those features in to a single fused image. The source images are first decomposed using *filter subtract decimate (FSD)* in laplacian domain. The *sum-modified-Laplacian (SML)* is used for obtaining the depth of focus in the source images. The selected images are then blended together using monotonically decreasing *soft decision blending (SDB)*, which enables smooth transitions across region boundaries. The resulting fused image utilizes focus information that is greater than that of the constituent images, while retaining a natural verisimilitude. Experimental results show the performance of the depth of focus extension using consumer video camera outputs.

1 Introduction

Recently, image fusion has become an important research topic in image analysis and computer vision [1, 2, 3]. Image fusion refers to the image processing techniques that produce a new, enhanced image by combining images from one or more sensors. The fused image is then made more suitable for human/machine perception, and for further image processing tasks such as segmentation, feature extraction and object recognition.

In this paper we propose a new pattern selective image fusion method that extends the depth of field of the sensor through the manipulation of multiple images at the same scene. An interesting observation motivating this approach is that, even though any single image may not have the entire scene in focus, the settings of the sensor and sensor optics can usually be adjusted so that at least some portion of the scene has the desired visual quality. The challenge, therefore, is to generate a set of images with varying apertures and focus settings, then to combine these images in to a single result where each scene feature has maximal focus.

*This work was supported by Korean Ministry of Science and Technology under the National Research Laboratory Project and by Korean Ministry of Information and Communication under the Chung-Ang University HNRC-ITRC program.

Prior work on the image fusion process has focused on operating on multiple intensity images based on wavelet and discrete cosine transformations [4, 5, 6] or use of a known camera point spread function (PSF) [5]. Other methods use pyramid based representation to decompose the source image into different spatial scales and orientations [7]. Similar results, although with more artifacts and less visual stability can be achieved with the use of other basis functions [8]. Another technique similar to pyramid representation, have been based on wavelet transform as a means to decompose the image into various sub bands [5, 6]. From the decomposed sub-bands the output is generated through selecting the sub bands that have maximum energy and reconstructing the fused sub-band. This representation, however, has an inherent limitation due to the sensitivity of the wavelet transform to translation and rotation and therefore is not particularly suitable for the fusion of images where, even after registration, residual motion is present.

Pattern selective image fusion method proposed in this paper is mainly composed of four step process: *pyramid construction, feature saliency computation, blending function, and reconstruction of the fused images*. This four-step process provides an overview of how images can be fused with maximum flexibility. The rest of the paper is organized as follows. Existing techniques and problem formulation are described in Section 2. The proposed pattern selective fusion algorithm is described in Section 3. Simulation results and comparisons are shown in Section 4. Finally, concluding remarks are outlined in Section 5.

2 Problem Formulation

The basic crux of the problem is deciding which portions of each image are in better focus than their respective counterparts in the associated frames and combining these regions to form the synthesized extended depth of focus image. In short, due to low pass filtering nature of the modified Bessel function present in the defocused images, the discrimination method of choice invariably involves quantification of high frequency content [1, 3].

Scenes containing large local changes in illumination and objects at greatly varying distances are impossible to image with high image quality throughout the scene [2, 4]. Known methods for adjusting a sensor's integration time and aperture, including methods such as automatic gain control and automatic iris selection, are commonly used to adjust for overall illumination in a scene, but cannot compensate for large local variations in the scene brightness. Likewise, physical limitation in standard sensor optics result in finite depth of field, which enables features within the depth of field to be in focus while scene contents outside the depth of field suffer from progressively increased blurring as the features are further and further from the depth of focus [9].

Another optical phenomenon which presents a possible obstacle to the fusion of differently focused images is image misalignment due to magnification or misregistration. The former occurs when the sensor plane, is moved between frames, thereby changing the effective magnification of the imaged object. On a related note, misregistration can also occur as a result of slight camera movement between frames. In this paper we will focus on the fusion issue and the source images are assumed to be already registered.

3 Proposed Pattern Selective Fusion Algorithm

The problem is solved as follows: Given N images of a static scene obtained at different depth of focus using a stationary camera, it is required to combine the images into a single image that has the maximum information content without producing details that are non-existent in the given images. The approach proposed here selects the most informative image for each local area and blends the selected images to create a new image. The foundation for combining multiple images into a single, enhanced result is the pattern selective fusion process itself. To simplify this discussion, we assume the fusion process is to generate a composite image C from a pair of source images denoted with A and B .

