Research Journal of Applied Sciences, Engineering and Technology 6(24): 4602-4606, 2013
ISSN: 2040-7459; e-ISSN: 2040-7467

Submitted: January 31, 2013 Accepted: March 02, 2013 Published: December 25, 2013

A Visual Attention Model Based Image Fusion

[1]Rishabh Gupta, [2]M.R.Vimala Devi and [2]M. Devi
[1]School of Electrical Sciences, VIT University, Vellore
[2]Department of ECE, SASTRA University, Thanjavur, Tamilnadu 613401, India

Abstract: To develop an efficient image fusion algorithm based on visual attention model for images with distinct objects. Image fusion is a process of combining complementary information from multiple images of the same scene into an image, so that the resultant image contains a more accurate description of the scene than any of the individual source images. The two basic fusion techniques are pixel level and region level fusion. Pixel level fusion deals with the operations on each and every pixel separately. The various pixel level techniques are averaging, stationary wavelet transforms, discrete wavelet transforms, Principal Component Analysis (PCA). But because of less sensitivity to noise and mis-registration, the region level image fusion is an emerging approach in the field of multifocus image fusion. The most appreciated approaches in region-based methods are multifocus image fusion using the concept of focal connectivity and spatial frequency. These two methods works well on still images as well as on video frames as inputs. A new region based technique is been proposed for the multifocus images having distinct objects. The method is based on the visual attention models and results obtained are appreciating for the distinct objects input images. The Proposed method results are highlighted using tenengrade and extended spatial frequency as performance parameters by taking several pairs of multi-focus input images like microscopic images, forensic images and video frames.

Keywords: Focal connectivity multifocus, misregistration, pixel level fusion, tenengrade, wavelet transform

INTRODUCTION

The objective of multifocus image fusion is to combine the source images of the same scene to form one composite image that contains the more accurate description of the scene than any of the individual image. In applications of digital cameras, when a lens focuses on a subject at a certain distance, all subjects at that distance are sharply focused. Subjects not at the same distance are out of focus and theoretically are not sharp. Multifocus image fusion consists of two critical steps. First, is to identification of the focussed and unfocussed region in the source images and the second, is to extract focussed region from the source images and combine them to form all focussed image. All multifocus image fusion techniques can be characterized under, pixel by pixel image fusion method and region based image fusion method (Hui *et al*., 1994).

Pixel by pixel image fusion: Pixel by Pixel Image fusion Method (PBM) involves operation on each and every particular image pixel. The simplest image fusion method just takes the pixel-by- pixel gray level average of the source images. This, however, often leads to undesirable side effects such as reduced contrast. Other pixel by pixel image fusion techniques involve multiscale transforms which are very useful for analyzing the information content of images for fusion purposes. Various methods based on the multiscale transforms have been proposed, such as Laplacian pyramid-based, gradient pyramid-based, ratio-of-low-pass pyramid-based, Discrete Wavelet-based (DWT) (Sasikala and Kumaravel, 1994). The basic idea is to perform a multiresolution decomposition on each source image, then integrate all these decompositions to form a composite representation and finally reconstruct the fused image by performing an inverse multiresolution transform. However, one limitation of pixel or decomposition coefficients based methods is that they are sensitive to noise or misregistration. And thus the region based image fusion techniques comes into picture (Yufeng *et al*., 2007).

Region based image fusion: A region based image fusion (RBM) is based on the principle of seeing an image as a combination of different objects present in it just like a human eye perceives an image. So this method involves the identification of various objects present in source images separately and then with operations trying to identify that object as focussed or unfocussed among the various images of the source. The region based image fusion consist of image segmentation and then image fusion (Shutao *et al*., 2001; Fred, 2004; Hariharan *et al*., 2007). Image segmentation will be the most crucial step as the efficiency of any image fusion algorithm depends on

Corresponding Author: M. Devi, Department of ECE, SASTRA University, Thanjavur, Tamilnadu 613401, India

