An Image Patch Matching Method Based on Multi-feature Fusion

Xiangru Yu[1,2], Yukun Guo[1,2], Jinping Li[1,2*], Fudong Cai[3]
1. School of Information Science and Engineering University of Jinan, 250022, Jinan, China
2. Shandong Provincial Key Laboratory of Network Based Intelligent Computing University of Jinan, 250022, Jinan, Shandong, China
3. Shandong Senter Electronic Co., Ltd, 255000, Zibo, Shandong, China
*Corresponding Author: ise_lijp@ujn.edu.cn

Abstract—**Appropriate features are very important for the robustness and effectiveness of matching algorithms. In general, current algorithms depend on descriptors like SIFT, SURF, which makes them only care about information of key points while ignoring the knowledge of the whole image, thus those methods are easy to result in false matches. We propose a novel matching method called multi-feature fusion, which takes full advantage of geometric, gray, color and texture features. Then we validate the effect of our method using images captured from practical applications. Experiments show the method can effectively complete the matching task of image patch.**

Keywords: image patch; matching; multi-feature fusion; texture energy

I. INTRODUCTION

For the problem of detecting changed areas caused by man-made factors, image matching is an effective and appropriate mean. In this paper, all the images are taken from the position-fixed and focal length fixed cameras, and these images are taken in different time for the same scenes. From the Fig. 1, we can see that these factors which result in change can be divided into two types: natural factors such as illumination in Fig. 1 (a), vegetation growth in Fig. 1 (b) and so on, man-made factors like ground construction in Fig. 1 (c), vehicles and cargo accumulation in Fig. 1 (d).

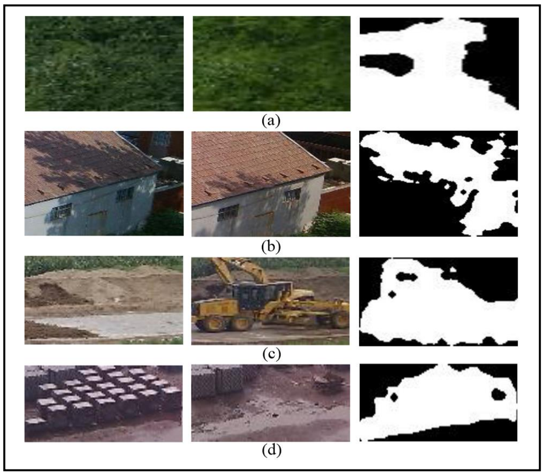


Figure 1. Image patch exists changes caused by natural and man-made factors.

The third column in Fig. 1 is calculated by multiple color spaces difference fusion [1,2]. For people, the most concerned regions are the changed regions caused by man-made factors, because hidden dangers usually exist in these regions. From Fig. 1, we can find features in geometric, gray, color and texture are different between changed regions caused by man-made factors and changed regions caused by natural factors, thus these features are important for image patch matching.

Image matching is a fundamental task in image processing and computer vision. Up to now, algorithms proposed in image matching can be divided into two types, one is based on feature detector and descriptors, the other is based on traditional features like contour, characteristics of frequency domain, etc.

SIFT (Scale-invariant Feature Transform) is a common image matching algorithm. It is proposed by D. G. Lowe in 1999 and 2004, which performs the following 4 basic steps [3, 4]: firstly, searches potential points over all scales; secondly, selects key points according to their stability; thirdly, assigns the direction of key points; finally, determines the key points descriptor. Although SIFT has made success in image matching, the computational burden and false match severely limit its application. To speed up the computation, Kusamura and Sheikh proposed hardware acceleration method [5,6].

To improve the matching effect, H. Bay proposed SURF, which uses integral image and box filter [7]. E. Rublee put forward ORB (Oriented FAST and Rotated BRIEF) [8], which uses FAST (Features from Accelerated Segment Test) to detect key points [9]; then BRIEF (Brief Binary Robust Independent Elementary Features) is used to describe the properties of key points [10]; the last step is matching of key points.

