The Comparison and Analysis of Scale-Invariant Descriptors based on the SIFT

JIAN GAO
College of Electrical Engineering & New Energy
China Three Gorges University
No8. DaXue Road, Yichang, Hubei Province
P.R. CHINA
gaojian@ctgu.edu.cn

Abstract: - Based on the feature matching theory about SIFT (Scale-Invariant Feature Transform) keypoints, the concentric circle structure and the color feature vector of scale-invariant descriptor are proposed in this paper. In the concentric circle structure, the radiuses of the concentric circles are proportional to the scale factor, which can achieve the scale invariance. To achieve the rotation invariance, the coordinates of descriptor are also rotated in relation to the point's orientation. Compared with the square structure of SIFT descriptor, the concentric circle structure not only has simpler computation, but also is more robust to image rotation. The color feature vector chooses the mean values of different color components R, G, B in each subregion of descriptor as the vector's elements. Compared with the gray feature vector of SIFT descriptor, the color feature vector fully utilizes the image's color information, having stronger rotation invariance, and obviously decreasing the vector's dimension, with less computation. After the theory analyses, the experimental results have certified their validity, too.

Key-Words: - computer vision; feature matching; scale-invariant keypoint; scale-invariant descriptor; concentric circle structure; color feature vector;

1 Introduction

Texture feature, line feature and point feature are three classical image features. The texture feature is fit for the processing of image information with obvious textures, and has been widely utilized in the remote sensing image processing. Line feature can be used to describe the beeline or curve figures with regularity in images. Many cutting-edge methods, such as Susan detector, Canny detector, Sobel detector and their extended, have been applied in the image detection and image segmentation. The texture feature and line feature are both the global features and sensitive to image clutter, occlusion and scale changes. As a kind of local feature, the point feature is more adapted to clutter and occlusion. Especially, the scale-invariant point feature is robust to many image transformations, such as illumination changes, image rotation and scale changes, and has been applied in many fields, such as object recognition and classification [1,2], object tracking [3] and panorama building [4].

The scale-invariant keypoints detecting and matching algorithms include three steps: the first step is to detect the interest points invariant to many image transformations. The second step is to use the scale-invariant descriptor to compute the feature vector of each point. The last one is to match the feature vectors in two images' vector spaces.

The Harris corner is one of the classical point features, and has been used in image matching or object recognition successfully. It has strong robustness for many image transformations, but can't be adapted to scale changes. To solve the problem, Carneiro and Jepson [5] have extended the Harris detector to image scale space. Mikolajczyk and Schmid [6] have proposed the Harris-Laplacian detector too. However, the scale-invariant keypoints extended from Harris corner require much computation. Lowe [7] presented the SIFT detector, which selects the local extreme points in the output of DoG (Difference of Gaussian) filter in scale space as the interest points. The SIFT detector and the LoG (Laplacian of Gaussian) detector both have better real time performances.

After the keypoints' location, scale and orientation are assigned, we need to describe every point by feature vector. In fact, the interest points matching in two images are the matching of their feature vectors. According to the biology theory, Lowe [7] presented the SIFT descriptor. It quantizes the gradient orientation histogram values in each subregion to form the feature vector. Mikolajczyk and Schmid [8] have extended the SIFT descriptor to bring the GLOH (gradient location-orientation histogram) descriptor. Yu and Morel [9,10] have also proposed the Affine-SIFT which is robust to affine distortions. Compared with other descriptors, the scale-invariant descriptors based on the SIFT theory perform the best. However, the descriptors are high-dimensional descriptors and evidently increase the complexity of points matching computation. The PCA (Principal Components Analysis) can be used to reduce the matching computation of high-dimensional descriptors [8,11], but need plenty of image samples. Li [12] has introduced the fixed scale feature transformation, which is a scale space building method, to decrease the computation, but it doesn't change the high-dimensional descriptor.

The calculation of SIFT feature vectors mainly utilizes the image's gray information. To utilize the image's color information, Abdel-Hakim and Farag [13] have proposed the CSIFT as a colored point feature descriptor. Alitappeh et al [14] have also introduced the SIFT descriptor in HSI color space to enhance the illumination invariance. Verma et al [15] have extended the SIFT descriptor to different color spaces too. The color SIFT descriptors can identify the color images better and achieve stronger illumination invariance, but the dimensions of their feature vectors are higher, which leads to more computation.

