A REGION-BASED TECHNIQUE FOR FUSION OF HIGH RESOLUTION IMAGES USING MEAN SHIFT SEGMENTATION

Li Shuang[a], Li Zhilin[a, b]

[a] LIESMARS, Wuhan University, P.R. China - lishuang129@gmail.com
[b] Dept. of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong - lszlli@polyu.edu.hk

Commission VII, WG VII/6

KEY WORDS: Image processing, Sharpening, Image understanding, Fusion, Geography

ABSTRACT:

This paper describes a region-based technique for fusion of high-resolution images. In this technique, mean shift segmentation is adopted to extract the features for high resolution image as a substitution of other segmentation methods (e.g. Canny operator) and Structure Similarity Index Metric (SSIM) is used to measure the region similarity. Experiments on IKONOS images are carried out to compare the results obtained from this new technique and those with Canny segmentation. It has been found that the results obtained this new technique is much better than the conventional ones in terms of spatial sharpness and spectral reservation.

1. INTRODUCTION

Image fusion is a process to combine two or more different images to form a new image by using certain algorithm (Phol and Genderen,1998). It takes place at three levels: pixel, feature and decision. Image fusion at pixel-level is the lowest processing level considering individual pixels or associated local neighbourhoods of pixels for fusion decision. In the past decades, large number of pixel level image fusion methods is proposed, e.g. Brovey, Intensity-Hue-Saturation, Principle Component Analysis, Wavelet transforms, etc (Zhen and He,2004). However, such methods often introduce colour distortion and/or block effect to high resolution image fusion. This is because a pixel is only a basic unit of information with no semantic significance. At the feature level, features from the input images will be first extracted (e.g. using segmentation procedures); and then fusion of these features will be operated by some rules. Comparing with pixel level image fusion, feature-level method is more meaningful. Because it can fully explore the characteristics of features to guide the image fusion process, such as region activity level, region similarity match measures and so on.

More recently, a number of region-based feature-level image fusion techniques have been proposed (Zhang,1997; Piella,2003; Lewis,2007). These techniques first transform the source images A and B to multi-scale representations by wavelet transforms; segmentation is carried out on the source image to get region representations of both images. Then by overlaying the two region representations, a shared region representation for these two images is obtained. And region activity level and similarity match measurements are calculated from each region to guide the fusion process. During the whole process, segmentation is the most important part because it directly influences the effect of fusion result. Previous work employs the watershed segmentation or Canny edge detection method and the results for fusion of high-resolution images are not very good. Therefore, this study aims to develop a new technique for fusion of high-resolution images.

It has been found (e.g. Mo et al,2006) that the mean shift segmentation is more suitable for the segmentation of high resolution image and thus will be adopted in this study. Moreover, the Structure Similarity Index Metric (SSIM) proposed by Wang,(2002) for image quality assessment will be used (instead of region match measure which is commonly used) to guide the fusion process.

Section 2 reviews the region based fusion. Mean shift segmentation for feature extraction is introduced in section 3. SSIM used for fusion decision making is described in section 4. Section 5 describes the evaluation of the proposed method and conclusions are made in section 6.

2. REGION BASED FUSION: AN OVERVIEW AND A PROPOSAL

The concept of region based fusion was first introduced by Zhang et al.,(1997) and developed by Piella et al.,(2003). Piella’s generic region based image fusion framework is shown in Figure 1.

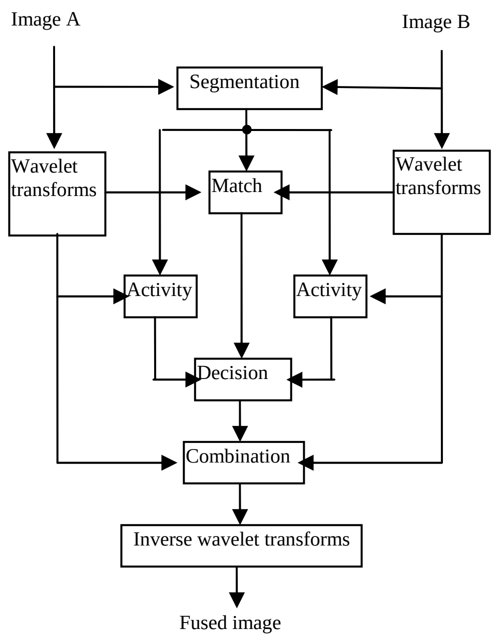

Figure 1. Generic framework of region based image fusion (Piella,2003)

