MULTI-SENSOR IMAGE FUSION USING THE WAVELET TRANSFORM

Hui Li

HNC Software Inc.
5930 Cornerstone Court West
San Diego, CA 92121

B.S. Manjunath and Sanjit K. Mitra

Dept. of Electrical and Computer Engineering
University of California
Santa Barabra, CA 93106

ABSTRACT

In the image fusion scheme presented in this paper, the wavelet transforms of the input images are appropriately combined, and the new image is obtained by taking the inverse wavelet transform of the fused wavelet coefficients. An area-based *maximum* selection rule and a consistency verification step are used for feature selection. A performance measure using specially generated test images is also suggested.

1. INTRODUCTION

With the availability of multi-sensor data in many fields such as remote sensing, medical imaging, machine vision and military applications, sensor fusion has emerged as a new and promising research area. The current definition of sensor fusion is very broad and the fusion can take place at the signal, pixel, feature, and symbol level [1, 2]. In this paper we address the problem of pixel-level fusion or the so-called image fusion problem. Multi-sensor data often presents complementary information about the region surveyed, so image fusion provides an effective method to enable comparison and analysis of such data. The goal of image fusion is to create new images that are more suitable for the purposes of human visual perception, object detection and target recognition. The use of multi-sensor data such as visible and infrared images has led to increased recognition rate in applications such as automatic target recognition [3].

A prerequisite for successful image fusion is that multi-sensor images have to be correctly aligned on a pixel-by-pixel basis. Image registration techniques have been discussed extensively in the literature and we assume here that the images to be combined are already perfectly registered. The simplest image fusion method is to take the average of two input images. However when this direct method is applied, the contrast of features uniquely presented in either of the images is reduced. In order to solve this problem, several Laplacian pyramid based (or variants of it) fusion schemes have been proposed in recent years [4, 5, 6]. The basic strategy here is to use a feature selection rule to construct a *fused* pyramid representation from the pyramid representations of the original data. The composite image is obtained by taking an inverse pyramid transform.

The Laplacian pyramid based image fusion techniques have certain drawbacks. Often the fused images contain blocking artifacts in the regions where the multi-sensor data are significantly different, as we found in our experiments. In contrast, the wavelet transform based approach proposed in this paper produces more naturally fused image even when the images to be combined are very different. An area-based maximum selection rule and a consistency verification step are proposed for feature selection. Better fusion results, both visually and quantitatively based on a proposed performance measure, have been achieved using the wavelet transform.

For details on the wavelet transform interested readers are referred to [7, 8, 9]. Section 2 of this paper describes the proposed image fusion scheme and a performance measure to evaluate different fusion algorithms. Section 3 presents experimental results.

2. THE IMAGE FUSION SCHEME

2.1. The Basic Algorithm

Figure 1 shows the schematic diagram of the basic structure of the proposed image fusion scheme. First the wavelet transforms of the images are computed. The wavelet transform contains the low-high bands, the high-low bands and the high-high bands of the image at different scales. Since larger absolute transform coefficients correspond to sharper brightness changes (and thus the "salient features" in the images), a good integration rule is to select, at every points in the transform domain, the coefficients whose absolute values are higher. In this way the fusion takes place in all the resolution levels and the more dominant features at each

0-8186-6950-0/94 $4.00 © 1994 IEEE

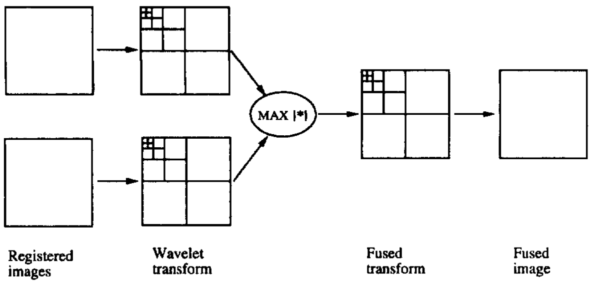


Figure 1: The block diagram of the image fusion scheme.

scale are preserved in the new multiresolution representation. Subsequently a new image is constructed by performing an inverse wavelet transform.

Due to its compactness, orthogonality and the availability of directional information, the wavelet transform can effectively extract salient features at different scales. As a result, the proposed fusion scheme produces better results than the Laplacian pyramid based methods. One important criterion for evaluating an image fusion scheme is the stability of the inverse transform from the fused multiresolution representation. The reconstruction of the Laplacian pyramid can be unstable especially in the regions where the two images appear significantly different. As a result, artifacts such as blocking effects are often visible. However no similar artifacts have been observed when the wavelet transform is adopted for image fusion.