3.1 Pyramid Construction for Image Fusion

The pyramid representation can be used both for assessing the saliency of the source image features, and for the reconstruction of the final image result. The following definitions for the pyramid are used. The fusion method described within this paper use a Laplacian pyramid representation. Laplacian pyramids are constructed for each image using the *filter subtract decimate (FSD)* method [8]. Thus the k^{th} level of the FSD Laplacian pyramid, L_k , is constructed from the corresponding Gaussian pyramid level k based on the relationship.

$$L_k = G_k - wG_k = G_k(1 - w), \quad (1)$$

where w represents a standard binomial Gaussian filter, usually of 5×5 spatial pixels extent. When constructing the FSD Laplacian, due to the decimation process and the fact that w is not an ideal filter, a reconstruction of the original image based on the FSD Laplacian pyramid incurs some loss of information. To partially correct for this effect, an additional correction term is added to the Laplacian. This term is obtained by subtracting the filtered Laplacian from the original Laplacian, and results in the corrected FSD Laplacian given by,

$$\tilde{L}_k = L_k + (1 - w)L_k = (2 - w)(1 - w)G_k. \quad (2)$$

The addition of this term allows the reconstruction to restore some of the frequency information that would be otherwise lost. Throughout this paper, while referring to Laplacian representation of the image, the corrected FSD Laplacian defined above should be assumed.

3.2 Feature Saliency Computation

The feature saliency computation process, expresses a family of functions that operate on the pyramids of both images yielding saliency pyramids. In practice, these functions can operate on the individual pixels or on a local region of pixels within the given pyramid level. The saliency function captures the importance of what is to be

fused. When combining images having different focus, for instance, a desirable saliency measure would provide a quantitative measure that increases when features are in better focus. Various such measures, including image variance, image gradients, have been employed and validated for related applications such as auto focusing [3, 4, 5]. The saliency function only selects the frequencies in the focused image that will be attenuated due to defocusing. Since defocusing is a low pass filtering process, its effects on the image are more pronounced and detectable if the image has strong high frequency content. One way to high pass filter an image is to determine its Laplacian or second derivative in our case.

$$\nabla^2 L_k = \frac{\partial^2 L_k}{\partial x^2} + \frac{\partial^2 L_k}{\partial y^2}, \tag{3}$$

Also we know that in the case of Laplacian the second derivatives in the x and y directions can have opposite signs and tend to cancel each other. In the case of textured images, this phenomenon may occur frequently and the Laplacian at times may behave in an unstable manner. We overcome this problem by defining absolute Laplacian as

$$\nabla^2 L_k = \left| \frac{\partial^2 L_k}{\partial x^2} \right| + \left| \frac{\partial^2 L_k}{\partial y^2} \right|, \tag{4}$$

Note that the modified Laplacian is always greater or equal in magnitude to the Laplacian. In order to accommodate for possible variations in the size of texture elements, we compute the partial derivative by using a variable spacing between the pixels used to compute the derivatives. Hence a discrete approximation to the modified Laplacian is given by,

$$ML(i, j) = |2I(i, j) - I(i - 1, j) - I(i + 1, j)| + |2I(i, j) - I(i, j - 1) - I(i, j + 1)|, \tag{5}$$

Finally, the focus measure at a point (i, j) is computed as the sum of modified Laplacian values, in a small window around (i, j) , that are greater than a threshold value.

$$F(i, j) = \sum_{x=i-N}^{i+N} \sum_{y=j-N}^{j+N} M_k(x, y) \text{ for } M_k(x, y) \geq T_1. \tag{6}$$

The parameter determines the window size used to compute the focus measure. In contrast to auto focusing methods, we typically use a small window of size, i.e. $N = 1$. The above equation can be referred to as *sum modified Laplacian (SML)*.

3.3 Soft Decision Blending and Reconstruction

The reconstruction process, operates on each level of the pyramid of the original images in conjunction with sum modified Laplacian to generate the composite image C.

The reconstruction process iteratively integrates information from the lowest to the highest level of the pyramid as follows:

$$L_{ck} = F_k \cdot L_{Ak} + (1 - F_k) L_{Bk}, \tag{7}$$

$$C_k = L_{ck} + w[C_k + 1] \uparrow 2. \tag{8}$$

Where C_k represents the reconstructed image from level N , the lowest level, to level k and $\uparrow 2$ refers to the expand process. The expansion process consists of doubling the width and height of the image by introducing columns and row in the original and then convolving the resulting image by the w filter. A typical problem that can occur with any type of image fusion is the appearance of unnatural borders between the decisions regions due to overlapping blur at focus boundaries. To combat this, *soft decision blending (SDF)* can be employed using smoothing or low pass filtering of the saliency parameter F_k . In this paper Gaussian smoothing has been used for obtaining the desired effect of blending. This creates weighted decision regions where a linear combination of pixels in the two images A and B are used to generate corresponding pixels in the fused image C. Then we have,

$$L_{ck} = \begin{cases} L_{Ak}, \tilde{F}_k < l, \\ L_{Bk}, \tilde{F}_k > h, \\ \tilde{F}_k \cdot L_{Ak} + (1 - \tilde{F}_k) L_{Bk}, \text{ otherwise.} \end{cases} \tag{9}$$

where \tilde{F}_k is now a smoothed version of its former self.