To decrease the calculation complexity and increase the robustness about image transformations in the matching process of scale-invariant keypoints, we present the concentric circle structure and the color feature vector of scale-invariant descriptor in this paper.

On one hand, the new concentric circle structure takes the interest point as the center, divides the sample area into 16 subregions, and rotates the coordinates in relation to the point's orientation. The presented and extended SIFT descriptors [7-10] are all square and divided into 4x4 array subregions or more. The new descriptor is circular, and has less calculation error about image rotation. Therefore, compared with the SIFT descriptor, the new descriptor structure has stronger rotation invariance.

On the other hand, the new color feature vector calculates the mean values of different color components R, G, B in each subregion as the feature vector's elements. The proposed color SIFT descriptors [13-15] all calculate the gradient orientation histogram values on different color component levels. The new descriptor only calculates the mean values of different color components, and has small vector's dimension. Therefore, compared with the SIFT descriptor, the color feature vector has less computation.

This paper is organized as follows: Section 2 introduces the SIFT detector and the LoG detector, which can detect scale-invariant point features. Section 3 proposes the concentric circle structure and the color feature vector of scale-invariant descriptor, and compares them with the square structure and gray feature vector of SIFT descriptor respectively. Experimental results are provided in Section 4 and Section 5 draws the conclusions.

2 Keypoint Detector

In normal conditions, the image transformations, which should be considered in the detection of scale-invariant keypoints, include image translation, illumination changes, image noise, image rotation, scale changes and so on. Especially in real applications, the points' matching should be adequately robust to these usual image transformations. The difference operation has been used to achieve translation invariance and illumination invariance, the isotropic derivative operators have been used to achieve rotation invariance, the Gaussian convolution operation has been used to decrease image noise, and the interest points detecting in image scale space has been an effective approach to achieve scale invariance.

2.1 Scale Space

As shown in Fig.1(a), the image scale space is a set of different representations of one image in different resolution levels. $I(X)$ is the original image and $g(X;\sigma)$ is the Gaussian kernel function. Then, the resolution representation $L(X;\sigma)$ in the scale space of $I(X)$ is created by the Gaussian convolution: $L(X;\sigma) = I(X)*g(X;\sigma)$. The standard deviation σ of $g(X;\sigma)$ represents the scale factor of $L(X;\sigma)$.

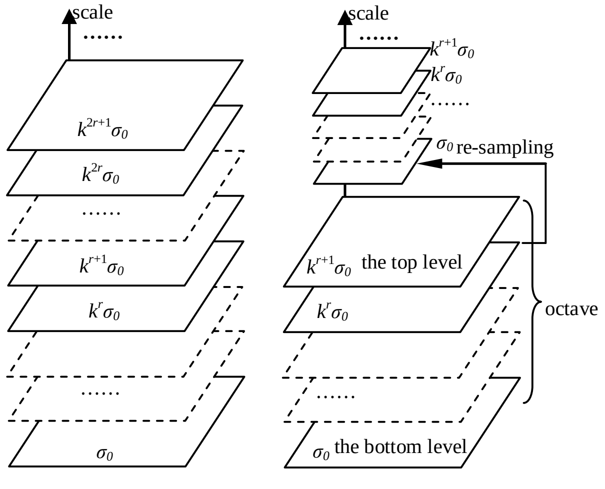


Figure 1 Image scale space.

In real operations, the scale space is sampled in scale orientation and the scale factors of different image resolution levels are distributed exponentially: $\sigma_w = k^w \sigma_0$. σ_0 is the initial scale factor and k is the scale factor ratio between the neighboring image resolution levels in scale space.

The Ref.[16] has analyzed the γ-normalized derivatives' characteristic of image pixel. When γ is equal to 1, the pixel's γ-normalized derivatives lead to perfect scale invariance. Therefore, the pixels, whose values of normalized derivative function are the local extrema in scale space, are scale invariant and often chosen as the interest points.

