## Research Article
## SIFT Feature Matching Algorithm with Local Shape Context

Gu Lichuan, Qiao Yulong, Cao Mengru and Guo Qingyan
School of Information and Computer, Anhui Agricultural University, China
Anhui Key Laboratories of Agricultural Informatics University, Hefei 230036 Anhui, China

**Abstract:** SIFT (Scale Invariant Feature Transform) is one of the most effective local feature of scale, rotation and illumination invariant, which is widely used in the field of image matching. While there will be a lot mismatches when an image has many similar regions. In this study, an improved SIFT feature matching algorithm with local shape context is put forward. The feature vectors are computed by dominant orientation assignment to each feature point based on elliptical neighboring region and with local shape context and then the feature vectors are matched by using Euclidean distance and the $\chi^2$ distance. The experiment indicates that the improved algorithm can reduce mismatch probability and acquire good performance on affine invariance, improves matching results greatly.

**Keywords:** Elliptical neighboring region, feature matching, local shape context, SIFT algorithm

### INTRODUCTION

Image matching (Rajendra Acharya and Vinitha Sree, 2011) is one of the key technologies of computer vision, image reconstruction, pattern recognition and other fields. Its main task is to extract the stable image features and describe to make these features have characters as distinction, reliability, independence. Algorithms commonly used are: A variety of wavelet algorithm (Evans and Liu, 2006), moment invariants algorithm (Tomas, 2010), corner point detection algorithm (Schmid *et al*., 2000), the geometry feature algorithm (Ding, 2007). Among these, SIFT (Scale Invariant Feature Transform), Lowe (1999, 2004) and Warren and Ghassan (2009) is a better algorithm on the field of feature matching both domestic and international.

The algorithm describe the feature points has invariance of the translation, rotation, scale and illumination However, when an image has many local areas of similarity, the SIFT algorithm will go wrong with a large number of false matching points. Therefore, the SIFT algorithm has been improved in this study, joined the local shape content, making the algorithm contains the shape of the curve within a certain range, thereby reducing the false matching rate. The original algorithm uses a field with a fixed shape, such as circular domain of the Feature points. So it is clear that it cannot handle the geometry distortion brought about by the affine conversion, because the fixed area can not cover the same image content after deformation. Recently Mikolajczyk and Schmid (2004) Pointed out that the feature points of the oval

neighborhood can get better effect with affine (Mikolczyk and Tuytelaars, 2005). Therefore, this study use elliptical neighborhood combined with shape information of the feature point neighborhood area, to describe the feature points, which greatly improved the matching efficiency.

### SIFT ALGORITHM

Lowe (2004) summarizes the existing invariant technology-based feature detection method, come up with a scale-invariant feature descriptor. Firstly, the algorithm make feature detection in the scale space and determine the location of the feature points and the scale of which and then use the main direction of the gradient of the feature point neighborhood, make it as the main direction, in order to achieve that the operator is independent of the scale and direction. The main steps are as follows:

**Detection of extreme points of the scale space:** Lowe use filtering method to detect feature points. For effective detection in scale space, to get a stable feature points, use a Gaussian difference (Difference, of the Gaussian, DoG) function to detect local feature points, the DoG function calculation is convenient and effective, is approximation of Normalized Gaussian Laplacian (Laplacian of Gaussian, LoG), the DoG defined as the difference of the two neighboring different scale Gaussian kernel, DoG operator is the following equation:

$$D(x,y,\sigma)=(G(x,y,k\sigma)-G(x,y,\sigma))*I(x,y)=L(x,y,k\sigma)-L(x,y,\sigma) \quad (1)$$

**Corresponding Author:** Gu Lichuan, School of Information and Computer, Anhui Agricultural University, China Anhui Key Laboratories of Agricultural Informatics, University, Hefei 230036 Anhui, China

Firstly set series of Gaussian image in scale space of each level and then calculate the differential of adjacent Gaussian image to get the Gaussian difference image. Each sampling points need to be compared with a total of 26 pixels, which are in own scale of eight neighborhood and up and down adjacent to the corresponding 3×3 regional of each sampling point ,to ensure the detection of local extreme in scale space and two-dimensional image space simultaneously. Then remove the feature of low contrast points and the edge response point of instability (because of the DOG operator will have a strong edge response), to enhance matching stability, improve the noise immunity.