Besides algorithms stated above, some improved matching methods has been proposed, such as PCA-SIFT (principle component analysis) [11]. GMS (Grid-based motion statistics) incorporates the smoothness constraints into feature matching [12], GMM (Gaussian mixture model) is incorporated into image matching [13]. For the image patch matching, the number of key points in image patch is not enough to accomplish the robust matching.

These methods based on features detectors and descriptors achieved good performance in some applications. but false match is an inevitable problem because these methods focuses

Fund Projects: The Major Project of Science and Technology in Shandong Province (Emerging Industry) (2015ZDXX0801A03), The Science and Technology Program Funded Project for Higher Education in Shandong Province (J14LNXX). Scientific Research Fund of University of Jinan (XKY1412).

978-1-5386-1937-7/17/$31.00 ©2017 IEEE

on the information of key points while ignoring the knowledge of whole image.

Matching methods based on traditional features are shown as below. Hui Li proposed a matching algorithm using image contour [14]. This method is fit for images that have clear contour information. D. I. Barnea proposed a similarity measurement method using template matching [15], the matching threshold is difficult to define. P. E. Anuta put forward an image registration technique to calculate cross correlation similarity [16], which is fit for multispectral and multitemporal image. Multi-region matching depends the context complexity of image [17]. Spatial structure constraint and shape properties are used in remote sensing image matching respectively in [18] [19].

Although many methods have been proposed, those methods have respective limitations, which makes them fit for some particular scenes but not fit for the image patch matching in this paper.

To improve the effect of image matching, we propose a multi-feature fusion method. After observing a large number of samples, we find the features which exist obvious difference between changed areas caused by man-made and natural factors include geometric, color, gray and texture features. But these features are available for the changed regions, not for the whole image. Changed regions are called ROI (region of interest) in the next context. Before matching, the image patches to match are obtained according to the ROI. Multi-feature fusion method for image patch matching performs the following 5 basic steps: firstly, we use area, eccentricity and solidity of ROI to describe the properties of geometric features; secondly, we calculate gray features like gray mean, gray variance and the correlation of gray histograms; thirdly, the correlation coefficients of different color components are used to represent the features of color; fourthly, we get the texture features by multi-scale laws energy [20,21]; finally, we fuse those features stated above and get the image patch matching result.

II. IMAGE PATCH BASED ON MULTI-FEATURE FUSION ANALYSIS

A. Multiple Color Spaces Difference Fusion

For the images captured from camera whose focal length and position are fixed, image difference is used to get the mask of ROI. Because the different color channel in different color spaces has different degrees response to same changed areas, we determine to use the multiple color spaces difference fusion to get the ROI, and the weight of different channel image is equal when the difference image of different channel is fused. Then the fusion difference image is binarized by OTSU (maximal variance between clusters) to get the ROI mask. After these, obtain the bounding rectangle of ROI, and the image patch corresponding with the bounding rectangle of ROI is obtained from the images to match.

B. Geometric Features

The geometrical features extracted are used to describe the geometric parameters of ROI. The area of the ROI is the total number of white pixels in the mask of ROI. Because small connected area is usually caused by noise, so it is necessary to set an area thresh to filter such cases. False objects include building and high voltage line, edges of those false objects look like a flat strip. The eccentricity of the ROI is the eccentricity of the ellipse with the same standard second order central moment, which can describe the degree of flat of ROI. The solidity of the ROI represents the pixel ratio falling within the ROI region to minimum convex hull. Finally, the extracted geometric features are Geometry = {Area, Eccentricity, Solidity}. Fig. 2 lists the three values of geometric features in ROI, the difference of geometric features is obvious.

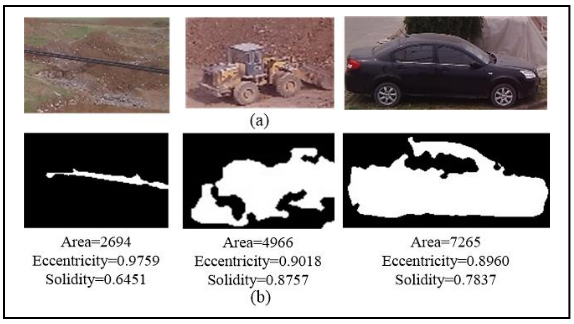


Figure 2. The geometric features of ROI

Fig. 2 (a) shows the image patch corresponding with the ROI mask. Fig. 2 (b) shows the parameters of geometric features for the ROI.