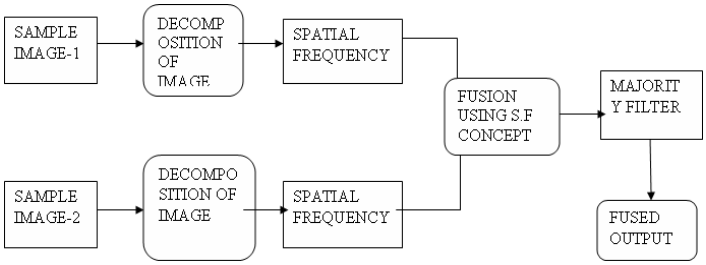


Fig. 1: Block diagram of block method using spatial frequency concept

the proper segmentation of the image into various objects present in it. For image segmentation region growing method or edge detection can be used. But the performance from these methods doesn't look promising. A method proposed in this study is visual attention estimator based on the gray value. Method is able to identify the accurate objects if they are distinct in the sample images. The second step is of decision making. The object images with higher spatial frequency value are combined together to form a focused image.

Image registartion: In practice, the images are usually captured by a handheld or mobile camera and fusion of the images requires registration of the images prior to fusion. It is the process of spatially aligning two or more images of a scene. The processing brings into correspondence individual pixels in the images. Therefore, given a point in one image, the registration processing will determine the positions of the same point in other image.

In this study, an efficient visual attention model based image fusion is proposed and its performance is highlighted using two performance parameters.

REGION BASED IMAGE FUSION METHODS

Most of the present study in multifocus image fusion is been done using region based approach.

Block method using spatial frequency concept: Decompose the source images A, B with size M×N. Denote the ith block of A, B as Ai and Bi. Compute the spatial frequency of each block. Now spatial frequency of the corresponding blocks from A and B is compared and the block with the highest spatial frequency value is to be selected as it is in shutao li research study. The image thus formed consists of pixels unaltered from either of the source images. The block diagram of this method is depicted in Fig. 1.

- **Spatial frequency concept:** Spatial Frequency (SF) is a numerical value, which can be calculated for a whole image or even for each and every pixel present in the image. Spatial frequency tells about how much the image will be perceivable to human eye. Higher the value of the spatial frequency the more perceivable it would be. For source images spatial frequency value is to be calculated block wise. So we should always select a region with a high spatial frequency value. The formulas used to calculate spatial frequency:

$$RF=\sqrt{\sum_{i=1}^{M}\sum_{j=2}^{N}[I(i,j)-I(i,j-1)]^2/(MN)} \quad (1)$$

$$CF=\sqrt{\sum_{i=2}^{M}\sum_{j=1}^{N}[I(i,j)-I(i-1,j)]^2/(MN)} \quad (2)$$

$$SF=\sqrt{RF^2+CF^2} \quad (3)$$

The image thus formed consists of pixels unaltered from either of the source images. Thus the fused image consists of the best details from the sample images without any change in their gray level value and thus would be a highly focused image. The fusion process should be tested for different decomposition sizes having different values of M and N. These values should neither be too small nor be comparable to the original size image. The values of M and N should be taken small if the image contains more details and should be bit high if the image has few objects. The performance of the above method depends on the right selection of M and N.

- **Majority filter:** Majority filter is introduced to correct and verify the results of the fusion obtained by using spatial frequency. Specifically if a center block comes from A and majority of the blocks around it are from B , then in this case center block will be replaced by the corresponding block from B and vice versa. This step increases the accuracy of the method by introducing one more check point.

Image fusion using focal connectivity: Focal Connectivity (FC) is established by isolating regions in an input image that fall on the same focal plane. This method uses focal connectivity and does not rely on physical properties like edges directly for segmentation. Method establishes sharpness maps to the input images, which are used to isolate and attribute image partitions to input images.

