2011 3rd International Conference on Signal Processing Systems (ICSPS 2011)
**IPCSIT vol. 48 (2012) © (2012) IACSIT Press, Singapore**
DOI: 10.7763/IPCSIT.2012.V48.22

# Multi-modal Image Matching Method Based on Optimized Wavelet Transform Modular Maximum and DAISY Descriptor

Zhenbing Zhao, Sasa Wang, Ping Yu and Qin Wang

School of Electrical and Electronic Engineering, North China Electric Power University, Baoding, China
Email: diligencyzhao@yahoo.com.cn; Wangsasa1988@hotmail.com

**Abstract.** The existing matching methods of multi-modal image have poor performance, so this paper proposes a kind of matching method of multi-modal image based on optimized wavelet transform modular maximum and DAISY descriptor. The method is divided into three steps: firstly, the optimized wavelet transform modular maximum is used to extract feature points; secondly, a fast local DAISY descriptor is introduced to describe feature points; finally, the Euclidean Distance(ED) and the RANdom SAmple Consensus (RANSAC) are used in coarse and fine matching for the descriptor. Experiment results show that the method has a higher matching precision than the traditional method.

**Keywords***:* Multi-modal image; Feature points extraction; Wavelet transform modular maximum; DAISY; Euclidean distance; RANSAC

## 1. Introduction

Searching for corresponding points between two or more images is called matching, which has turned out to be one of the hottest and most difficult issues in respect of image processing in recent decades. Currently, matching has been widely developed and applied in many fields, such as in the military, remote sensing, medicine, computer vision and so on [1].

As multi-modal images come from different imaging devices, they are able to provide richer and more comprehensive information than the single-modal images. For example, compared with images obtained by a single sensor, the fused images obtained by sensors of different types can maximally use the information of various information sources, reduce the uncertainty of the target perception of the scene, and improve the accuracy of target recognition. Consequently, the study about the multi-modal image matching method has become a popular trend.

Three kinds of image matching methods are as follows: based on gray correlation, based on feature and based on interpretation. Now the most studies are the image matching method based on feature. Due to the wavelet transformation having the properties of multi-resolution and localization, it is widely used for image feature point extraction. Literature [2] presents a kind of image feature matching method based on wavelet modulus maxima, and the method for single-modal image matching obtains good results. But the method for multi-modal image matching has poor results.

Aiming at solving the problem mentioned above, firstly, the optimized wavelet transform modular maximum method is proposed to extract feature points of multi-modal image in this paper, and then a fast local descriptor DAISY [3] is introduced to describe the feature points. Finally, the Euclidean Distance(ED) and the RANdom SAmple Consensus (RANSAC) are used to match. The experimental results demonstrate the effectiveness and precision of the method.

## 2. Proposed Matching Method

Two multi-modal images are given to match. The goal of multi-modal image matching is to rectify one image into the coordinate system of another image and to make corresponding coordinate points in the two images fit the same geographical location. In this paper, the matching process is carried out in the following three steps:

## 2.1. Feature Point Extraction

The traditional method mentioned in literature [2] should firstly use the wavelet transform modular maximum to obtain the image edge contour matrix; then the value of each pixel is replaced by the sum of each pixel in its 8-th neighborhood; finally, the feature points are extracted through finding the local maximum. But, we should not need to compute the sum of each pixel in its 8-th neighborhood. In fact, the image feature points are obtained directly through our optimized wavelet transform modular maximum method. So compared to the traditional method, our method has a small computing. At the same time, the threshold is set to be a simple value for the traditional method, but in our method the threshold is selected depending on the standard deviation of the gradient image. Moreover, we build the sliding window to eliminate insignificant feature points [4]. Our method is described as follows:

a) Given a scale, we compute the derivative of two-dimensional Gaussian function along the x and y directions, which are defined as the wavelet functions.
b) Image is convolved with two wavelet functions, and then we compute the gradient value G and angle of each pixel.
c) The angles obtained are divided into the four types: the first direction $0^o$ or $180^o$(level)，the second direction $90^o$ or $270^o$(vertical), the third direction $45^o$ or $225^o$(positive diagonal), and the fourth direction $135^o$ or $315^o$(negative diagonal).
d) Detect each pixel whether it is the maximum in the direction of the nearest angle. If so, record the gradient, and if not, set it to zero.
e) Set the threshold on the gradient image. A point is recorded if gradient valu $G > th$, where $th = c \times \delta^2$, and c is a parameter whose value is defined by the user, and $\delta$ is the standard deviation of the gradient image. Otherwise, the gradient value is set to zero.
f) In order to eliminate insignificant feature points, we find local maxima within a sliding window of size n×n(In this paper we choose a window of size $5\times5$) [5].