C. Gray Features

Grayscale features include gray mean, gray variance, and grayscale histogram. In general, the difference between the gray mean values of the two images with large difference is larger, so the gray scale mean can be used as a feature of image matching. The gray mean can be calculated by (1).

$$\mu = \frac{\sum_{i=0}^{m}\sum_{j=0}^{n}(I(i,j) \bullet S(M(i,j)))}{\sum_{i=0}^{m}\sum_{j=0}^{n}S(M(i,j))} \tag{1}$$

I is the gray image patch obtained according to the *M*, *M* is the mask image of *ROI*, *S* is the signal function, When the parameter value is 0, the function value is 0; otherwise the function value is 1.

The gray scale variance describes the degree of varying in the gray value of the pixel of the image. When calculating the gray variance, we get a more accurate variance with the ROI mask, and it is shown in (2).

$$\sigma^2 = \frac{\sum_{i=0}^{m}\sum_{j=0}^{n}((I(i,j)-\mu)^2 \bullet S(M(i,j)))}{\sum_{i=0}^{m}\sum_{j=0}^{n}S(M(i,j))} \tag{2}$$

In this paper, we calculate the mean and variance of the gray scale in the matching image and the reference image respectively, and then the difference between the mean gray scale and the difference between the variance gray scale can be calculated. These differences are taken as the final matching feature.

The gray statistic histogram describes the distribution of image grayscale. By comparing the histogram of two images, the matching degree of two images can be calculated. As shown in Fig. 3, the histogram of the gray statistic of the different images are quite different.

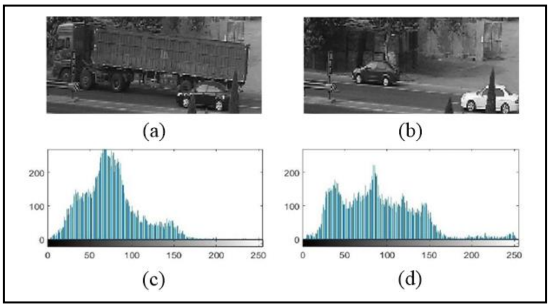


Figure 3. Histograms of gray scale.

Fig. 3 (a) and (b) are the images patches to match, and Fig. 3(c) and (d) are the histograms of (a) and (b) respectively.

D. Color Features

The image patches are mapped into multiple color spaces and the features are extracted in different color channel. The multi-color space includes RGB, HSV, YUV and Lab. The color mean can be calculated using (1). The color histogram is analogy to the grayscale histogram, except that the color histogram reflects one color channel information, and the grayscale histogram reflects the gray scale information of the image. The color histogram of the multiple color space used in this paper is shown in Fig. 4.

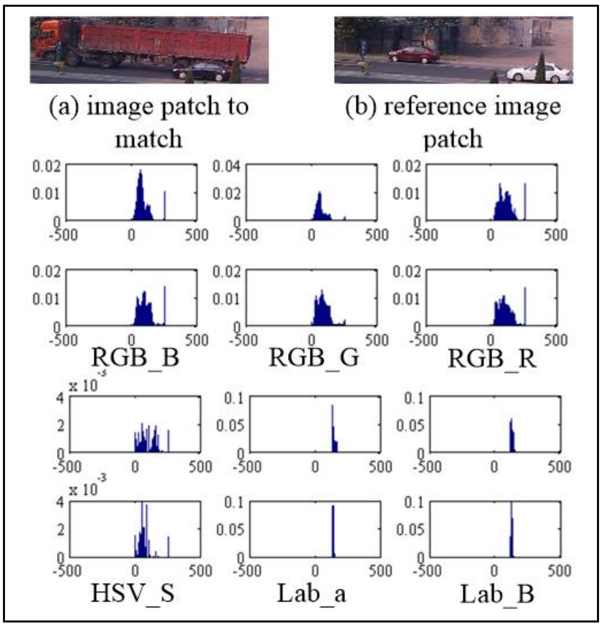


Figure 4. Different color channel histograms of multiple color spaces.