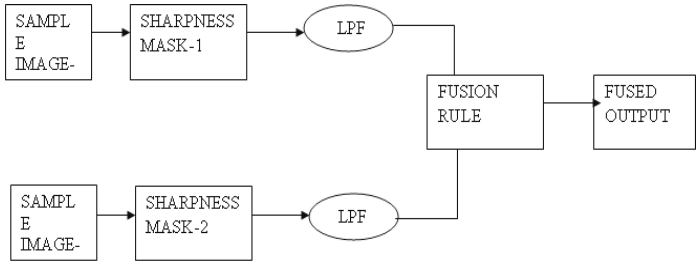


Fig. 2: Block diagram of focal connectivity method

- **Sharpness mask:** Sharpness map is calculated for every input image I{i} (x, y) V i = 1,2,…, N. As a precursor to this step, the images are filtered with sobel masks to approximate horizontal and vertical gradients, Ix{i}(x, y) and Iy{i}(x, y) respectively, where the subscripts x and y denote directional gradient operations. These are used to calculate the sharpness maps Si(x, y)'s for each of the N input images by, isolate and attribute such partitions to one particular input image. The chosen partition is in better focus than its relative counterparts from all the input images. Sharpness mask will be given by:

$$S(x, y) = [Ix(x,y)^2 + Iy(x,y)^2]^{1/2} \quad (4)$$

To make the system less vulnerable to fluctuations (e.g., noise), optics (e.g., magnification and side lobes), local contrast and illumination at the scene we low pass filter the sharpness maps. This increases the accuracy of the decisions to follow by ensuring that areas with better focus influence the decision of its neighbours.

- **Fusion rule:** The sharpness maps are examined for regions of higher focus with their respective counterparts. When the sharpness map of input image I{i}(x, y), of N input images, is compared with its N-1 counterparts, one focally linked region, P{i}(x, y) is isolated by:

$$P\{i\}(x, y) = S\{i\}(x, y) > S\{k \neq i\}(x, y) \quad (5)$$

The union of the such partitions, P{i}(x, y)'s, form the fused image space. Since, here sharpness maps will be able to differentiate the subtle differences in the source images this method works quite well for the video frames as input. This technique as depicted in Fig. 2.

Visual attention model and spatial frequency method: First the image is segmented into different objects present in it using the concept of visual attention model and then using the concept of spatial frequency the focused objects are selected from the source images and then combined together to make a all focused fused image.

- **Visual attention map:** Steniford model: The model of Visual Attention (VA) proposed by Stentiford henceforth referred to as the Stentiford model of visual attention. It functions by suppressing areas of the image with patterns that are repeated elsewhere. As a result flat surfaces and textures are suppressed while unique objects are given prominence. Regions are marked as high interest if they possess features not frequently present elsewhere in the image. The result is a visual attention map. The visual attention map generated tends to identify larger and smoother salient regions of an image.
- **Spatial frequency:** For each and every object segmented above spatial frequency is calculated (Shutao *et al*., 2001). The value of spatial frequency of the same object among the various source images are compared and the object is choose from that source image which give the highest spatial frequency value for that object. The partitions are mosaiced seamlessly to form the fused image.

RESULTS AND DISCUSSION

The results are explained quantitatively by using tabular columns of performance measurement parameter. The algorithms have been implemented using MATLAB 7.1.

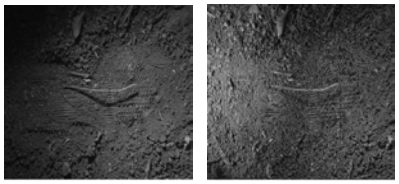
Performance measurement parameters:

Extended spatial frequency value: This parameter tells how much the image is perceivable to human eye. Higher value of it the better the image will be. Higher the value of it tells higher the information content of the image.