2.2. The Feature Selection Algorithm

Since the useful features in the image usually are larger than one pixel, the pixel-by-pixel maximum selection rule may not be the most appropriate method. In the fusion scheme proposed in [6], an area-based selection rule is used. The images are first decomposed into a gradient pyramid. The variance of each image patch over 3×3 or 5×5 window is computed as an activity measure associated with the pixel centered in the window. If the activity measures at the corresponding locations are close to each other, the average of the two is considered as the new value; otherwise the larger value is chosen.

The computation of the variance can be considered as a nonlinear high-pass filtering operation and the Laplacian pyramid can be considered as a linear high-pass filtering operation. Since the activity measure in [6] is equivalent to the cascade of a linear high pass filter with a non-linear high pass filter, it does not have a clear physical meaning. In our implementation we use the maximum absolute value within the window as an activity measure associated with the center pixel. In this way a high activity value indicates the presence of a dominant feature in the local area. A binary decision map of the same size as the wavelet transform is then created to record the selection results. This binary map is subject to a consistency verification. Specifically in the transform domain if the center pixel value comes from image A while the majority of the surrounding pixel values come for image B, the center pixel value is switched to that of image B. In the implementation, a majority filter is applied to the binary decision map; the map is then negated, followed by the application of a majority filter, and is negated again. A fused image is finally obtained based on the new binary decision map. This selection scheme helps to ensure that most of the dominant features are incorporated into the fused image.

2.3. A Performance Measure

In the current literature there is no quantitative performance measure for evaluating image fusion algorithms. The fusion results are mostly evaluated visually. The problem with defining a quantitative measure lies in the difficulty in defining an ideal composite image based on multi-sensor images or images taken at different times. Although the best criterion should be linked with the specific application, it is still desirable to design a performance measure for the purpose of comparing different algorithms.

A simple test is to create images for which the desired fusion result is known. A suitable candidate is a pair of images containing two objects with different distances to the camera; in one image the first object is focused and the second object is blurred; in the other image the first object is blurred and the second object is focused. The ideally fused image would contain both the well-focused objects and it can be created manually by cut and paste. A performance measure ρ is defined as the standard deviation of the difference image between the ideal image and the fused image:

$$\rho = \sqrt{\frac{\sum_{i=1}^{N}\sum_{j=1}^{N}[I_{pr}(i,j) - I_{fd}(i,j)]^2}{N^2}} \qquad (1)$$

where I_{pr} is the perfect result and I_{fd} is the fused image. One set of multi-focus test images taken at the UCSB Image Processing Laboratory is shown in Figures 4(a) and (b).

It should be noted that this method is restricted to specially created test images, and generally is not applicable to real multi-sensor data where the ideal fusion is ill-defined and can not be obtained manually.

Nevertheless, since the objective of selecting and incorporating most salient features into the fused images is the same as in many computer vision applications, the performance measure with respect to the multi-focus images provides a quantitative comparison of various image fusion techniques.

3. EXPERIMENTAL RESULTS

Figure 2 shows a pair of MRI and PET images, their wavelet transforms, the fused wavelet transform based on the maximum selection rule, and the resulting fused image. The MRI image provides anatomic information while the PET image provides functional information. In the fused image, the relative position of the functional information with respect to the anatomic landmarks is clearly displayed. This information may be very useful for physicians in medical diagnosis.

Figures 3(a) and (b) show a pair of registered Landsat TM image and Spot image. The Spot image has a higher resolution so it contains many finer features such as rivers that are not present in the Landsat image. On the other other hand, the Landsat image also contains some unique features such as the textural patterns. The image obtained by pixel averaging is displayed in Figure 3(c) and it can be seen that the contrast of the unique structures in either the images is reduced. This problem is nicely resolved by using the wavelet transform based image fusion scheme as shown in Figure 3(d). Although the two images look quite different in many regions, there are no artifacts in the fused image.