From Fig. 4, we can see that the distributions of pixels in the color histogram of the two image patches with large difference in content show the difference obviously. And the similarity of histograms can be expressed by the correlation of histograms, it is shown in (3) (4).

$$d_{corr}(H_1, H_2) = \frac{\sum_i (H_1^{'}(i) \bullet H_2^{'}(i))}{\sqrt{\sum_i (H_1^{'2}(i) \bullet H_2^{'2}(i))}} \tag{3}$$

$$H_k^{'}(i) = H_k(\mathrm{i}) \quad \frac{1}{N}(\sum_j H_k(j)) \tag{4}$$

where N is the number of bin in histogram. And the color features include the histogram correlation of *RGB_R*, *RGB_G*, *RGB_B*, *HSV_S*, *YUV_U*, *YUV_V*, *Lab_a* and *Lab_b*.

E. Texture Features

For image patch matching, if we only use the statistical features of gray and color information of an image patch, obviously it is insufficient. Although the statistical features can reflect the gray and color information of image pixels in the statistical aspect, but lose the pixel information in the spatial position of the relationship. And the texture feature can describe the rugged groove on the surface of object, the image patch appears as a pixel gray value of the bright and dark changes. Considering the ability of texture in describing image pixel spatial position and intensity information, we use grayscale correlation coefficients and multi-scale laws energy to describe texture.

The gray correlation coefficient is calculated using (5). In this paper, the window is used to calculate the correlation coefficient with the pixel as the central point.

$$C(x,y) = \frac{\sum_{i=\ k}^{k}\sum_{j=\ k}^{k}(I_1^{'}(x+i,y+j) \bullet I_2^{'}(x+i,y+j))}{\sqrt{\sum_{i=\ k}^{k}\sum_{j=\ k}^{k}(I_1^{'2}(x+i,y+j) \bullet I_2^{'2}(x+i,y+j))}} \tag{5}$$

where C is the correlation map, $I_k^{'}$ is calculated by (6), I_k means gray image.

$$I_k^{'}(x,y) = I_k(x,y) - \frac{1}{2k+1}(\sum_{i=-k}^{k}\sum_{j=-k}^{k} I_k(x+i,y+j)) \tag{6}$$

The ROI mask is used to guide the computation of correlation coefficient in image patch, as shown in (7).

$$D(x,y) = C(x,y) \bullet S(M(x,y)) \tag{7}$$

For ease of observation, the correlation coefficient has been normalized to 0 to 255. As shown in Fig. 5.

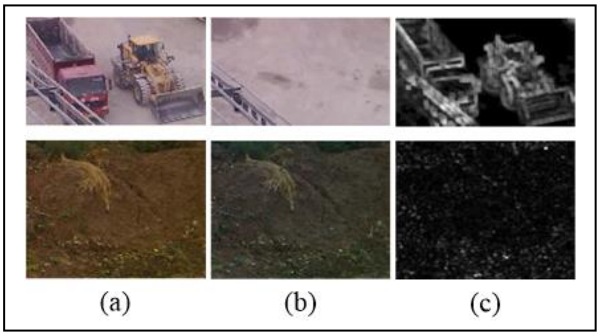


Figure 5. Image patches and its correlation map.

Fig. 5 (a) represents the image patches obtained in the matching image, Fig. 5 (b) represents the image patches obtained in the reference image, Fig. 5 (c) is the correlation map for (a) and (b).

As can be seen from Fig. 5, for image patches with similar content, the correlation coefficient image will appear similar. When the difference in image content is large, the correlation coefficient image is different obviously, especially in the edge part. The mean value of the correlation coefficient image is used as the final feature in image patch matching.

Because the same texture has different characteristics on different scales, the texture features at multiple scales will be better than the texture features at a single scale. Multi-scale laws energy is a better way to extract the texture features from image patches.

In this paper, we calculate the energy distribution map in four scales as shown in (8), $k \in \{1,2,3,4\}$, k means the size of scale, $2k+1$ represents the size of window which is used to extract texture.

$$E(x,y) = \frac{1}{(2k+1)^2} \sum_{i=-k}^{k} \sum_{j=-k}^{k} (I(x+i, y+j))^2 \tag{8}$$

where I is the original image, k is determined by the current image scale.

