Real Time Human Visual System Based Framework for Image Fusion

Gaurav Bhatnagar[1], Q.M. Jonathan Wu[2], and Balasubramanian Raman[3]

[1,2] University of Windsor, Windsor, ON, Canada N9B 3P4
[3] Indian Institute of Technology Roorkee, Roorkee-247 667, India
{goravb,jwu}@uwindsor.ca, balaiitr@ieee.org

Abstract. Image Fusion is a technique which attempts to combine complimentary information from multiple images of the same scene so that the fused image is more suitable for computer processing tasks and human visual system. In this paper, a simple yet efficient real time image fusion algorithm is proposed considering human visual properties in spatial domain. The algorithm is computationally simple and implemented very easily in real-time applications. Experimental results highlights the expediency and suitability of the algorithm and efficiency is carried by the comparison made between proposed and existing algorithm.

1 Introduction

In the recent years, operators are getting much more information than ever before due to the development of image sensors. The essential development of sensors has also resulted in the augmentation of the human observer workload. To deal with these situations, there is a strong need of developing an image processing technique which integrate the information from different sensors. At most all the advanced sensors of today, for example, optical cameras, millimetre wave (MMW) cameras, infrared cameras, x-ray imagers, radar imagers etc provide the information in the form of images. As a solution, Image fusion[1] comes to our help and is used very frequently. Hence, image fusion becomes very attractive research topic in the recent years. It greatly improves the capability of image interpretation and the reliability of image judgement which resulted in enhancing the accuracy of classification and target recognition.

Classical approaches of image fusion are based on additive technique[2], Principal Component Analysis(PCA)[3] and high pass filter merger[4]. Now a days, image fusion using the wavelet transform[5,6] has been a popular tool due to its ability to perform the consistency of the spectral and spatial resolution. Additive (also called spatial domain) technique is pixel level image fusion method which is done by taking the pixel-by-pixel average of the source images. In the PCA method, the weighings for each source image are obtained from the eigenvector corresponding to the largest eigenvalue of the source image co-variance matrix. In high pass filter merger, the high pass components of each source image is taken to enhance the local contrast and then fuses high pass components by weighted averaging. The wavelet transform methods are relatively popular due to their better localization property. The simplest way is to decompose all the source images by means of wavelet transform and the transformed parts are then combined using fusion rule decided by fuser followed by the inverse wavelet transform to

A. Elmoataz et al. (Eds.): ICISP 2010, LNCS 6134, pp. 71–78, 2010.

get the fused image. Additive and PCA techniques are less complex as no transform is used. However, these methods often produce undesirable side effects such as block artifacts, reduced contrast etc. High pass filter merger and wavelet based techniques preserve more spectral information but loose more spatial information. However, these techniques are quite expensive and time consuming. In this paper, a spatial domain technique is presented taking visual perception in consideration. It exploits the fact that human visual system is sensitive to local image properties. The Noise visibility function is used in the proposed technique. The fused image that is produced by this scheme presents a visually better representation than the input images. Moreover, the compatibility and superiority of the proposed method is carried out by the comparison made by us with the existing methods.

The rest of the paper is organized as follows. The human visual system model and the proposed technique are explained in section 2. The experimental results are presented in section 3. Finally, the concluding remarks are given in section 4.

2 Proposed Fusion Technique

In this proposed real time framework, the image fusion is performed at the pixel level, which denotes a fusion process generating a single image containing more accurate description and more information than any individual original image. In order to make fused image more suitable for human perception, object detection, target recognition and other computer-processing tasks; the characteristics of human visual system are used. Before going to the proposed framework, first the used human visual system model is described.