4 Experimental Results

In this section we experimentally demonstrate the effectiveness of the pattern selective fusion algorithm. Experiments were performed on a 256-level image of size 640x480. Here, each image contains multiple objects at different distances from the camera. Thus one or more objects naturally become out of focus when the image is taken. For example, the focus in on the clock in fig 1(a), while that in fig 1(b) is on the student.

Fig 1 and 2 shows the result of image fusion applied to two images having different depth of focus. The resulting composite merges the portions that are in focus from respective images. The result of salient focus measure is given in fig 1 (c), (d). The SML salient operator enhances the sharp edges and textures and makes them predominant for further selection and reconstruction. Fig 1(e) and 2(c) give the result of the patter selective fusion algorithm. The arbitrary threshold value in the range [0, 30], provides acceptable results in most cases. It can be seen the all objects are in focus in the final fused image and any undesired discontinuities at the region boundaries are prevented by soft decision blending function.

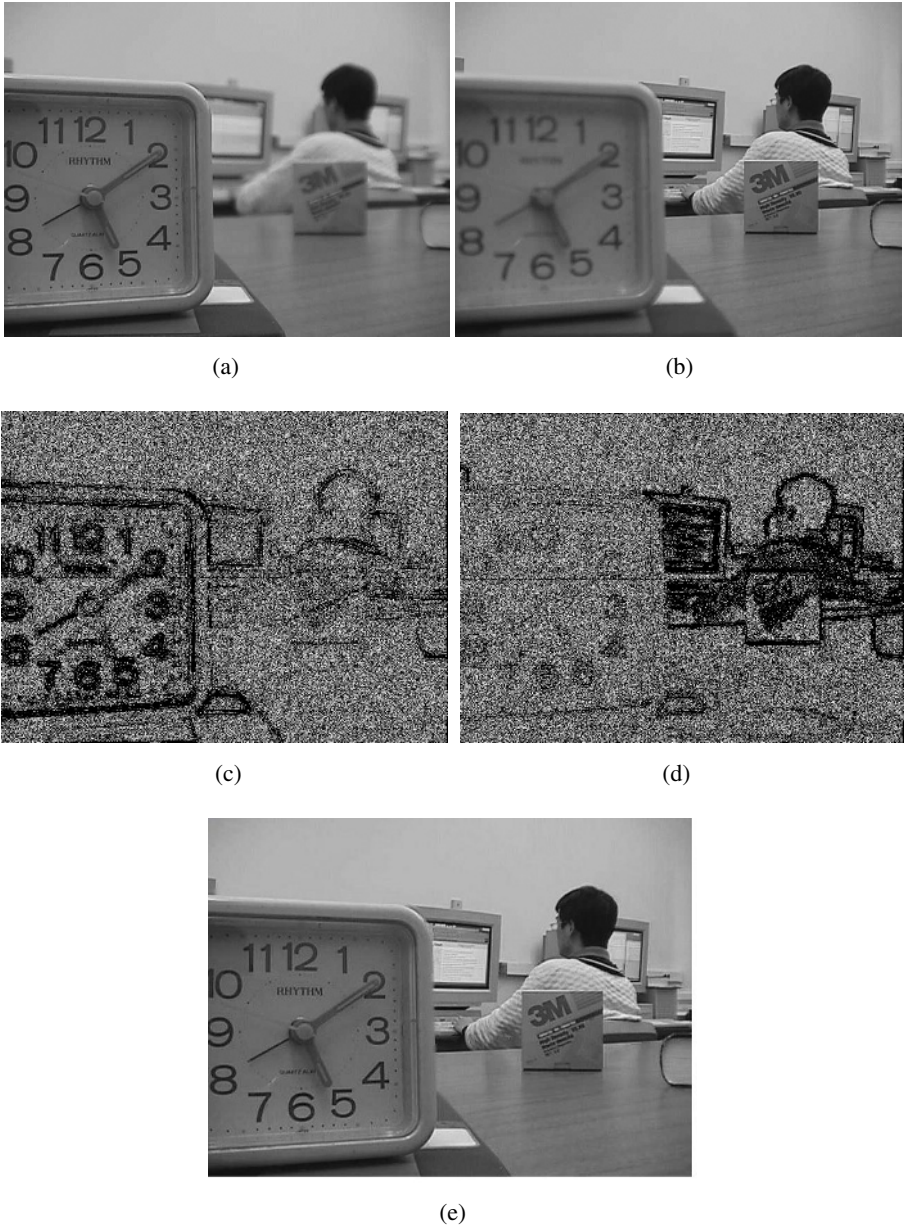


Fig. 1. The “Lab” source images and fusion results: (a) focus on the clock; (b) focus on the student; (c), (d) focus measure using *sum-modified-Laplacian*; (e) fused image with all objects in focus

