Adaptive Feature Extraction and Image matching Based on Haar Wavelet Transform and SIFT

[1*] Mengmeng Zhang, [2] Zeming Li, [3] ChangNian Zhang, [4] Huihui Bai
[1*] *College of Information Engineering, North China University of Technology, zmm@ncut.edu.cn*
[2] *College of Information Engineering, North China University of Technology, lzm@ncut.edu.cn*
[3] *College of Information Engineering, North China University of Technology, zhcn@ncut.edu.cn*
[4] *Institute of Information Science, Beijing Jiaotong University, luckybhh@gmail.com*

Abstract

Recently, Scale Invariant Feature Transform (SIFT) algorithm is widely used in feature extraction and image matching. However, it has some defects, such as large volume of computational data and low efficiency of image matching. To address these defects, adaptive feature extraction and image matching based on Haar Wavelet Transform and SIFT (AHWT-SIFT) is proposed in this paper. In view of the characteristics of Haar wavelet, the low-frequency components of image can be decomposed adaptively by DWT, which represents the main features of the image and avoids the high-frequency of instability redundant information. Then SIFT is applied in these low-frequency components to extract the feature points. Furthermore, nearest neighbor algorithm is utilized for image matching. The experimental results have shown that the proposed scheme not only retains the general characteristics of SIFT, but the speed and accuracy of feature points matching have been greatly improved.

Keywords: *Haar wavelet; SIFT; Image matching.*

1. Introduction

Nowadays, Content Based Image Retrieval (CBIR) is a hotspot in image processing. Early CBIR methods mainly focus on describing the bottom information of images, such as color, texture, shape and so on. These methods have some achievements but they have more difficulties to describe image scaling, rotation movement, affine and other features in detail. Therefore, scale-invariant feature extraction algorithm has become a promising choice for CBIR.

In recent years, many algorithms about scale-invariant feature point extraction have been proposed, such as Harris-Laplacian algorithm and SIFT algorithm. Harris-Laplacian algorithm [1] has been improved based on Harris-corner detection, which overcomes the defect that Harris corner can’t adapt to the change of scale-space. However, this operator is sensitive to image noise. SIFT algorithm [2] is proposed with many properties, such as rotation invariant, scale invariant and affine invariant, but its calculation speed is much slower and it is also sensitive to illumination. Later, a variety of descriptors of feature points are compared in [3], which is claimed that SIFT has the best performance. Furthermore in 2011 the invariance of SIFT scale space has been proved in [4] which has resulted in many applications based on SIFT [5], [6] and [7]. In [6] the SIFT features are extracted from the fused image of fingerprint and face. And [7] applies SIFT to extract the salient feature descriptors for face recognition. Both [6] and [7] have utilized SIFT to process DWT images. However, the decomposition level of DWT is fixed in [6] and [7], which may impact the speed and accuracy of image matching.

In order to satisfy the requirements of computing speed, in 2004, [8] proposed PAC-SIFT algorithm to further reduce the computing complexity. In 2006, [9] designed a fast approximated SIFT to increase the speed eight times, while the performance of feature extraction decreased slightly. Furthermore Bay drawn on the approximate idea used in SIFT in [10] and [11], which made the calculation speed three times faster than the conventional SIFT by using a box filter approximation of second-order Gaussian partial derivatives.

Our idea is to consider the image as a signal stream. In general, the stable component of the signal contains the main features of the image, and it is located in the low-frequency part of the signal, while the high-frequency component contains the noise and unstable edge information. In view of the characteristics of DWT, the DWT coefficients can adaptively describe the different frequency information of the image. Therefore, in this paper we decompose adaptively the image using DWT firstly, and then extract the low-frequency information of image. At the same time, this approach can filter out the instability edge points which may lead to edge effect. Consequently, less number feature points with more stable feature can be achieved by SIFT.

The rest of this paper is organized as follows. In Section 2, overview of the proposed scheme is presented and the process of feature extraction and image matching is described step by step. The performance of the proposed scheme is examined against some other relevant algorithm in Section 3. We conclude the paper in Section 4.