First, the source images are decomposed by wavelet to get the approximate and detailed sub-images; and then segmentation is carried for these sub-images to get the regions of each level. These regions are used to guide fusion process. The activity level and match degree measure of the wavelet coefficients of source images are computed in these regions; and the maximum value rule and the weighted average rule are respectively used to combine the coefficients of detailed sub-images and approximate sub-images. At last, the combination coefficients are inversely transformed by wavelet to obtain the final fusion image.

The choice of segmentation is vitally important because it directly influences the fusion decision. An appropriate segmentation will give useful information to image fusion, while an inappropriate segmentation will provide misleading information to guide the fusion process. Currently, the popular image segmentation methods used in the region based image fusion framework (Zhang,1997; Piella,2003; Wang,2005; Lewis,2005) are c-means clustering, watershed algorithms, and Canny edge detection method. But these segmentation methods can be substituted by others. The selection of appropriate segmentation method is the first issue to be considered.

Moreover, in the traditional region-based fusion framework, the effect of segmenting sub-images will be more serious than that of segmenting original images because sub-images contain lesser information as the number of decomposition level increases. There may be inaccuracy in segmented regions at each level no matter what segmentation methods are used. When inversely transformed by wavelet, the inaccuracy will increase level by level. To reduce the inaccuracy is the second issue to be considered.

Formalization of appropriate rules to guide the fusion process is the third issue to be considered.

To develop a more robust technique for the fusion of high-resolution, in this study, the following strategy is adopted:

- Use of mean shift segmentation to substitute Canny segmentation;
- use of the original input images to get the binary image of shared region and then map of the shared region image to each level by down-sampling to ensure the consistency of segmentation at each level; and
- use of Structure Similarity Index Metric (SSIM) proposed by Wang,(2002, 2004) to guide the fusion process instead of region match measure because SSIM has more physical meanings.

3. MEAN SHIFT SEGMENTATION FOR EXTRACTION OF FEATURES FROM HIGH-RESOLUTION IMAGES

Mean shift analysis is a newly developed nonparametric clustering technique based on density estimation for the analysis of complex feature spaces. It has found many successful applications such as image segmentation and tracking (Comaniciu,1999; Luo,2003).

The mean shift procedure is an adaptive local steepest gradient ascent method. The mean shift vector is computed by the following formula:

$$m_{h,G}(x) = \frac{1}{2} h^2 c \frac{\hat{\nabla} f_{h,K}(x)}{\hat{f}_{h,G}(x)} \quad (1)$$

Where the subscripts G and K are kernels, their corresponding profiles satisfy $g(x) = -k'(x)$; $\nabla \hat{f}_{h,K}$ is the density gradient estimator of kernel K ; $\hat{f}_{h,G}$ is the probability density of new kernel G ; h is the bandwidth and c is a constant; x is the centre of kernel(window).

It indicates that, at location x, the mean shift vector computed with kernel G is proportional to the normalized density gradient estimate obtained with kernel K . Therefore, to get the direction of $\hat{\nabla} f_{h,K}(x)$, only the vector $m_{h,G}(x)$ should be calculated. The mean shift vector thus always points toward the direction of maximum increase in the density.

The mean shift procedure is achieved by a 2-step iteration:

1) Compute the mean shift vector $m_{h,G}(x)$,
2) Translate the kernel (window) $G(x)$ by $m_{h,G}(x)$ until convergence.