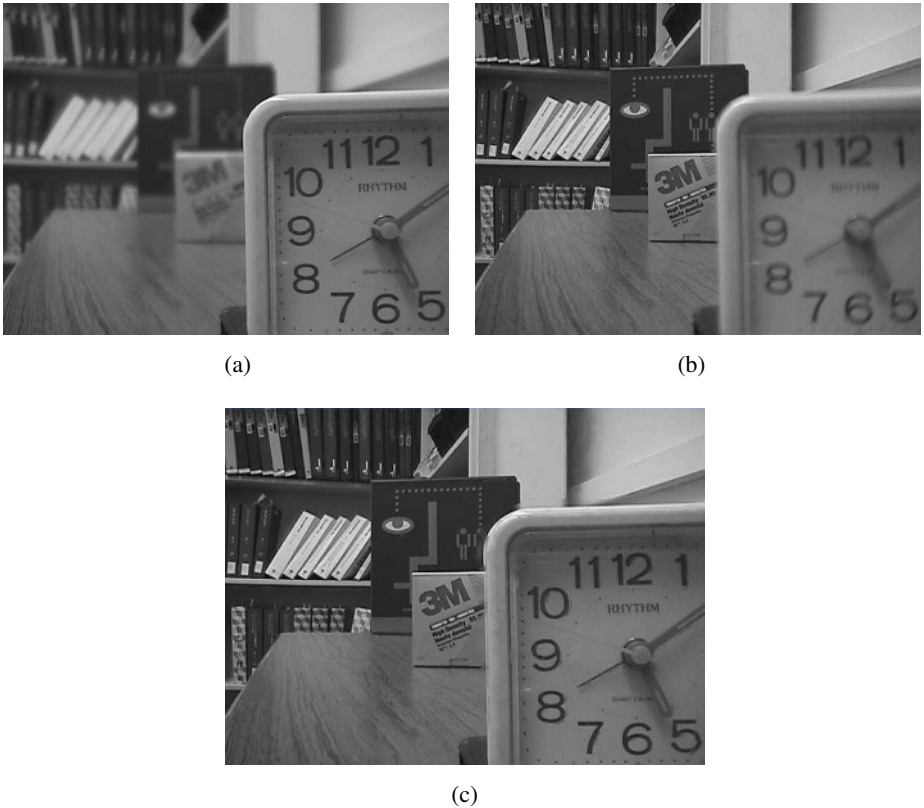


Fig. 2. The “Disk” source images and fusion results: (a) focus on the left; (b) focus on the right; (c) fused image

5 Conclusions

In this paper we proposed a pattern selective fusion algorithm for the synthesis of extended depth of focus imagery. The pyramid structure followed by sum-modified-Laplacian and soft decision blending make the algorithm effective and easy to implement. Also, no particular characteristics of the imaging is need to be known a priori; the main requirements are that underlying assumptions governing the fusion process. A natural extension to the work presented here will include the application of pattern selective fusion to image sequences and also for dynamic range enhancement.

References

- [1] J. K. Aggarwal, *Multisensor Fusion for Computer Vision*, Springer, Berlin, 1993.
- [2] A. Akerman, “Pyramid techniques for multisensor fusion,” *Proc. SPIE*, vol. 2828, pp. 124-131, 1992.

- [3] Z. Zhang and R. S. Blum, "A categorization of multiscale-decomposition-based image fusion schemes with a performance study for a digital camera application," *Proc. IEEE* vol. 87(8), pp. 1315-1326, 1999.
- [4] S. K. Kim, S. R. Park, and J. K. Paik, "Simultaneous out-of-focus blur estimation and restoration for digital AF system," *IEEE Trans. Consumer Electronics*, vol. 44, no. 3, pp. 1071-1075, August 1998.
- [5] G. Ligthart and F. Groen, "A comparison of different autofocus algorithms," *IEEE Int. Conf. Pattern Recognition*, pp. 597-600, 1992.
- [6] K. Aizawa, K. Kodama, and A. Kubota, "Producing object-based special effects by fusing multiple differently focused images," *IEEE Trans. Circuits, Systems for Video Technology*, vol. 10(2), pp. 323-330, 2000.
- [7] X. Yang, W. Yang, and J. Pei, "Different focus points images fusion based on wavelet decomposition," *Proc. Int. Conf. Information Fusion*, pp. 3-8, 2000.
- [8] S. Li, J. T. Kwok, and Y. Wang, "Combination of images with diverse focuses using the spatial frequency," *Int. Journal, Information Fusion*, vol. 2, no. 3, pp. 169-176, September 2001.
- [9] W. B. Seales and S. Dutta, "Everywhere-in-focus image fusion using controllable cameras," *Proc. SPIE*, vol. 2905, pp. 227-234, 1996.