2. Overview of the Proposed Scheme

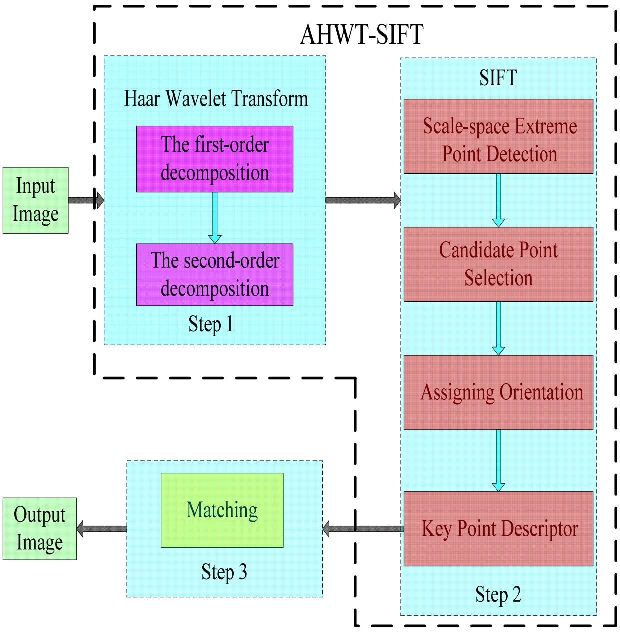

Figure 1. Steps of our Approach

First, the input image is decomposed twice by Discrete Wavelet (Haar) transform and extracted the low-frequency component; and then we extract the adaptive feature points from the low-frequency component by SIFT; in final, matching with the nearest neighbor distance. As shown in Figure 1. and the specific steps are as follows.

Step. 1 Haar Wavelet transform.

As we all know, CWT (Continuous Wavelet Transform) is the inner product of original signal and analyzing wavelet. In order to reduce the amount of computation of CWT, DWT (Discrete Wavelet Transform) is proposed. It is like a filter, which consists of a low pass filter and high pass filter. So we can get the low frequency components and high frequency components of original signal by DWT. In general, the low-frequency signal is the most important part, since the high frequency part of an "additive" effect.

As for DWT, it is the most important to find a good wavelet[12]. Mi chen [13] proposes that Haar wavelet has good character in image processing, for example: fast, easy processing, high ratio of image compression, good de-noising effect and good image features to maintain the characteristics. In this paper, we choose Haar wavelet and decompose original image by DWT[14].

$I(x, y)$ denotes the image information, $\psi(t)$ and $\phi(t)$ are wavelet function (1) and scaling function (2) approaching the 20 resolution ratio.

$$\psi(x)=\begin{cases}1 & 0\le x<1/2\\ -1 & 1/2\le x<1\\ 0 & \text{others}\end{cases} \tag{1}$$

$$\phi_i^j(x)=\phi(2^j x-i),\quad i=0,\cdots,(2^j-1) \tag{2}$$

Where j denotes scale factor, we can reduce or enlarge the graph of function by changing j; and i denotes translation parameter, while can make the graph of function translation along the axis.

In order to express the vector space, we define the vector space V^j , denoted by formula (3): it is composed of a group of basis functions $\phi_i^j(x)$.

$$V^j = sp\left\{\phi_i^j(x)\right\}\quad i=0,\cdots,2^j-1 \tag{3}$$

After DWT, the image is broken down into four sub-images containing different frequency information. Denoted by formula (4):

$$\begin{aligned}I(x,y)&=c_0^0\phi_0^0(x,y)+d_0^0\psi_0^0(x,y)\\&+d_0^1\psi_0^1(x,y)+d_1^1\psi_1^1(x,y)\end{aligned} \tag{4}$$

Among them, the four coefficients (c_0^0 、 d_0^0 、 d_0^1 and d_1^1) are obtained by Haar Wavelet transform. They are used to represent the different frequencies and different resolutions of coefficients. And the four functions ($\phi_0^0(x)$, $\psi_0^0(x)$, $\psi_0^1(x)$ and $\psi_1^1(x)$) constitute the basis of vector space V2. As shown in figure 1.