Since the control parameter has clear physical meanings, both gray level and color images are processed by the same algorithm. An image is typically represented as a 2-D lattice of p-dimensional vectors (pixels). When p=1, it denotes grey image. When p=3, it denotes color image. When p>3, it denotes multi-spectral image. The space of lattice is known as the spatial domain, while the gray level and multi-spectral information are represented in the color domain. When the location and color vectors are concatenated in the joint spatial-color domain of dimension d=p+2, thus, the multivariate kernel is defined as:

$$K_{h_s,h_r}(x) = \frac{C}{h_s^2 h_r^p} k\left[\left\|\frac{X^s}{h_s}\right\|^2\right] k\left[\left\|\frac{X^r}{h_r}\right\|^2\right] \quad (2)$$

Where X^s is the spatial part and X^r the color part of the feature vector; $k(x)$ is the common profile used in both of the two domains; h_s and h_r are the kernel bandwidths; and C is the corresponding normalization constant. The quality of segmentation is controlled by the spatial domain h_s and color domain h_r .

4. STRUCTURE SIMILARITY INDEX METRIC (SSIM) FOR FUSION DECISION MAKING

When the regions are obtained by the overlaying process, the regions' structure similarity index metrics are computed to guide the wavelet coefficient fusion. Wang,(2002) firstly proposed a Universal Image Quality Index (UIQI) which achieves satisfactory result for assessing compressed image quality. Afterward, Wang,(2004) improved on it and named it Structure Similarity Index Metric (SSIM) which is a global metric to measure the similarity of two images. In this paper, we use the SSIM to calculate the similarity of corresponding regions in two input images. It is defined as follows:

$$SSIM(x,y) = [l(x,y)]^{\alpha} \cdot [c(x,y)]^{\beta} \cdot [s(x,y)]^{\gamma} \quad (3)$$

The luminance, contrast and structure comparison measures were given as follows:

$$l(x,y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \quad (4)$$

$$c(x,y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \quad (5)$$

$$s(x,y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \quad (6)$$

C_1, C_2, C_3 are constants. When $\alpha = \beta = \gamma = 1$ in formula (3), the SSIM index is given by

$$SSIM = \frac{(2\mu_A \cdot \mu_B + C_1)(2\sigma_{AB} + C_2)}{(\mu_A^2 + \mu_B^2 + C_1)(\sigma_A^2 + \sigma_B^2 + C_2)} \quad (7)$$

where μ_A, μ_B are the mean values of regions in images A and B; and σ_A, σ_B are the variances of images A and B. σ_{AB} is the covariance of A and B. C_1, C_2 are two constants.

The SSIM index may be illustrated geometrically in a vector space of image components. These image components can be either pixel intensities or other extracted features. It is more meaningful to compare the similarity of two regions than to compare region match measures.

Another parameter -- the region activity level -- is defined as follows (Piella, 2003):

$$a(r) = \frac{1}{N}\sum_{c(i,j)\in r} \mathrm{c(i,j)}^2 \quad (8)$$

Where N denotes the pixel number in region r and $c(i, j)$ denotes the corresponding wavelet coefficient at location (i, j) .

For each region, the coefficients are fused according to the SSIM. If SSIM is less than a threshold α , we will perform selection; and otherwise we will perform averaging.

$$c_F(r) = \begin{cases} c_A(r), & if\ a_A(r) \ge a_B(r), SSIM(r) \le \alpha \\ c_B(r), & if\ a_A(r) < a_B(r), SSIM(r) \le \alpha \\ \frac{c_A + c_B}{2}, & SSIM(r) > \alpha \end{cases} \quad (9)$$

For each edge, the fusion rule is as follows:

$$c_F(i,j) = \begin{cases} c_A(i,j), & only\ if\ c_A(i,j)\ is\ at\ an\ edge \\ c_B(i,j), & only\ if\ c_B(i,j)\ is\ at\ an\ edge \\ \frac{c_A(i,j) + c_B(i,j)}{2} & both\ are\ at\ edges \end{cases} \quad (10)$$

Once the composite coefficients are obtained, the fused image can be produced by inverse wavelet procedure.

5. EVALUATION OF PROPOSED METHODS

To evaluate the proposed method, two sets of images are used. One is the built-up area and the other is rural area. Both are from IKONOS-2 sensors. The dimensions of Panchromatic (Pan) and Multi-Spectral (MS) images are 512*512 and128*128, respectively. The proposed fusion result is compared with the conventional region-based image fusion using Canny segmentation by both visual inspection and objective analysis. Four objective measures are used in this evaluation, i.e. Entropy, Mutual Information (MI), Spatial Frequency (SF) and Relative Dimensionless Global Error in Synthesis (ERGAS) (Eskicioglu, et al,1995).