Figures 4(a) and (b) show a pair of test images containing two clocks with different distances from the camera; and only one clock in either image is in focus. The ideally fused image, obtained by manually putting together the well-focused clocks, is shown in Figure 4(c). Figure 4(d) is the fused image using the wavelet transform and the area-based feature selection scheme. Figure 4(e) is the normalized difference image between the automatically fused image and the manually fused image. The difference image appears to be quite uniform and the performance measure is $\rho = 3.28$. By comparison the performance measures for the direct averaging method and the Laplacian pyramid method are $\rho = 5.44$ and $\rho = 4.67$, respectively. The binary decision map shown in Figure 4(f) displays how the new multiresolution representation is generated from the two input sources. The dark pixels indicate the image in (a) is selected, while the bright pixels indicate the image in (b) is selected. In can be seen that the sharper clocks are favored at all the resolution levels.

The proposed fusion algorithm performed equally well when applied to optical-to-SAR, visible-to-infrared and infrared-to-range image fusion. This algorithm can be straightforwardly extended to the fusion of more than two multi-sensor images.

4. ACKNOWLEDGMENTS

This work was supported in part by a University of California MICRO grant, with matching funds from ESL Inc., Digital Instruments Inc., Tektronix Inc. and Rockwell International, and in part by the Lawrence Livermore National Laboratory. We thank Professor Richard Leahy, University of Southern California, for providing the multi-sensor medical images, and Dr. Eric Rignot, JPL, Caltech, for providing the multi-sensor satellite images used in this research.

5. REFERENCES

[1] R. Luo and M. Kay, "Data fusion and sensor integration: state-of-the art 1990s," in *Data Fusion in Robotics and Machine Intelligence* (M. Abidi and R. Gonzalez, eds.), Academic Press, 1992.

[2] J. Aggarwal, *Multisensor Fusion For Computer Vision.* Springer-Verlag, 1993.

[3] Y. Zhou and R. Hecht-Nielsen, "Target recognition using multiple sensors," in *Proc. IEEE Workshop: Neural Networks for Signal Processing,* pp. 411–420, 1993.

[4] A. Toet, "Hierarchical image fusion," *Machine Vision and Applications,* pp. 1–11, March 1990.

[5] A. Toet, "Multiscale contrast enhancement with application to image fusion," *Optical Engineering,* vol. 31, pp. 1026–1030, May 1992.

[6] P. Burt and R. Lolczynski, "Enhanced image capture through fusion," in *Proc. the Fourth International Conference on Computer Vision,* pp. 173–182, 1993.

[7] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-11, pp. 674–693, July 1989.

[8] I. Daubechies, "Orthonormal bases of compactly supported wavelets," *Commun. Pure Appl. Math.*, vol. 41, pp. 909–960, November 1988.

[9] M. Vetterli and C. Herley, "Wavelets and filter banks: theory and design," *IEEE Trans. Signal Processing*, vol. 40, pp. 2207–2232, September 1992.

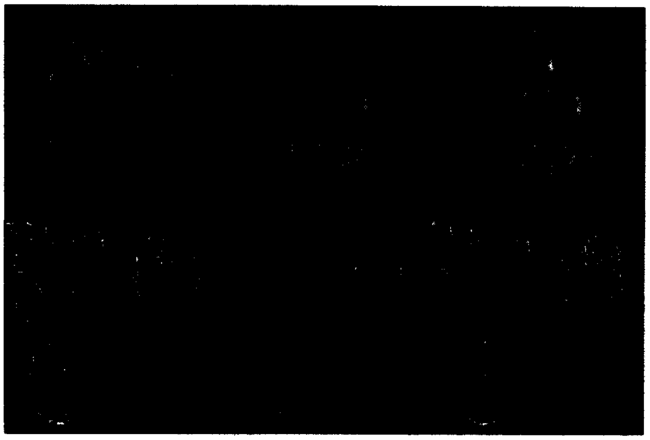
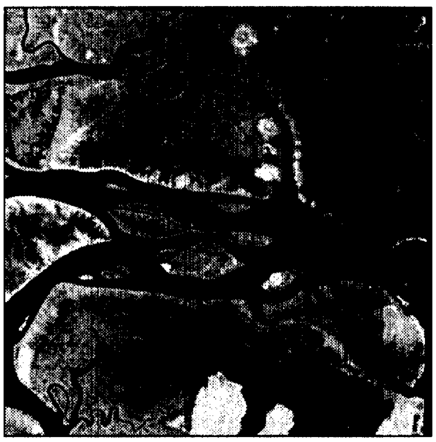


Figure 2: A pair of MRI and PET images, the fused image and their wavelet representations.