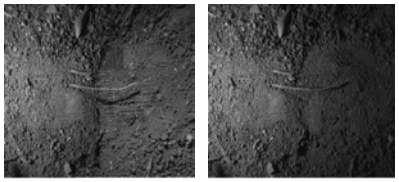
$$RF = \sqrt{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=2}^{N}[I(i,j) - I(i,j-1)]^2}, \quad (6)$$

$$CF = \sqrt{\frac{1}{MN}\sum_{j=1}^{N}\sum_{i=2}^{M}[I(i,j) - I(i-1,j)]^2}, \quad (7)$$

$$MDF = \sqrt{w_d \cdot \frac{1}{MN}\sum_{i=2}^{M}\sum_{j=2}^{N}[I(i,j) - I(i-1,j-1)]^2} \quad (8)$$



Input 1 Input 2



SF method FC method

Fig. 3: Results-forensic image; (a): Input1; (b): Input1; (c): SF method; (d): FC method

Input1 Input2 SF method

FC method VA method

Fig. 4: Results: tower image

$$SDF = \sqrt{w_d \cdot \frac{1}{MN} \sum_{j=1}^{N-1} \sum_{i=2}^{M} [I(i,j) - I(i-1,j+1)]^2} \quad (9)$$

$$SF = \sqrt{(RF)^2 + (CF)^2 + (MDF)^2 + (SDF)^2} \quad (10)$$

So extended spatial frequency not only counts changes in horizontal and vertical directions but also in diagonals.

Tenengrade parameter: Tenengrade tells about the sharpness of the image. So higher the value of this parameter higher will be the detail components in the image which in turn implies better the fused output image:

$$T = \sum_{i=1}^{m} \sum_{j=1}^{n} \sqrt{F_x^2(x,y) + F_y^2(x,y)} \quad (11)$$

Table 1: Performance analysis (footprint image)

Methods applied	EXT. spatial frequency	Tenengrade parameter
PCA	43.877	9.66×10^4
WT (max activity)	33.510	6.53×10^4
Focal connectivity	62.608	1.33×10^5
Block mothed+S.F	53.862	1.2×10^5

Table 2: Performance analysis (tower image)

Method applied	EXT. spatial frequency	Tenengrade parameter
Visual attention model	11.5978	1.82×10^4
Block method+SF	11.4549	1.48×10^4

Fx = Mask operation in x-direction.
Fy = Mask operation in y-direction.

Results: Forensic Image (Fig. 3)
Tower image (Fig. 4)

DISCUSSION

Visual attention method: Principal Component analysis and Discrete wavelet transform are taken as comparison methods. The performance of spatial frequency and focal connectivity using forensic image as input is quantatively analyzed in Table 1. The performance of the proposed technique visual attention based Multifocus image fusion is highlighted using Tower image in Table 2.

CONCLUSION

The implemented techniques are tested on a wide range of images and the results imply that:

- For video frames as input, Focal connectivity is the best method.
- For sample images having distinct objects, Visual attention model will be the most efficient method.
- For general multifocus images, both Focal connectivity and Block method using spatial frequency are giving the best fused outputs.RBM gives better results than PBM for any type of images and video frames but on the cost of:
- Higher complexity
- More execution time
- Requirement of human interference for the best results. So need of the hour is to develop fast and less complex [RBM], which can be easily and efficiently implemented in real time applications.

REFERENCES

Fred, S., 2004. A visual attention estimator applied to image subject enhancement and colour and grey level compression. Proceedings of the 17th International Conference on Pattern Recognition (ICPR'04), 3: 638-641.

Hariharan, H., A. Koschan and M. Abidi, 2007. Multifocus image fusion by establishing focal connectivity. Proceedings of IEEE Computer Society Conference on Computer and Pattern Recognition (ICIP-2007), San Antonio, TX, 3: 321-324.

Hui, L., B.S. Manjunath and K.M. Sanjit, 1994. Multi sensor image fusion using wavelet transform. Graph. Mod. Im. Proc., 57(3): 235-245.

Sasikala, M. and N. kumaravel, 1994. A comparitve analysis of feature based image fusion methods. Inform. Technol. J., 6(8): 1224-1230, ISSN: 1812-5638.

Shutao, L., T.K. James and W. Yaonan, 2001. Combination of images with diverse focuses using spatial frequency. Inform. Fusion, 2(3): 169-176.

Yufeng, Z., A.E. Edward, H. Bruce and M.H. Andrew, 2007. A new metric based on spatial frequency and its application to dwt based fusion algorithm. Inform. Fusion, 8: 177-192.