Built-up area data sets

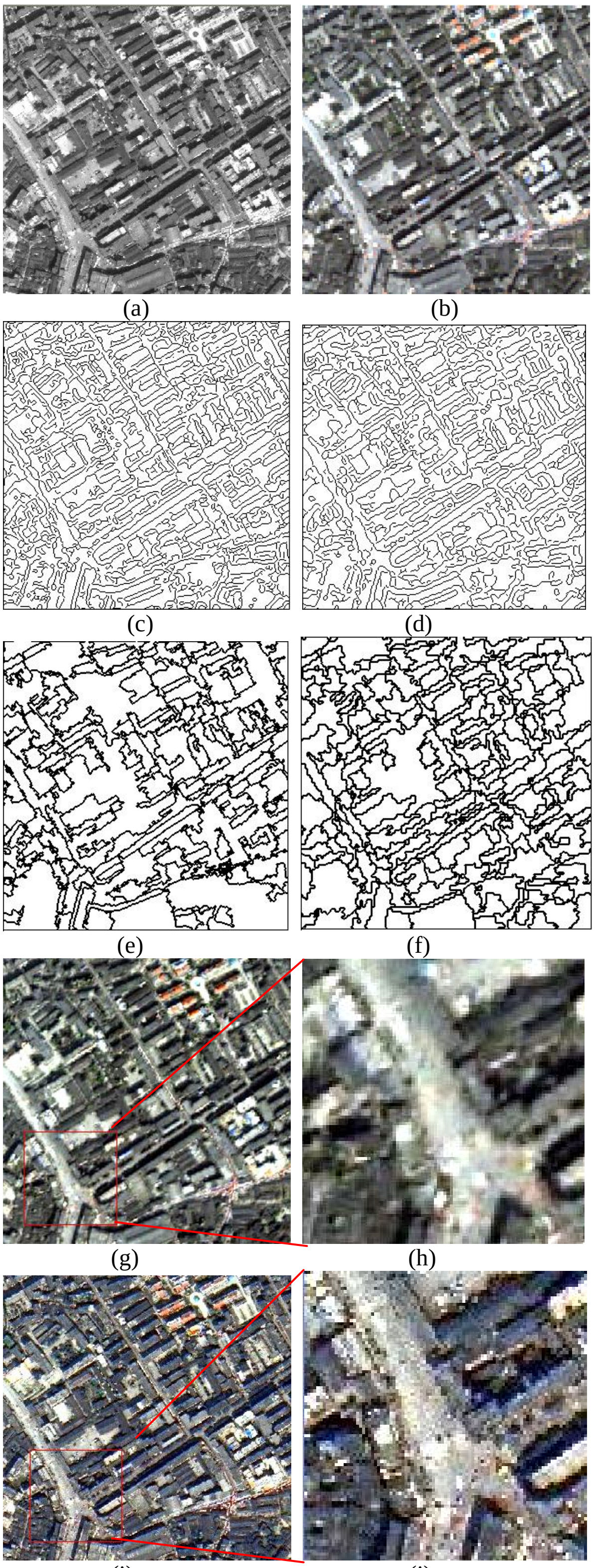

Figure 2. Fusion of built-up area images. (a)Original Pan image, (b)Original MS image, (c) Canny segmentation of Pan image, (d) Canny segmentation of MS image, (e) mean shift segmentation of Pan image, (f) mean shift segmentation of MS image, (g) Canny segmentation fused result, (i) Our proposed fused result

A visual inspection of results shown in Figure 2 reveal that the Canny segmentation of Pan and MS images produces over segmented results and the resultant regions are too tinny to tell which one is useful. In contrast, the mean shift segmentation produces sounder results. The segmentation trends of Pan and MS images are nearly the same. The total number of regions in MS image is slightly more than that in Pan image with the mean shift segmentation but the relationship is reversed with Canny segmentation . It is clear that the image fused with mean sift segmentation is clearer than that with Canny segmentation. The spatial texture of images fused with Canny segmentation is disturbed due to over segmentation.