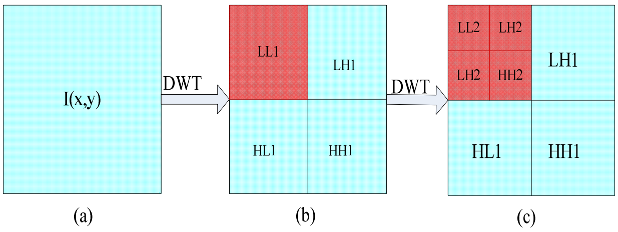

Figure 2. Second-order decomposition

From figure 2, we can see that the image generates the low-frequency components (LL1), the vertical components of the image (LH1), the horizontal components of the image (HL1) and the diagonal components (HH1) of the image after using DWT. LL1 contains large amounts of image's energy and the main features, while LH1, HL1 and HH1 consist of vertical edge information and horizontal edge information that will be generated edge effect problems in the SIFT. Because LL1 can be still decomposed, we can get second class decomposed sub-image (LL2, LH2, HL2 and HH2). Therefore, the low-frequency components of image decomposed is act as input image which can compress image data and improve computing speed of SIFT.

Step. 2 SIFT feature point detection.

After the image compression, we delete some edge information which can products edge effects, but some pseudo-edge information also generates edge effects, so we still need the step of removing edge effects in the SIFT feature point extraction. Following we describe what is SIFT in brief.

2.2.1 scale-space extreme point detection

Firstly, Lowe uses difference of Gaussian (DoG) function to generate differential Gaussian scale space. DoG is denoted by formula (5).

$$D(x,y,k\sigma)=\left[G(x,y,k\sigma)-G(x,y,\sigma)\right]*I(x,y)\ =L(x,y,k\sigma)-L(x,y,\sigma) \tag{5}$$

Where k: a constant multiplicative factor; and the compressed image $I(x,y)$; the Gaussian function $G(x,y,\sigma)$ and its standard deviation or scale-space factor σ.

Then we detect extreme point in the Gaussian scale space. As shown in figure 3, maxima and minima of the difference-of-Gaussian image are detected by comparing a pixel (marked with dark circle) to its 26 neighbors in 3×3 regions at the current and adjacent scales (marked with red circles). If the pixel marked with dark circle is maxima or minima in the 27 pixels, it is a feature candidate point.

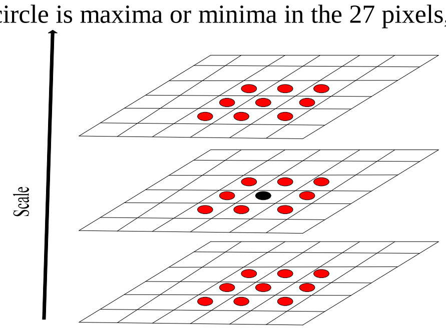

Figure 3. Extreme points detected

2.2.2 Candidate point selection, obtaining the stable key points.

1) Excluding low-contrast candidate points

$D(x,y,k\sigma)$, shifted using the second -order Taylor expansion, as follow.

$$D(x) = D + \frac{\partial D^T}{\partial x} x + \frac{1}{2} x^T \frac{\partial^2 D}{\partial x^2} x \tag{6}$$

The extremum of $D(x)$ can be got by derivation of equation (6), and than get equation (7),

$$D(\hat{x}) = D + \frac{\partial D^T}{\partial x}\hat{x} \tag{7}$$

By doing experiments Lowe gets the result that if all extremum with a value of $|D(\hat{x})|$ are more than 0.03, they should be reserved, or deleted.

2) Delete the false edge points

We can remove the pseudo-edge points by judging whether or not its principal curvature less than a threshold. The principal curvature can be computed from a 2x2 Hessian matrix, H(x, y):

$$H(x,y) = \begin{bmatrix} D_{xx}(x,y) & D_{xy}(x,y) \\ D_{xy}(x,y) & D_{yy}(x,y) \end{bmatrix} \tag{8}$$

The eigenvalues of H(x, y) are proportional to the principal curvature of D(x). now let α be the eigenvalue with the largest magnitude and βbe the smaller one, and $\alpha = \gamma\beta$.