2.1 Human Visual System Model: Noise Visibility Function

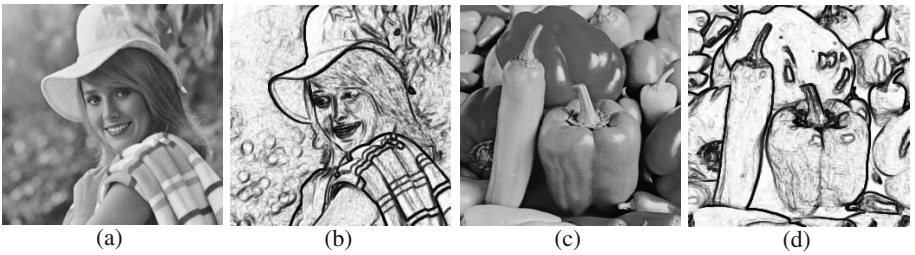
Human visual system (HVS) research offers the mathematical model about how humans see this world. A lot of work has been explored to understand the HVS and then applying in image processing applications. Human visual system based model[7] was first introduced in image compression algorithms. For better quality of the compressed image, these models have played a remarkable role. Now a days, these models are used very frequently in digital watermarking in order to embed an invisible watermark. The watermark is embedded in the sufficient amounts to maximize the strength of the watermark and to guarantee the imperceptibility. Some of the models are modeled specially for image compression and some are for watermarking. One of the popular model is Noise Visibility Function (NVF), which is based on the noise visibility of an image and modeled specially for watermarking by Voloshynovskiy[8]. Considering the original image as a random variable with either non-stationary or stationary generalized Gaussian pdf, the NVF at each pixel position can be written as

$$NVF(i,j) = \frac{1}{1 + \theta\, \sigma_x^2(i,j)} \tag{1}$$

where $\sigma_x^2(i,j)$ is the local variance of the image in a square window centered on the pixel with coordinates (i,j) and defined as

$$\sigma_x^2(i,j) = \frac{1}{S_p \times S_p} \sum_{k=i-\frac{S_p}{2}}^{i+\frac{S_p}{2}} \sum_{l=j-\frac{S_p}{2}}^{j+\frac{S_p}{2}} [x(k,l) - \widetilde{x}]^2 \tag{2}$$

with $\widetilde{x} = \frac{1}{S_p \times S_p} \sum_{k=i-\frac{S_p}{2}}^{i+\frac{S_p}{2}} \sum_{l=j-\frac{S_p}{2}}^{j+\frac{S_p}{2}} x(k,l)$ and θ is a tuning parameter which depends on the image and is given by $\theta = \frac{D}{\sigma_{x,max}^2}$, where $\sigma_{x,max}^2$ is the maximum local variance for a given image and $D \in [50, 100]$ is an experimentally determined parameter.



(a) (b) (c) (d)

Fig. 1. Noise Visibility Function of Lady and Peppers Image

It is clear by definition that NVF characterizes the local image properties and identifies texture, edges and flat regions. Figure 1 shows an example of NVF. Higher values of NVF indicate flat regions whereas smaller values indicate texture regions or regions with edges. Flat regions are the brighter regions whereas texture regions or regions with edges are darker regions in Figure 1(b,d).

2.2 Framework for HVS Based Fusion

In this section, we have discussed some motivating factors in the design of our approach to image fusion. The proposed technique employees in spatial domain. For our convenience, only two source images are considered in our experiment. Let they be f_1 and f_2. For fusion, the basic condition is that the size of all source images is same. Hence without loss of generality, let us consider, source images are of size $M \times N$. The fusion algorithm is given as follows:

- First, NVF of all source images are calculated, denoted by NVF_{f_i}, where $i = 1, 2$ is the number of image.
- Fused all source images as follows:

$$f_{fused}(i,j) = \begin{cases} f_1(i,j) & NVF_{f_1} < NVF_{f_2} \\ f_2(i,j) & NVF_{f_1} > NVF_{f_2} \\ median[f_2(i,j), f_2(i,j)] & NVF_{f_1} = NVF_{f_2} \end{cases} \tag{3}$$

In case of more than two images, two images are fused first and then the fused image is fused with the third one and this process continues till all the images are fused. For the merging of equivalent region, median is used instead of averaging. The main reason

for choosing median is come from statistical theory. According to the statistical theory, averaging has some drawbacks neither it can be determined by inspection nor it can be located graphically, it cannot be determined when one or more data is missing and mainly it is affected very much by extreme values. Where as median is the middle of a distribution: half the values are above the median and half are below the median. The median can be viewed as a value such that the number of data above is equal to the number of data below. Hence, median is a positional average. Moreover, median is determined when one or more data is missing and it is not at all affected by extreme values and this makes it a better measure than the mean. In the case of two data, median coincides with average value of given data.