Mean and variance are shown in (9) and (10).

$$\sigma^2 = \frac{1}{(2k+1)^2} \sum_{i=-k}^{k} \sum_{j=-k}^{k} (I(x+i, y+j) - \mu(x,y))^2 \tag{9}$$

$$\mu^2 = \frac{1}{(2k+1)^2} \sum_{i=-k}^{k} \sum_{j=-k}^{k} (I(x+i, y+j))^2 \tag{10}$$

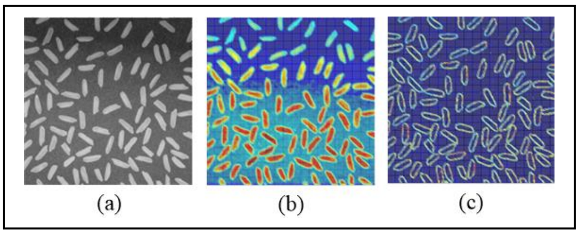


Figure 6. The distribution map of texture energy, k=2. (a) raw image. (b) energy map of (a). (c) distribution map of energy variance.

From Fig. 6 (b), we can see that energy calculated by (8) is very high. But the point we concern is the intensity of gray fluctuation reflected by the energy map. As shown in Fig. 6 (c), Eq. 9 also has a good corresponding to the edge of the rice grain while the energy response to the rice grain is very low, because the fluctuation of the pixel value within the rice grain is not large. Therefore, (9) is more accurate than (8) in reflecting the distribution of texture energy.

The three features (correlation, means and variance of energy) in texture distribution on four different scales are used to accomplish the image patch matching. Results as shown in Fig. 7.

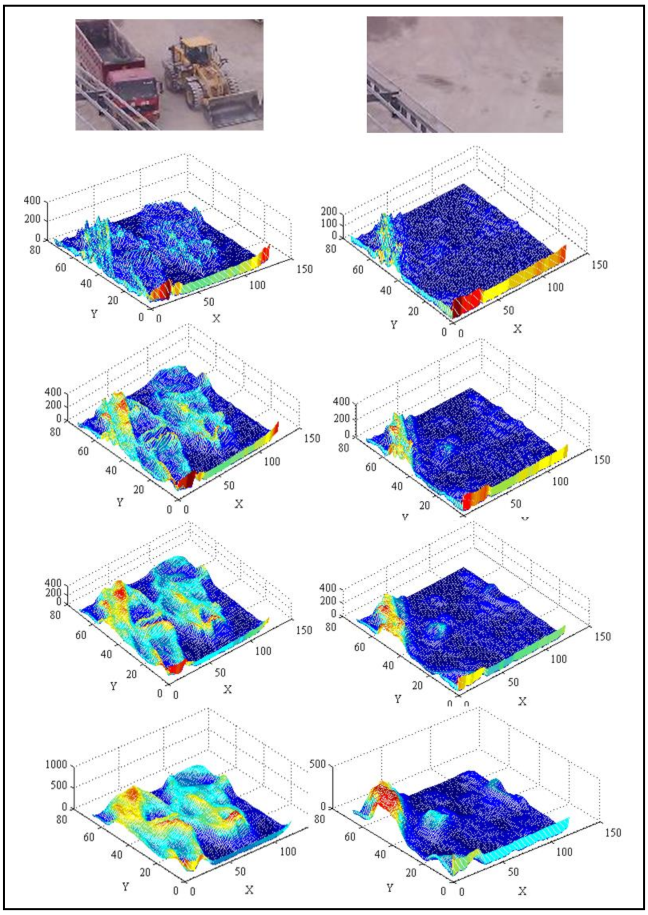


Figure 7. Energy distribution map of multi-scale

In Fig. 7, the first row shows the image patches to match. The second row to the fifth row show the texture energy distribution at the scale of 1, 2, 3, and 4 of two image patches respectively.

F. Multi-feature fusion

The features used in image patch matching represent the different properties of image patch, and the features are independent of each other. The decision fusion for four types features is used to fusion these features. And the weight of feature is equal. The fusion process is shown in Fig. 8.