Lowe considers that we only need to judge whether or not satisfies the formula: $\frac{tr(H)^2}{Det(H)} < \frac{(\gamma+1)^2}{\gamma}$, where γ =10, tr(H) is the trace of H(x,y) and Det(H) is its product form the determinant. If meet, the point is retained, or deleted.

2.2.3 Assigning orientation for each key point

The orientation of each key point is assigned by using the gradient histogram feature in the key point areas, so that the operator has rotation invariance

2.2.4 Key points descriptor generation

In the compressed image, firstly, rotate the axis to the direction of key points, to ensure rotation invariance.

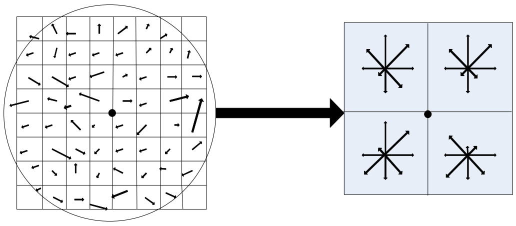

Figure 4. Key points feature vector generation.

In figure 4, the center point is the position of the current key point on the left. Each small grid represents a pixel which is a neighborhood of the key point in the scale space. The direction of the arrow represents gradient direction of the pixel and the arrow length represents the size of the gradient mode. These are weighted by a Gaussian window, indicated by the overlaid circle. Then accumulate into orientation histograms summarizing the contents over 4 × 4 sub-regions, as shown on the right, there are four seed points around the key point location.

And each seed point contains eight-direction vector information. In order to enhance the stability of matching, Lowe recommends using 4 × 4 seed points to describe the each key point. Therefore, the experiments in this paper use a 4×4×8=128 element feature vector for each key point.

Step 3 Key points matching

In this paper, the similarity of key points in different images is measured by calculating the Euclidean distance of two key points' feature vector. Let R1(s1，s2…sn) represents the feature vector of key points of original image, and R2(d1,d2…dm) represents the feature vector of key points of the target image. Then,

$$D(i) = \left\| R1(i) - R2(j) \right\| \quad i \in n\,,\ \ j \in m \tag{9}$$

Where m and n represents the number of their key points.

The mean of the formula is to calculate the Euclidean distance of the i-th feature point of original image and the j-th feature point of target image. When the Euclidean distance of the two key points meets that the ratio of the minimum distance and the second close distance is less than a threshold, we consider them as a pair of matching points. In our paper, we choose value of the threshold: 0.7.

3. Experimental Results and Analysis

Firstly, this experiment extracts the feature points of original target object (Figure 4), and than match with different scene (Figure 9), in final, we analyze results of experiment form the quality of feature points extracted, time-consuming of matching and matching precision. Matching accuracy (MA) is defined, then,

$$MA = 1 - \frac{n}{N} * 100\% \ . \tag{10}$$

Where n is the number of matching error, and N is the total of the matching points.
(In this paper, we choose sets of objects to test, due to space limitations, we analysis the results and phenomena according to a group of images. As shown below.)

3.1 Visual quality

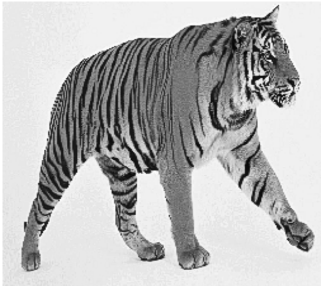

Figure 5. The Original Image

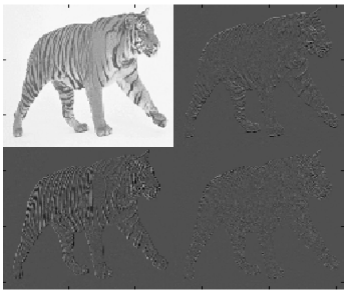

Figure 6. The First-order Decomposed by DWT

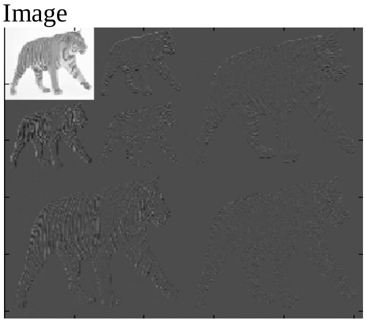

Figure 7. The Second-order Decomposed by DWT

From figure 5 and 6, we can see that the original image is decomposed into four sub-images through the first DWT decomposition. Sub-image at the top-left is the original image of the low frequency component and relatively stable component. While low-frequency sub-image still can be decomposed. As shown in figure 7, we can get more stable low-frequency components