**Determination of the direction of feature points:** Sample on neighborhood window centered on the feature point and use histogram to statistics gradient direction of neighborhood pixel and then select the histogram peak direction as the direction of the feature points. Gradient direction and magnitude of the neighborhood pixels are calculated by the following equation:

$$m(x,y)=\sqrt{(L(x+1,y,\sigma)-L(x-1,y,\sigma))^2+(L(x,y+1,\sigma)-L(x,y-1,\sigma))^2} \quad (2)$$

$$\theta(x,y)=\tan^{-1}((L(x,y+1,\sigma)-L(x,y-1,\sigma))/(L(x+1,y,\sigma)-L(x-1,y,\sigma))) \quad (3)$$

where, m (x, y) and $\theta(x,y)$ denote the magnitude and direction, the image $\theta(x,y)$ use smooth scales of characteristic point, so that the calculation of scale is invariant. The direction of the gradient orientation histogram is O a 360°, the column histogram, a total of 36, each column contains 10°. Histogram peak position represents the main direction characteristics of the neighborhood Gradient, used as the direction of the feature points.

**Local area description:** First use Axis rotation as the direction of feature point, to ensure the invariance of direction. Then take 16×16 window with feature centered, divide into 16 pieces of 44 Sub-block, on which calculate Gradient accumulator of directions of 00, 450, 900, 1350, 1800, 2250, 2700, 3150 to draw the gradient histogram. A 4×4 sub-blocks can get eight direction descriptors. So, the 16x16 sub-blocks can get the 128 direction descriptors. So that for each feature point can produce a length of 128 data and ultimately appear the formation of the SIFT feature vector with a length of 128.

## THE SIFT ALGORITHM COMBINED WITH THE LOCAL SHAPE INFORMATION

To overcome Shortcomings of these SIFT algorithm, SIFT algorithm has been improved, for each detected feature points, a vector consists of two parts is

set: part of SIFT descriptor to describe the local characteristics by using the elliptical neighborhood and the other part of vector used to distinguish the local shape information between similar local features. Therefore, the improved feature vector of the SIFT algorithm as defined in (4):

$$F = \begin{bmatrix} \alpha\,E \\ (1-\alpha)\,S \end{bmatrix} \quad (4)$$

Here, E is 128-dimensional SIFT vector, S is the 60-dimensional shape vector, $\alpha$ is the relative weighting factor.

**Use the SIFT descriptor of the elliptical neighborhood:** For each feature, SIFT is the point of a small circular neighborhood to calculate the dominant gradient direction. Neighborhood size is determined by the point scale, but its shape is not affine invariant, fixed the shape of a region cannot handle the geometric distortion caused by viewpoint changes. Such as that Image structure in a circle after the affine changes may be mapped to an oval area. If we make circle into the corresponding location of the image after the change, you will find the difference between the image structural information contained in the circle and before, this would make any of the same operator distortion. Recently Mikolajczyk and Schmid (2004) pointed out that using the elliptical area to calculate the dominant direction is more stable, because after transformation two elliptical areas are very close. Second-order moment is usually used to describe the local image structure. The second moment of intensity gradient to describe the distribution of the gradient of the local neighborhood of feature points, which can decide the shape of the local neighborhood. So that we can use second-order moments to estimate the elliptical neighborhood of the feature points.