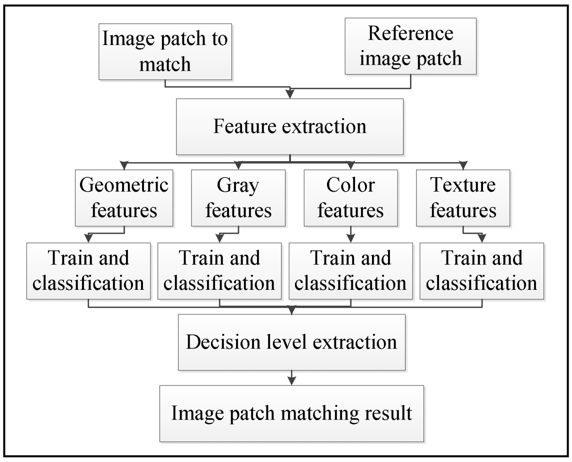


Figure 8. The fusion of multi-feature

III. EXPERIMENTS AND DISCUSSION

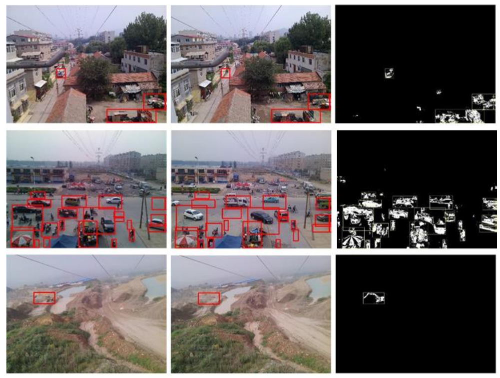
All the experiments are carried out using an Intel Core i5 1.60GHz with a NVIDLA GEFORCE GT 750M, 12GB memory. To validate the effectiveness of out method for the image patch matching, we compare our method with other matching methods using only one kind of feature. We use 23278 images from 5 different scenes that contain different kinds of objects. The average time for one image is about 200ms, which includes the image difference and image patch matching.

TABLE I. CLASSIFICATION ACCURACY OF DIFFERENT FEATURES

Method ID	**Accuracy**
geometric features	64.35%
gray features	60%
color features	76.09%
texture features	71.74%
multi-feature fusion	87.54%

It can be seen from Table 1 that classification result using grayscale feature is the worst, only slightly better than the random classification. The classification result using geometric features is also not ideal. The classification accuracy using color features and texture features is higher than 70%, which is higher than the geometric and gray features. The correct rate of classification result using four categories of features reaches 87.54%. After the features are fused, the classification accuracy has been improved significantly. Fig. 9 shows the matching results in experiments.

Fig. 9 (c) shows the ROI mask. When the image patch corresponding with ROI mask cannot find its matching image patch, it will be marked with red rectangle. In the first row of Fig. 9 (a), the changed region caused by plant shaking finds its matching image patch successfully, it demonstrates that the algorithm in this paper is robust for the change caused by natural factors. In the second row of Fig. 9, the change caused by man-made factors like vehicles and pedestrian has been detected. In the third row of Fig. 9, the only difference is the excavator in the distance, although the excavator is similar to the background, and the area proportion of excavator is small compared with the image, the algorithm is still able to detect it.



(a) matching image (b) reference image (c) ROI image

Figure 9. Result of image patches matching.

IV. CONCLUSIONS

We present a novel matching method for image patch, the experimental results show that the proposed method can effectively match image patch and decrease the rate of wrong match caused by natural factors such as shadow, vegetation growth, etc. But there are still some problems in our method, for example, occlusion of objects will affect the result of image patch matching. We will address these issues in future research.

REFERENCES

[1] Z. Zhang, J. Li, P. B, “Mobile object detection algorithm based on information fusion of multiple color spaces”. Journal of University of Jinan(Social Science Edition), vol. 25, no. 2, pp. 191-194, 2011.

[2] M. Qin, Y. Guo, J. Li, “Ojectionable video detecting based on multiple color space”, Journal of Jilin University (Engineering and Technology Edition), vol. 43, no. S1, pp. 283-286, 2011.