3 Experimental Results

Some general requirements for fusion algorithm are: (1) it should be able to extract complimentary features from input images, (2) it must not introduce artifacts or inconsistencies according to Human Visual System and (3) it should be robust and reliable. Generally, these requirements are often very difficult to achieve. Hence, first an evaluation index system is established for evaluating the proposed fusion algorithm. These indices are determined according to the statistical parameters. This evaluation index system includes mean, standard deviation, entropy, peak signal to noise ratio (PSNR) and spectral distortion. Among these mean, standard deviation, entropy reflect spatial details information whereas PSNR and spectral distortion reflects spectral information. Mathematical definitions of these indices are given below:

1. *Mean and Standard Deviation:* In statistical theory, mean and standard deviation are defined as follows:

$$\begin{aligned} \hat{\mu} &= \tfrac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} x_{i,j}, \\ \hat{\sigma}^2 &= \tfrac{1}{(M-1)\,(N-1)} \sum_{i=1}^{M} \sum_{j=1}^{N} (x_{i,j} - \hat{\mu})^2 \end{aligned} \tag{4}$$

   where MN is the total number of pixels in the image and $x_{i,j}$ is the value of the ij^{th} pixel.
2. *Entropy:* Entropy is the measure of information quantity contained in an image. If the value of entropy becomes higher after fusion then the information quality will increase. Mathematically, entropy is defined as:

$$E = -\sum_{i=1}^{M} \sum_{j=1}^{N} p(x_{i,j}) \ \ln p(x_{i,j}) \tag{5}$$

   where $p(x_{i,j})$ is the probability of the occurrence of $x_{i,j}$.
3. *Average Gradient:* The average gradient is given by:

$$\bar{g} = \frac{1}{MN} \sum_{i} \sum_{j} \sqrt{\frac{\triangle I_x^2 + \triangle I_y^2}{2}} \tag{6}$$

where $\triangle I_x$ and $\triangle I_x$ are the differences in x and y direction. The larger the average gradient, the sharper the image.

4. *Peak Signal to Noise Ratio:* The PSNR indicates the similarity between two images. The higher the value of PSNR, the better fused image is. Mathematically, PSNR is defined as:

$$PSNR = 10 \lg \frac{255^2}{\frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} [x_{i,j} - y_{i,j}]^2} \tag{7}$$

where MN is the total number of pixels in the image, $x_{i,j}$ and $y_{i,j}$ are the values of the ij^{th} pixel in original and degrade image.

5. *Structural Similarity:* Structural similarity (SSIM) is designed by modeling any image distortion as the combination of loss of correlation, radiometric distortion and contrast distortion. SSIM is defined as:

$$SSIM = \frac{\sigma_{xy}}{\sigma_x \, \sigma_y} \frac{2\mu_x \, \mu_y}{\mu_x^2 + \mu_y^2} \frac{2\sigma_x \, \sigma_y}{\sigma_x^2 + \sigma_y^2} \tag{8}$$

where μ_x, μ_y are mean intensity and σ_x, σ_y, σ_{xy} are the standard deviation. In equation 8, first term is the correlation coefficient between x and y. The second term measure how close the mean grey level is, while third term measure the similarity in contrast of x and y. The higher the value of SSIM, the better fused image is.

The performance of the proposed fusion algorithm is demonstrated using MATLAB by taking different set of gray-scale images as experimental images. In the experiments,