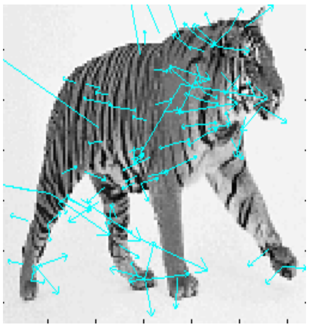

Figure 8. Our Approach

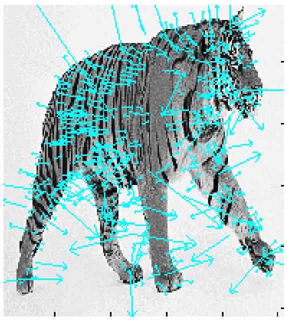

Figure 9. The Original SIFT

Figure 8 and 9 shows that the number of feature points extracted by our algorithm is less than the number of feature points extracted by original SIFT, but it is more stable part in the original feature points. This view can be confirmed in the match.

Figure 10. Matching of the Original SIFT Feature Points.

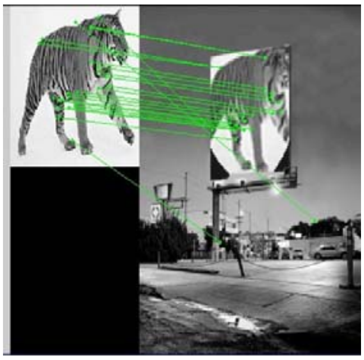

Figure 11. Matching after First-order Decomposition.

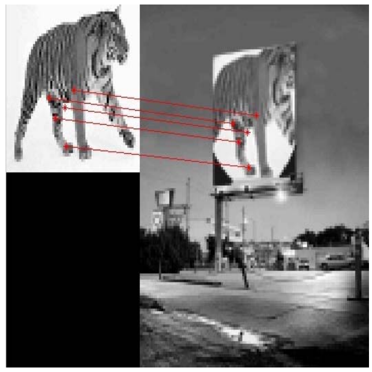

Figure 12. Matching after Second-order Decomposition.

From figure 10, 11 and 12, we can see that the matching accuracy is improved after the first WDT decomposition, but there are still more false match points. After the second DWT decomposition, we can see that matching accuracy has been greatly improved, even reaching to 100%. SIFT feature points of this paper are better quality, more stable, and more accurate matching.

3.2 Analysis of performance

In the experiment, hardware environment: Intel, Pentium Dual-core CPU, 2.6GHz, 1G memory. Development environment: Matlab 7.1 and Matlab and C++ language mixed. Experimental data obtained as follows:

Table 1. Time-consuming of Compute

Groups	Size	AHWT-SIFT cost(/ms)	SIFT cost(/ms)
1st	256*256	1.2	1.7
	441*552	1.7	4.4
2nd	256*256	1.2	2.3
	441*552	2.0	4.4
3rd	256*256	1.4	2.1
	441*552	2.3	4.9

Table 1 shows image size decreases after DWT, so time-consuming of SIFT relatively reduces.