(a)

(b)

(c)

(d)

Figure 3: (a) and (b) show a pair of registered Landsat and Spot images, (c) and (d) show the fused images obtained through pixel averaging and the proposed multiresolution scheme, respectively.

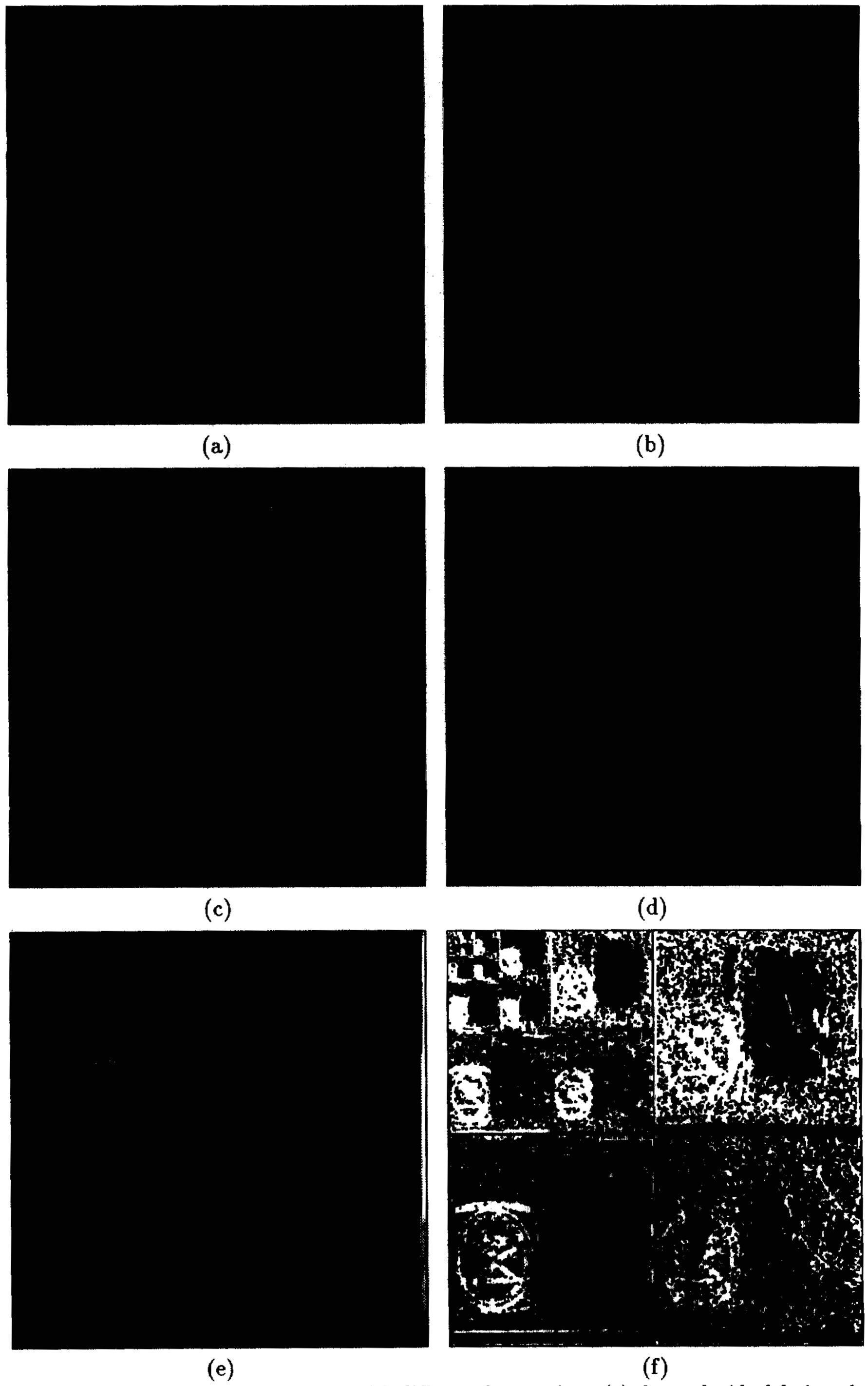


Figure 4: In (a) and (b) a pair of registered images with different focus points, (c) shows the ideal fusion obtained by manual cut and paste, (d) is the fusion result using the wavelet transform, (e) shows the normalized difference image between (c) and (d), and (f) shows a binary decision map corresponding to the outcome of the proposed feature selection rule.