Table 1. Evaluation Indices for Fused Images

Images		Mandrill	Payaso	Pepper	Lab	Medical	Gun
Time Taken	DWT	51.9358	50.7060	52.0318	41.2769	32.0462	31.9971
(in sec.)	Proposed	10.7415	10.6577	9.9431	8.3922	6.6555	6.1712
Mean	Original	129.1438	99.7830	120.2164	122.9372	29.9628	25.6251
	DWT	129.2793	99.4065	119.7593	122.7133	27.2458	31.3983
	Proposed	129.3688	99.6213	119.6156	123.9435	41.1227	52.2069
Standard Deviation	Original	10.3859	10.5246	10.8560	46.0241	38.7900	32.7680
	DWT	10.3703	9.9039	10.7842	9.6350	8.8112	8.0418
	Proposed	42.1657	63.0997	50.9855	47.2410	35.1843	55.2290
Entropy	Original	5.1000	5.2635	5.2635	4.7847	2.8922	3.3344
	DWT	5.0854	5.3771	5.2542	4.8079	4.0565	4.2379
	Proposed	5.1004	5.3678	5.2558	4.9537	4.4472	4.6353
PSNR	DWT	34.0813	33.3704	31.6640	28.2273	15.9483	17.0233
	Proposed	35.3932	33.3838	26.3765	27.6945	15.6526	17.9351
SSIM	DWT	2.7971	3.1574	4.0271	4.8788	27.0301	19.8870
	Proposed	1.8513	2.6500	3.0346	3.6889	26.3006	19.0912