## 2.2. Feature Point Description by DAISY

The schematic structure of DAISY descriptor is showed in Fig. 1. DAISY uses circular grids instead of rectangular ones, and it shows that the former has better localization property.

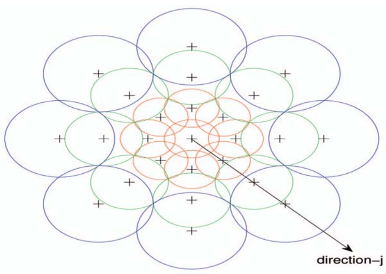


Fig.1. Structure of DAISY descriptor

For a given input image, we first compute H number of orientation maps, $G_o$, $1 \le o \le H$, one for each quantized direction, where $G_o(u,v)$ equals the image gradient norm at location $(u,v)$ for direction $o$ if it is bigger than zero, else it is equal to zero. This preserves the polarity of the intensity changes. Formally, orientation maps are written as $G_o = \left(\frac{\partial I}{\partial o}\right)^+$, where I is the input image, $o$ is the orientation of the derivative, and $(\cdot)^+$ is the operator such that $(a)^+ = \max(a,0)$.

Each orientation map is then convolved several times with Gaussian kernels of different $\Sigma$ values to obtain convolved orientation maps for different sized regions as $G_o^{\Sigma} = G_{\Sigma} * \left(\frac{\partial I}{\partial O}\right)^+$ with $G_{\Sigma}$ a Gaussian kernel. Different $\Sigma$ are used to control the size of the region.

Finally, a whole DAISY described by concatenating the previously computed normalized vectors of itself and its neighbor sample points of outer rings is as follows,

$$
\begin{aligned}
D(u_0, v_0) = [&\tilde{h}^T_{\Sigma 1}(u_0, v_0),\\
&\tilde{h}^T_{\Sigma 1}(l_1(u_0, v_0, R_1)), \ldots, \tilde{h}^T_{\Sigma 1}(l_T(u_0, v_0, R_1)),\\
&\tilde{h}^T_{\Sigma 2}(l_1(u_0, v_0, R_2)), \ldots, \tilde{h}^T_{\Sigma 2}(l_T(u_0, v_0, R_2)),\\
&\ldots\\
&\tilde{h}^T_{\Sigma Q}(l_1(u_0, v_0, R_Q)), \ldots, \tilde{h}^T_{\Sigma Q}(l_T(u_0, v_0, R_Q))]^T
\end{aligned}
$$

Where $l_j(u, v, R)$ is the location with distance $R$ from $(u, v)$ in $j$ direction. In this paper, the proposed parameters are as follows: $R = 15, Q = 3, T = 8, H = 8$. Thus the total number of the sample points is $S = Q * T + 1 = 25$, and the dimensionality of a DAISY descriptor in this case is $D = S * H = 200$.

### 2.3. Feature Point Matching

In the matching stage, Euclidean Distance is used as the metric to measure the distance between one descriptor and another. We apply in the matching method based on NNDR(Nearest Neighbor Distance Ratio). To further eliminate the error matching points and enhance matching accuracy, we use RANdom SAmple Consensus (RANSAC) [6].

## 3. Experiment and Results

The experiment is conducted using a sliding window of size 5×5 when extracting feature points through finding local maxima. In order to demonstrate the capabilities of the proposed feature points based on wavelet transform modular maximum for multi-modal image matching, the performance of the proposed matching method is evaluated with the accuracy ratio which is defined as :

The accuracy ratio =number of total matching interest points / number of correct matching interest points in common area.