[3] D. G. Lowe, “Object recognition from local scale-invariant features”, Proceedings of the Seventh IEEE International Conference on Computer Vision, Kerkyra, vol. 2, pp. 1150-1157, 1999.

[4] D. G. Lowe, “Distinctive image features from scale-invariant key points”, International Journal of Computer Vision, vol. 60, no. 2, pp. 91-110, 2004.

[5] Y. Kusamura, Y. Kozawa, T. Amagasa and H. Kitagawa, "GPU Acceleration of Content-Based Image Retrieval Based on SIFT Descriptors," 2016 19th International Conference on Network-Based Information Systems (NBiS), Ostrava, 2016, pp. 342-347..

[6] I. Sheikh, A. Oviedo, A. Emerio and N. Mekhiel, "Parallel implementation of image matching with MPI," 2016 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), Vancouver, BC, 2016, pp. 1-4.

[7] H. Bay, Ess A, Tuytelaars T, et al, “Speeded-up robust features (SURF)”, Computer Vision and Image Understanding, vol.110, no.3, pp. 346-359, 2008.

[8] E. Rublee, V. Rabaud, K. Konolige and G. Bradski, “ORB: An efficient alternative to SIFT or SURF”, 2011 International Conference on Computer Vision, Barcelona, 2011, pp. 2564-2571.

[9] E. Rosten, and T. Drummond. “Machine Learning for High-Speed Corner Detection”, European Conference on Computer Vision Springer-Verlag, 2006, pp. 430-443.

[10] M. Calonder, V. Lepetit, C. Strecha, et al, “BRIEF: Binary Robust Independent Elementary Features”, European Conference on Computer Vision. Springer-Verlag, 2010, pp. 778-792.

[11] Yan Ke and R. Sukthankar, “PCA-SIFT: a more distinctive representation for local image descriptors”, Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 506-513, 2004.

[12] J. W. Bian, W. Lin, Y. Matsushita, S. K. Yeung, T. D. Nguyen, M. M. Cheng, “GMS: Grid-based Motion Statistics for Fast, Ultra-robust Feature Correspondence”, Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[13] K. Sun, L. Liu and W. Tao, “Image Matching via Feature Fusion and Coherent Constraint”, in IEEE Geoscience and Remote Sensing Letters, vol. 14, no. 3, pp. 289-293, Mar. 2017.

[14] Hui Li, B. S. Manjunath and S. K. Mitra, “A contour-based approach to multisensor image registration” in IEEE Transactions on Image Processing, vol. 4, no. 3, pp. 320-334, Mar. 1995.

[15] D. I. Barnea and H. F. Silverman, “A Class of Algorithms for Fast Digital Image Registration” in IEEE Transactions on Computers, vol. C-21, no. 2, pp. 179-186, Feb. 1972..

[16] P. E. Anuta, "Spatial Registration of Multispectral and Multitemporal Digital Imagery Using Fast Fourier Transform Techniques," in IEEE Transactions on Geoscience Electronics, vol. 8, no. 4, pp. 353-368, Oct. 1970.

[17] C. M. Pun, C. Yan and X. C. Yuan, “Image Alignment-Based Multi-Region Matching for Object-Level Tampering Detection”, in IEEE Transactions on Information Forensics and Security, vol. 12, no. 2, pp. 377-391, Feb. 2017.

[18] J. Jiang and X. Shi, “A Robust Point-Matching Algorithm Based on Integrated Spatial Structure Constraint for Remote Sensing Image Registration”, in IEEE Geoscience and Remote Sensing Letters, vol. 13, no. 11, pp. 1716-1720, Nov. 2016.

[19] Y. Ye, L. Shen, M. Hao, J. Wang and Z. Xu, “Robust Optical-to-SAR Image Matching Based on Shape Properties”, in IEEE Geoscience and Remote Sensing Letters, vol. 14, no. 4, pp. 564-568, Apr. 2017.

[20] Laws, K. I. “Rapid texture identification”, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. vol. 238, pp. 376-380, 1980.

[21] Y. Guo, X. Yu and J. Li, “A classification method of epithelial cells and clue cells based on multi-scale texture analysis”, 2016 9th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), Datong, 2016, pp. 432-436..