We usually use the histogram to describe the data distribution of the feature point neighborhood. To ensure that each sample point of the elliptical neighborhood of feature points is mapped to the correct block, we have the ellipse parameters from second moments of the point and then the elliptical area is owned by one into a circle. Using the square root of the second-order moments, image data can be mapped to the circle. Each sample point position X fallen on elliptical neighborhood can be mapped to a location $X' = M^{-1/2}X$ within the standardized round. Then, set the main direction based on the normalized circular neighborhood like SIFT, Then sampling the gradient vector of a circular neighborhood of standardization of feature points, then set the establishment of a 4×4 histogram array, each bar in the eight directions, so You
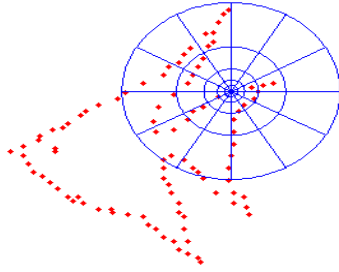
Fig. 1: Shape information diagram

can generate a 128-dimensional feature vector. And then the 128-dimensional vector normalized to unit length.

**Local shape descriptors:** Shape information can be expressed more information which can be distinguished, when an image have many places of local similarity, there will be a matching ambiguity, in order to overcome the shortcomings, Belongie *et al.* (2002) proposed a technique called shape information (shape context) by examining the distribution of points around the point to be matched to characterize the local similarity of the feature points. First feature points to be investigated as shown in Fig. 1, construct a set of pole pairs number grid, using the very purpose of the number of grid to stress that the closer the relative position of feature points have effect on the value of shape information. The information value of the feature points is defined as the number of feature points within each grid and then the shape information of a feature point can be expressed as the distribution measurement of the feature point related.

We use descriptor of the feature points formed by using shape information method. The key issue is to take the neighborhood with the appropriate diameter. Mortensen *et al.* (2005) use Curve of the other features point to describe a feature point, called the global information, but the diameter of the feature point neighborhood is the diagonal length of the entire image, the size of the shape vector function of the entire image size, instead of point of interest and global shape information does not have a full scale invariance, is not a good robustness to the transformation of the image. Here, we use the feature points of the curve near the information to generate the shape of a feature point descriptor, but the diameter of the neighborhood of the feature points is determined by the circular neighborhood of standardized point. The shape information of such a point descriptor has scale invariance.

In standardized circular domain, each curvature of feature points near the principal can be calculated. Given (x, y) the principal curvatures C (x, y) can be

obtained by the 2×2 Hansen matrix H (x, y), it is Hansen matrix's largest Eigen value. Hansen matrix is defined as follows:

$$H(x,y) = \begin{pmatrix} L_{xx} & L_{xy} \\ L_{xy} & L_{yy} \end{pmatrix} = I(x,y) * \begin{pmatrix} g_{xx} & g_{xy} \\ g_{xy} & g_{yy} \end{pmatrix} \tag{5}$$

where,
$L_{xx}$, $L_{yy}$ : Differences of the x y direction of image
$L_{xy}$ : The difference of Cross section

Second derivative can be obtained by Gaussian nuclear convolution with $\sigma_D$ and image. Set e (x, y) equal as the maximum Eigen value of Hansen matrix of the absolute value. The main curvature of the image can be defined as:

$$c(x,y) = |e(x,y)|$$

For each feature point, with its established as the central logarithmic polar coordinates and make the point where the scale standardized circular domain as the described area, along the radial direction is divided into five, divided into 12 constructed along the rotation angle of the direction of pole pairs, so the number of grid divided into 60 regions, the local shape information is a 5×12 histogram, cumulative curvature values in the histogram of each grid. Gaussian function weighted curvature value of each pixel inverted, the weight function (6):

$$w(x,y) = 1 - \exp(-((x-x_0)^2 + (y-y_0)^2 / (2\sigma_D^2))) \tag{6}$$

$(x_0, y_0)$ is the location of the feature points $\sigma_D$ take the same neighborhood weighted with the SIFT local feature scale.