image		Entropy	MI	SF	ERGAS
1	B	6.8396	1.2108	20.2839	12.9820
	G	7.5488	1.2141	25.9663	
	R	7.8040	1.2184	29.3771	
	Nir	8.1422	1.2126	32.6054	
2	B	7.1512	1.1520	32.2683	12.0265
	G	7.5669	1.1626	39.6594	
	R	7.6649	1.1792	42.1544	
	Nir	7.9947	1.1862	41.3579	

Table 1. Quantitative results on built-up area. (Image 1 from Canny segmentation fusion and image 2 from proposed fusion)

The objective evaluation is made to verify the visual analysis and the results are given in Table 1. From Table 1 we can see, that the Entropy in image 1 is larger than that in image 2 except for the blue band. The MI in image 1 is also a slightly larger than that in image 2. Of course, this doesn't mean the quality of image 1 is better than of image 2 as the noise will also increase the information which can be verified by visual analysis. But the SF in image 2 is much higher than in image 1. This means the clarity in former image is higher than latter images which are consistent with the visual inspection. The smaller ERGAS implies error in all the bands is smaller.

Rural area data sets

(a) (b)

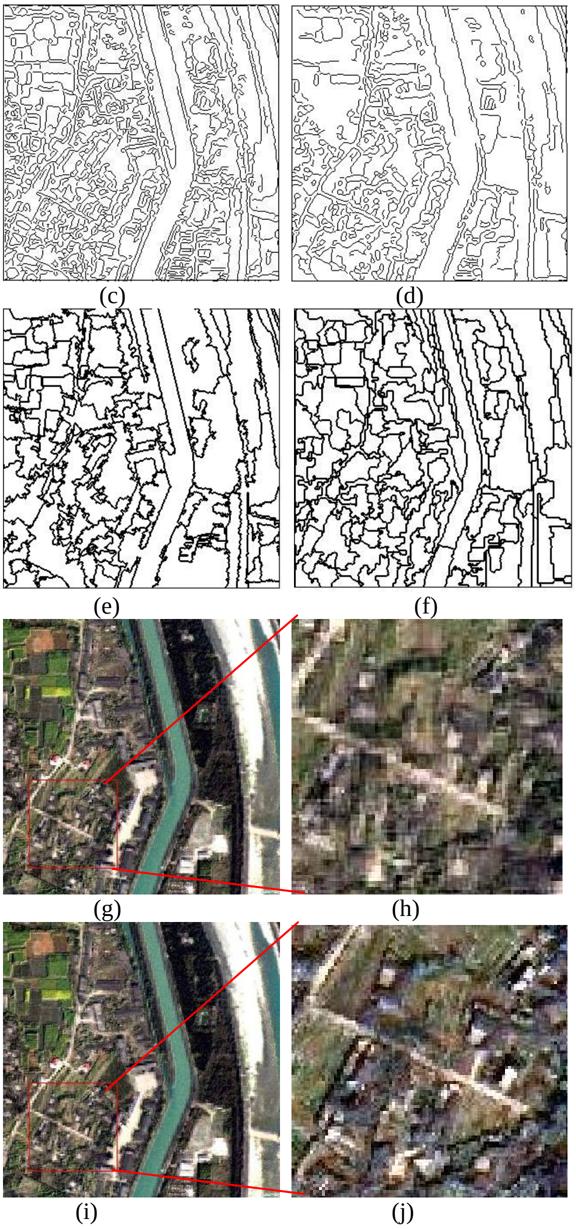

Figure 3. Fusion of rural area images. (a)Original Pan image, (b)Original MS image, (c) Canny segmentation of Pan image, (d) Canny segmentation of MS image, (e) mean shift segmentation of Pan image, (f) mean shift segmentation of MS image, (g) Canny segmentation fused result, (i) Our proposed fused result