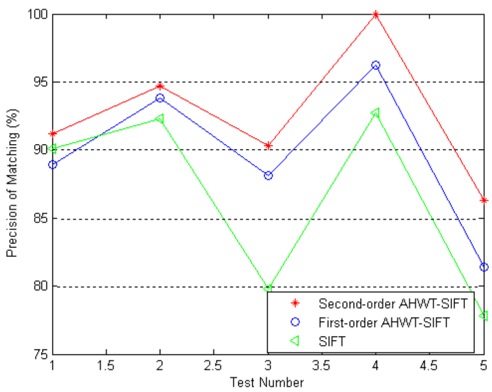

Figure 13. Matching Accuracy

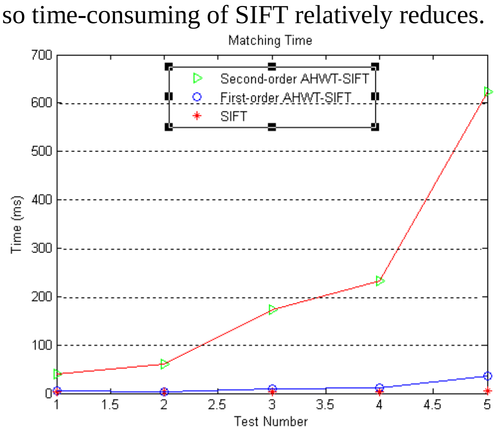

Figure 14. The Time-consuming of Match

From figure 13 and 14, we can see both the matching speed and the matching accuracy are improved, so this method can improves matching speed of the image retrieval.

4. Conclusions

The experiment shows that the adaptive feature points extracted by our approach are less quantity and more stable than that extracted by SIFT. Therefore, our method can improve the speed and accuracy of image matching, and make up the defect that it is low efficiency to retrieve image using SIFT. Although the feature points are extracted from more stable component of image, there are still unstable feature points. Therefore, accurate selection of stable feature points still need further study.

5. Acknowledgment

This work is partially supported by the National Natural Science Foundation of China (No.61103113, No.60903066) and PHR (IHLB) (PHR201008187).

6. References

[1] Mikolajczyk K. , Schmid C , "Indexing based on scale invariant interest points", Proceedings of the 8th International Conference on Computer Vision, Vancouver, pp.525-531, 2001.
[2] Lowe D G. ,"Distinctive image features from scale-invariant keypoints", International Journal of Computer Vision, vol. 60, no. 2, pp.91- 110, 2004.
[3] Mikolajczyk K., Schmid C., "A performance evaluation of local descriptors", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.27, no. 10, pp.1615- 1630, 2005.
[4] Morel J. M., Yu G. S., "Is SIFT Scale Invariant?", Inverse Problems and Imaging, vol.5, no.1, pp.115-136, 2011.
[5] Yi Z., Zhiguo C., Yang X., "Multi-spectral remote image registration based on SIFT", Electronics Letters, vol.44, no.2, pp.107-108, 2008.
[6] Karanwal S, Kumar D and Maurya R. "Fusion of fingerprint and face by using DWT and SIFT", International Journal of Computer Applications, vol.2, no.5, pp.33-36, 2010.
[7] Eleyan A. , Demirel H. "Face recognition using SIFT descriptors extracted from multiresolution Images", 2010 IEEE 18th Signal Processing and Communications Applications Conference (SIU), pp.109-112, 2010.
[8] Ke Y. and Sukthankar R., "PCA-SIFT: A More Distinctive Representation for Local Image Descriptors", Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.506-513, 2004.
[9] Grabner M, Grabner H, and Bischof H. , "Fast Approximated SIFT", Asian Conference on Computer Vision, vol. 1, pp. 918-927, 2006.
[10] Bay H., Tuytelaars T. and Gool L V. , "SURF: speeded up robust features", ECCV, pp. 404-417, 2006.
[11] Bay H, Tuytelaars T, and van Gool L. , "speeded-up robust features (SURF)", Computer vision and image understanding, pp. 346-359, 2008.
[12] Yanyan Gao, Honggang Zhang, and Jun Guo, "A Novel Keypoint Delection in Wavelet Pyramid Space", JDCTA, vol. 5, no. 6, pp. 55-62, 2011.
[13] Chen Mi and Weifeng Lan. ,"Haar Wavelet Transform application in image processing", Ningxia Engineering Technology, 2003.
[14] [14] Michael Z and Xingming Sun, "DWT-PAC(EVD) Based Copy-move Image Forgery Detection", JDCTA, vol. 5, no. 1, pp. 251-258, 2011.