When the scale is small, w (x, y) larger, making a larger proportion of smaller local neighborhood corresponding to the shape information; when the scale is larger, w (x, y) smaller, thereby to reduce the proportion of shape information. Therefore, the weighting function makes the local area can be separate through the description of the shape information Specifically, if the feature points are, $X_0 = (x_0, y_0)^T$ the main direction is $\theta$, then:

$$a = \left[ \frac{6}{\pi} (\arctan(\frac{y-y_0}{x-x_0}) - \theta) \right] \tag{7}$$

$$d = \max(1, \left[ \log_2(\frac{\|x-x_0\|}{r}) + 6 \right]) \tag{8}$$

where,

a, d : The discrete values of the angle and radial distance

r : The radius of the shape information

Let $N_{a,d}$ be the collection of the polar coordinates within the grid point values of a and d, then the grid corresponding to the histogram can be calculated using the following equation:

$$G = \sum_{(x,y) \in N_{a,d}} C(x,y) \qquad (9)$$

$C(x, y)$ is the curvature of the image. Finally, the shape information vector normalization makes the algorithm invariant to illumination changes.

**Image matching:** Based on matching feature points, is to find the point of the same name of the two images, described by the calculation of the two image features vector similarity to determine the feature points of the same name point. Feature vector consists of two parts, this study uses the Euclidean distance as the similarity measurement of the two images SIFT part:

$$d_L = |L_i - L_j| = \sqrt{\sum_k (L_{i,k} - L_{j,k})^2} \qquad (10)$$

$L_{i,k}$ and $L_{j,k}$ are SIFT descriptors in two image's feature points. Statistical test function $\chi^2$ measure part of the similarity of the local shape information:

$$d_s = \chi^2 = \frac{1}{2} \sum_k \frac{(h_{i,k} - h_{j,k})^2}{h_{i,k} - h_{j,k}} \qquad (11)$$

where, $h_{i,k}$, $h_{j,k}$ were the two image feature points $p_i$ and $q_j$ local shape descriptors. And the $d_L$ and $d_S$ smaller, indicating that the two points more match. The final distance of:

$$d = wd_L + (1-w)d_s \qquad (12)$$

With the same as formula (6), used W to control the proportion of local shape information. Here using recent feature points of neighbor and second nearest neighbor distance ratio to reduce the mismatch. According to the ratio of the minimum distance Dist1 between Dist2 the 2 descriptions vectors Dist1/Dist2 to exclude unreliable point. If the nearest distance and near distance ratio is less than a threshold Td that point to match point, otherwise discard. Reduce the threshold to match the point; number will be reduced, but more stable.
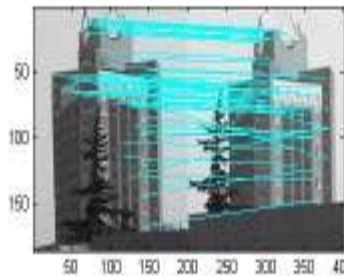
## EXPERIMENTAL RESULTS

To test the accuracy and efficiency of the algorithms, a large number of experiments we have done. Threshold Td is 0.5, the weighting factor of 0.5, the experimental results as shown in the Fig. 1.
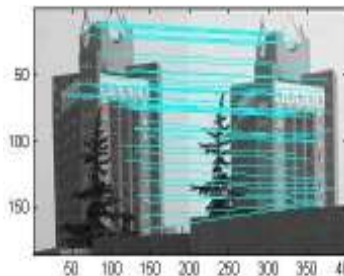