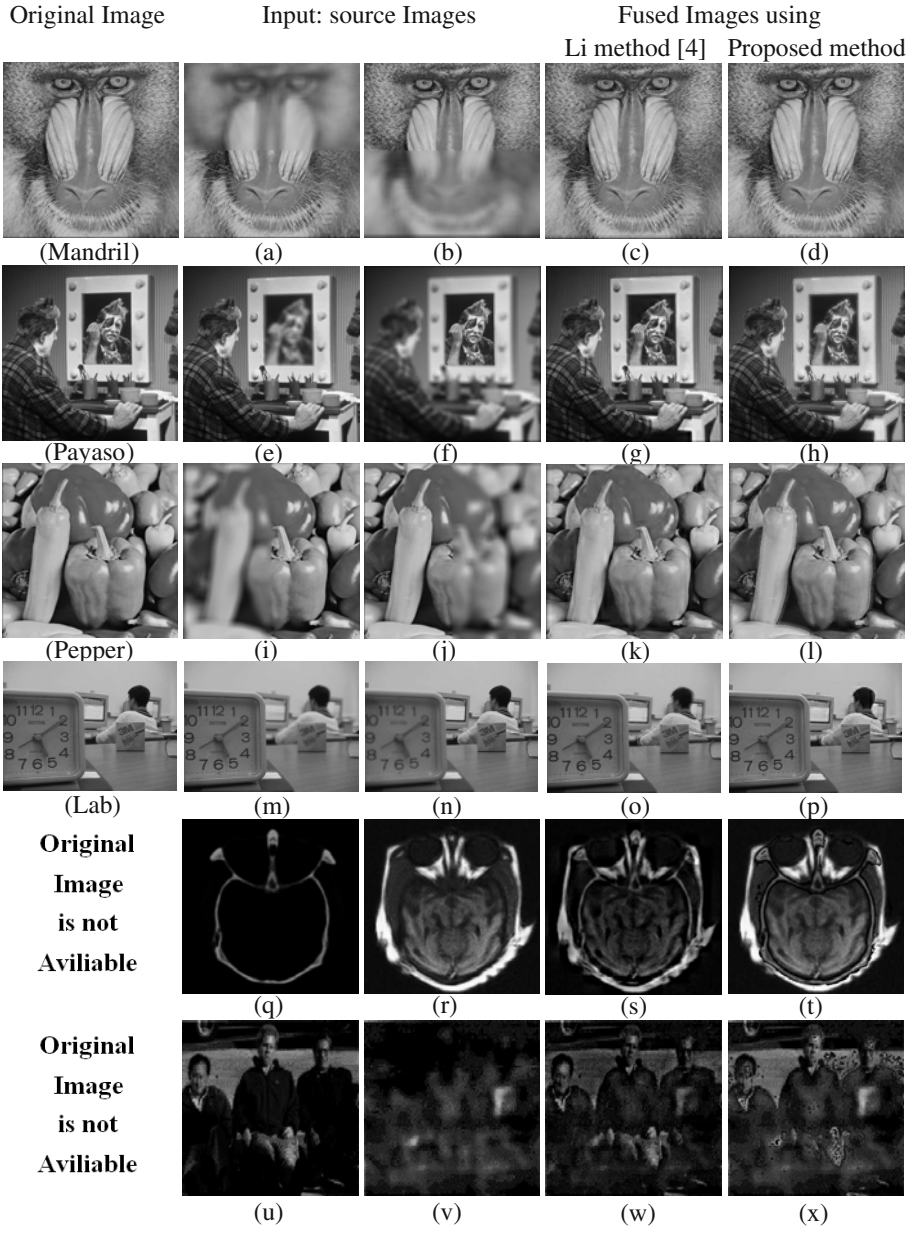


Fig. 2. Results for all experimental Image

mandrill, payaso, peppers, lab, medical images and gun images are taken as experimental images these are having size of 512×512, 512×512, 512×512, 384×288, 256×256 and 256×256 respectively. In mandrill images, we have concentrated on upper half and lower half part. In pepper images, we have concentrated on left half and right half part. In payaso images, we have concentrated on middle and outer part. Lab and medical images are the examples of multi focus images. Finally, gun images are the very famous example of the Concealed Weapon Detection. To evaluate the robustness and superiority of the proposed method, the comparison is made by us with very well known fusion algorithm given by Li [5] which is based on the wavelet transform and activity measure. The activity measure of each coefficient is computed in a 3×3 or 5×5 window and fusion is done by consistency verification along with area based activity measure and maximum selection. For Medical and Gun images, proper original images are not available to compare our results. Hence, we evaluate all indices with the help of all input images and then take the average of all values as the results for the original image. Table 1 shows the values of evaluation indices for original and fused images using Li [5] and proposed methods and visual results are shown in figure 2. The first row of the table shows the time taken in the fusing images of different sizes. It is clear that the time taken by the proposed method is very less. The proposed methods perform better compared to Li *et al.* [5]. To reach this conclusion the main stress is focused on the mean, standard deviation, entropy, PSNR and SSIM which have the highest/lowest values between the compared methods.

4 Conclusions

The fusion method described in this paper cover a large variety of practical applications. In the proposed technique, human visual system is considered and applied in spatial domain which make it well suitable candidate for real time applications. The main benefit is that the use of no transform makes the proposed technique less complex and very easy. Hence, it has very low computation cost. It can do a equivalent/better job than existing wavelet based fusion method[5] as shown by the experiments and comparison made by us. This shows the algorithm robustness and accuracy in fusing several types of real images obtained from different kinds of sensors, for different purposes. Hence, the proposed algorithm will be suitable for implementation in real-time processing.

Acknowledgement

The work is supported by the Canada Research Chair program, the NSERC Discovery Grant.

References

1. Hall, D.L., Llinas, J.: An Introduction to Multisensor Data Fusion. Proceedings of the IEEE 85(1), 6–23 (1997)
2. Xydeas, C. S., Petrovic, V.: Objective Pixel-level Image Fusion Performance Measure. In: Procedding of SPIE, Sensor Fusion: Architectures, Algorithms, and Applications IV, vol. 4051, pp. 89–98. Society of Photographic Instrumentation Engineers (2002)

3. Chavez, P.S., Kwarteng, A.Y.: Extracting spectral contrast in Landsat thematic mapper image data using selective principal component analysis. Photogrammetric Engineering and Remote Sensing 55, 339–348 (1989)
4. Burt, P.J., Kolczynski, R.J.: Enhanced image capture through fusion. In: Proceedings of International Conference on Computer Vision, Berlin, Germany, pp. 173–182. IEEE Press, Los Alamitos (1993)
5. Li, H., Manjunath, B.S., Mitra, S.K.: Multisensor image fusion using the wavelet transform. Graph Models Image Processing 57(3), 235–245 (1995)
6. Qu, G., Zhang, D., Yan, P.: Medical Image Fusion by Wavelet Transform Modulus Maxima. Optics Express 9, 184–190 (2001)
7. Levicky, D., Foris, P.: Human visual system models in digital watermarking. Radioengineering 13(4), 38–43 (2004)
8. Voloshynovskiy, S., Herrigel, A., Baumgartner, N., Pun, T.: A stochastic apporach to content adaptive digital image watermarking. In: Pfitzmann, A. (ed.) IH 1999. LNCS, vol. 1768, pp. 211–236. Springer, Heidelberg (2000)