As figure 3 shows, the test on rural area shows a similar result to the built-up area. The image in Figure 3(c) is over segmented. In Figure 3(d), some parts are over segmented and some are under segmented. The segmentation results shown in Figure 3(e) and (f) are more reasonable than those in Figure 3(d) and (e). The fused result based on Canny segmentation is blurred as can be seen from Figure 3(g) and (h). The sharpness and spectral reservation of the images fused by this new method are better. More detailed quantitative evaluation result is given below:

image		Entropy	MI	SF	ERGAS
1	B	6.6494	1.2668	8.4064	12.0409
	G	7.4298	1.2648	10.2679	
	R	7.4779	1.2725	10.9206	
	Nir	8.4651	1.2521	18.0159	
2	B	6.9650	1.1759	26.8664	11.7954
	G	7.4381	1.1892	30.5736	
	R	7.4379	1.2147	25.9148	
	Nir	8.5994	1.2240	37.7858	

Table 2. The quantitative results on rural area. (Image 1 from Canny segmentation fusion and image 2 from proposed fusion)

6. CONCLUSIONS

In this paper, a new technique for the fusion of high-resolution images has been described. In this technique, mean shift segmentation is adopted to extract the features from images. SSIM is used to measure the region similarity which has more physical meaning. The SSIM is then used to guide the decision making in the fusion process. Experimental evaluations have been conducted for built-up areas and rural areas. The results show this new technique performs better than the conventional technique with Canny detection operator. It has been found that, for high resolution image, the Canny detection tends to produce unstable segmentation, i.e. over segmentation in a sub-region and under segment in another region of the same images. On the other hand, mean shift segmentation is more reliable. The sharpness and spectral reservation of images fused by our proposed technique are better than those by conventional method with Canny segmentation.

This conclusions made here are based on the limited tests. More comprehensive tests will be conducted in the future.

ACKNOWLEDGEMENT

This research was supported by State 973 project grant 2006CB701304.

REFERENCES

Comaniciu, D., Meer, P.,1999. Mean shift analysis and applications. The proceedings of the seventh IEEE international conference on Computer Vision, pp. 1197-1203.

Eskicioglu, A.M., Fisher, P.S.,1995. Image quantity measures and their performance. IEEE Transactions on Communications, 43 (12), pp.2959-2965.

Lewis, J.J., O'Callaghan, R.J., Nikolov, S.G., Bull, D.R., Canagarajah, N.,2007. Pixel- and region-based image fusion using complex wavelets. Information Fusion (8), pp. 119-130.

Luo, J.B., Guo, C.E.,2003. Perceptual grouping of segmented regions in color images. Pattern Recognition, 36(12), pp. 2781-2792.

Mo, D.K., Lin,H., Li, J.P., Sun, H., Xiong, Y.J.,2006. VHR Imagery Multi-Resolution Segmentation Based on Mean Shift. Journal of Guangxi Normal University, 24(4), pp. 247-250.

Piella, G.,2003. A general framework for multiresolution image fusion: From pixels to regions. Information Fusion, 4(4), pp. 259-280.

Pohl, C., Genderen, J. L. V.,1998. Multisensor image fusion in remote sensing: concepts, methods and applications. International Journal of Remote Sensing, 19(5), pp. 823-854.

Wald, L., Ranchin, T., Mangolini, M.,1997. Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images. Photogrammetric Engineering & Remote Sensing, 63(6), pp. 691-699.

Wang, R., Gao, L. Q., Yang, S., Chai, Y.H., Liu,Y.C.,2005. An Image Fusion Approach Based On Segmentation Region. International Journal of Information Technology, 11(7), pp. 92-100.

Wang, Z., Bovik, A.C.,2002. A universal image quality index. IEEE Signal Processing Letters, 9(3), pp. 81-84.

Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.,2004. Image quality assessment: From error visibility to structural similarity. IEEE Transaction on Image Processing, 13(4), pp. 600-612.

Zhang, Z., Blum, R.S.,1997. A Region-based image fusion scheme for concealed weapon detection. Proceedings of 31st Annual Conference on Information Sciences and Systems, March, pp.168-173.

Zhen, J., He, G.J.,2005. Shortages of the data fusion approaches to high-resolution satellite images and expected improvements. Remote sensing information, April, pp.45-48.