Fig. 2: The original image
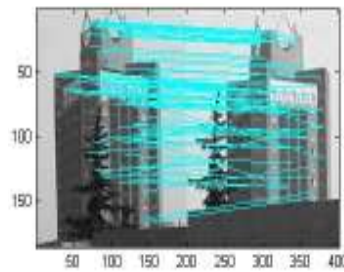


(a) The SIFT method matches



(b) The SIFT+GC method match results



(c) The method of our matching results

Fig. 3: Matching results with three different algorithms

Figure 2 is the original image of size 200 185, although the same building, but the two images are different in scale, perspective and light. These two images match in three ways, shown in Fig. 3, Fig. 3a of the match, the detected feature points 272 LOWE SIFT algorithm, matching the 210 point, remove mismatching points, match points to 112, the correct rate of about 53%; Fig. 3b, Mortensen *et al.* (2005) a combination of global information SIFT algorithm (SIFT+GC) matching results, the correct matching points 158, the correct rate of 75%; Fig. 3c is the result of our algorithm matches the correct match for the 189, the correct rate of 90%. From the experiment can be seen that the invariance of the SIFT algorithm has good scale, pan, light, but a number of similar regions in the image, more mismatch; the SIFT+GC algorithm by adding global information, improved false matching rate of SIFT algorithm, but the range described by the global vector size is fixed and the size of the image does not have the scale invariance; our local shape information vector, its size is standardized by the feature points round the neighborhood, the size of range described by vector can be changed, which has a scale invariance. Figure 3, when the rotation changes, the noise, the perspective changes, our algorithm is better than the classic SIFT algorithm and improve algorithm matching rate higher, especially when the perspective changes greater than 30°, our method matches the rate significantly better than the other algorithms. The matching rate of our algorithm can reach 40% when the viewing angle is changed to 80° while SIFT algorithm is reduced to below 10%, else improved SIFT algorithm matching rate is 20%, showing the changes in perspective of the other algorithms very sensitive, the proposed algorithm improves the stability of the perspective changes, the match rate is greatly improved, significantly improved matching results.

## CONCLUSION

SIFT descriptor has a good scale, rotation, illumination invariant. However, when an image has many local areas of similarity, it appears that a large number of false matching points. This study presents a SIFT matching algorithm with combination of local shape information, use the neighborhood of the ellipse of the point instead of circular field to describe the feature points and join the local shape information of shape descriptors. The theoretical analysis and experimental results show that the improved algorithm can not only solve an image with many similar local region of mismatching, but also can enhance affine invariance of the SIFT algorithm, thereby improving

the matching rate. Algorithm can also be improved in the following areas:

- In the case of color images, these must be converted to gray scale to use this algorithm, if the two images are similar, just different colors, the algorithm would be difficult to achieve good results. The problem of invariance of the color image can be considered.
- The algorithm add the local shape information descriptor, calculation is larger, so the algorithm can be simplified without affecting the effect of the algorithm.

## ACKNOWLEDGMENT

## REFERENCES

Belongie, S, J. Malik and J. Puzicha, 2002. Shape matching and object recognition using shape contexts [J]. IEEE T. Pattern Anal., 24(4): 509-522.

Ding, X.M., 2007. New method for detecting and tracking of moving target based on difference and invariant [J]. Opt. Precision Eng., 15(4): 570-576.

Evans, A.N. and X.U. Liu, 2006. A morphological gradient approach to color edge detection [J]. IEEE T. Image Proc., 15(6): 1454-1463.

Lowe, D.G., 1999. Object recognition from local scale-invariant features [C]. Proceeding of the Seventh IEEE International Conference on Computer Vision, 2(20-27): 1150-1157.

Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints [J]. Int. J. Comp. Vision, 60(2): 91-110.

Mikolajczyk, K. and C. Schmid, 2004. Scale and affine invariant interest point detectors. Int. J. Comp. Vision, 60(1): 63-86.

Mikolczyk, K. and T. Tuytelaars, 2005. A comparison of affine region detectors. Int. J. Comput. Vision, 65(1-2): 43-72.

Mortensen, E., H. Deng and L. Shapiro, 2005. A SIFT descriptor with global context. Proceeding IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1: 184-190.

Rajendra Acharya, U. and S. Vinitha Sree, 2011. Application of recurrence quantification analysis for the automated identification of epileptic EEG signals [J]. Int. J. Neural Syst., 21(3): 199-211.

Schmid, C., R. Mohr and C. Bauckhage, 2000. Evaluation of interest point detectors. Int. J. Comp. Vision, 37(2): 151-172.

Tomas, F., 2010. Mixture of Gaussians exploiting histograms of oriented gradients for background subtraction [C]. ISVC'10 Proceedings of the 6th International Conference on Advaces in Visual Computing, part II, LNCS 6454: 716-725.

Warren, C. and H. Ghassan, 2009. n-SIFT: n-Dimensional scale invariant feature transform. IEEE T. Image Proc., 18(9): 2012-2022.