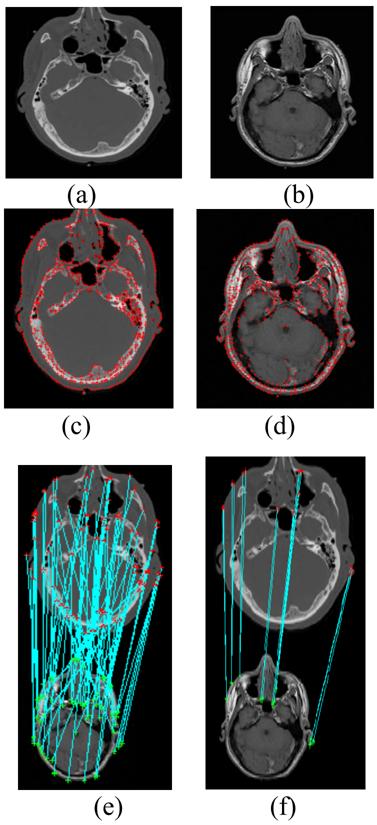


Fig.2. Experiment results

(a) The CT image (b) The MR image
(c) The result of extracting feature points on the CT image
(d) The result of extracting feature points on the MR image
(e) The coarse matching result (f) The fine matching result

In this experiment, we choose different multi-modal images, which include several typical infrared/visible image, medical image and remote image. In the meanwhile, we compare the proposed method with the traditional method. Our experiment results are illustrated in Fig. 2 and Table 1. From the Fig. 2, we know that the optimized method can extract good local feature points and can obtain a good matching result. From the Table 1, we can clearly see that the proposed matching method—DAISY descriptor combined with feature points extraction based on optimized wavelet transform modular maximum has a higher matching precision.

Table1. Feature Points Matching Results.
"Total" and "Correct" in this table denote the number of total matches and correct matches;
"Ratio" denotes the ratio of correct match number to total match number.

| **Method** | **Image Pair 1:Infrared/Visible** | | | **Image Pair 2:CT/MR** | | | **Image Pair 3:Remote** | | |
|---|---|---|---|---|---|---|---|---|---|
| | ***Total*** | ***Correct*** | ***Ratio*** | ***Total*** | ***Correct*** | ***Ratio*** | ***Total*** | ***Correct*** | ***Ratio*** |
| The traditional method | 18 | 12 | 66.7% | 6 | 1 | 16.7% | 34 | 21 | 61.8% |
| The proposed method | 52 | 39 | 75% | 13 | 6 | 46.2% | 70 | 53 | 75.7% |

## 4. Conclusions

A multi-modal image matching method based on wavelet transform modular maximum feature points extraction and a DAISY descriptor has been presented in this paper, and a detailed description of the method is given. Experiment results show that it is proved that the feature point extraction method by using the optimized method not only has a small computing but also is robust for multi-modal image matching. What's more, adding the DAISY descriptor improves the matching speed because its computing speed is very fast when describing the image. But there are still some problems, because the proposed method is the better for some multi-modal images which contain the entire contents of the scene. Future work will be concentrated on the multi-modal images which contain the local contents of the scene, which will be very useful in object recognition and robot applications in computer vision [7].

## 5. Acknowledgment

The research is supported by "the Fundamental Research Funds for the Central Universities" under grant number 09QG10.

## 6. References

[1] Hongmei Wang, Yanjun Li, Ke Zhang, "An image matching algorithm based on contourlet transform", Journal of Astronautics, Vol. 29, No. 5, September. 2008, pp. 1643-1647.

[2] Xiaowei Hong, Shoudong Shi, Dan Kang, "Image feature matching based on wavelet transform modular maximum", Journal of NINGBO University(NSEE), Vol. 22, No. 1, Mar. 2009, pp. 66-69.

[3] Engin Tola, Vincent Lepetit, and Pascal Fua, "DAISY: An efficient dense descriptior applied to wide-baseline stereo", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 32, No. 5, May. 2010, pp. 815-830.

[4] Yue Li, Xiaodong Yang, Yudong Wang, Xiaoming Zhu, "An improved denoising method based on wavelet modular maximum method for DS-UWB systerms", Wireless, Mobile and Sensor Nerwork, Dec. 2007.

[5] Chahira Serief, Mourad Barkat, Youcef Bentoutou, Malek Benslama, "Robust feature points extraction for image registration based on the nonsubsampled contour transform", International Journal of Electronics and Communications, Aug, 2009, pp. 148-152.

[6] Jing Zhang, Zhuangzhi Yan, Muyun Wang, Liming Wang, "Improved method of feature extraction and matching for image mosaic", Journal of Applied Sciences-Electronics and Information Engineering, Vol. 26, No. 3, May. 2008, pp. 288-294.

[7] Jinsha Yuan, Zhenbing Zhao, Qiang Gao, Yinghui Kong, "Review and prospect on infrared/visible image registration"，Laser & Infrared. Vol. 39, No.7, July. 2009, pp. 693-699.