2.2 The LoG Detector

The LoG detector chooses as the keypoints the pixels whose scale normalized LoG are the local extrema in scale space. The normalized Laplacian is:

$$\sigma^2\left(\partial_{x^2}L(X;\sigma)+\partial_{y^2}L(X;\sigma)\right) \quad (1)$$

The comparing calculation is among the centre point and its 26 neighboring pixels. The 26 neighboring pixels include the 8-neighbourhood pixels at the same level, the 9 corresponding pixels at the immediate lower level and the 9 corresponding pixels at the immediate upper level.

2.3 The SIFT Detector

Because of the Gaussian convolution operation, the scale space building process needs much computation. The SIFT detector utilizes the SIFT pyramid frame, which includes Gaussian pyramid and DoG pyramid, to reduce the computation and storage in scale space [7]. The differences of the neighboring resolution levels in Gaussian pyramid form the resolution levels in DoG pyramid:

$$D(X;\sigma) = I(X)*(g(X;k\sigma) - g(X;\sigma)) \quad (2)$$

The pixels, whose DoG values are local extrema, are detected as the keypoints. Just as the SIFT detector, the LoG detector can be used in pyramid frame to reduce the computation.

Location, orientation and scale are three characteristic parameters of point feature. The location is denoted by the point's coordinates in the original image. The scale is indicated by the scale factor of the corresponding image level. Furthermore, the orientation histogram, which has been presented in Ref.[7], can be used to calculate the point's orientation.

3 Scale-Invariant Descriptor

After the characteristic parameters of all keypoints are computed, we need utilize the local feature descriptors to describe the points. The matching of feature vectors decides the corresponding matching of the points. Therefore, the descriptors not only should cover the distinctive image information around points, but also have strong robustness for many image transformations just like the corresponding keypoints. Corresponding to the scale-invariant points, the descriptors should be scale-invariant. The computation of descriptor includes the design of descriptor's structure and the choice of feature vector's type. The descriptor's structure can confirm the figure and area of local image, which is used to compute feature vectors, around points. The calculation of feature vectors can utilize the image gray information such as the gradient magnitude and orientation, and can also utilize the image color information such as the mean values of different color components.

3.1 Descriptor Structure

The square structure has been selected in the SIFT descriptor. However, the concentric circle structure is proposed in the paper and compared with square structure.

3.1.1 Square Structure

Among the proposed descriptors, the SIFT-based descriptors have been proved better than others [8]. Based on the biological vision theory, the SIFT descriptor, which creates the square structure as shown in Fig.2, takes the keypoint as the center and choices a square area for feature vector computation on the corresponding image resolution level. The total square descriptor area has been divided into 4X4=16 subregions, and the calculation of feature vector is executed in each subregion. On the other hand, the descriptor's coordinates are rotated relative to the point's orientation, and the descriptor's radiuses are proportional to the point's scale factor.

If the 8-bin orientation histogram is selected to compute feature vectors, then the feature vector's dimension of the descriptor with square structure has reached or exceeded 16X8=128. The high-dimensional descriptors would increase the matching accuracy to a certain degree, but would also result in high computational complexity of feature matching in a large database.

3.1.2 Concentric Circle Structure

As shown in Fig.3, a new concentric circle structure is proposed in the paper. According to the angle interval of 90 degree, the inner circle is divided into

4 subregions: 0°~90°, 90°~180°, 180°~270° and 270°~360°. According to the angle interval of 30 degree, the ring between the outer circle and the inner circle is divided into 12 subregions: 0°~30°, 30°~60°, 60°~90°, 90°~120°, 120°~150°, 150°~180°, 180°~210°, 210°~240°, 240°~270°, 270°~300°, 300°~330° and 330°~360°. Therefore, the descriptor includes 16 subregions. If the radius of the outer circle is two times longer than that of the inner circle, then the 16 subregions' area are the same. As shown in Fig.3, according to the anticlockwise orientation, the 4 subregions in the inner circle are respectively named as S1~S4, and the 12 subregions in the ring are respectively named as S5~S16. The reference direction of the descriptor is set to point to the 0° direction.

During the calculation of feature vectors, the keypoint is taken as the descriptor's center to achieve translation invariance. The coordinates of the descriptor are rotated in relation to the point's orientation to achieve rotation invariance. The radiuses of the concentric circles are proportional to the point's scale factor to achieve scale invariance. The Gaussian convolutions are used to reduce the influence of image noise.

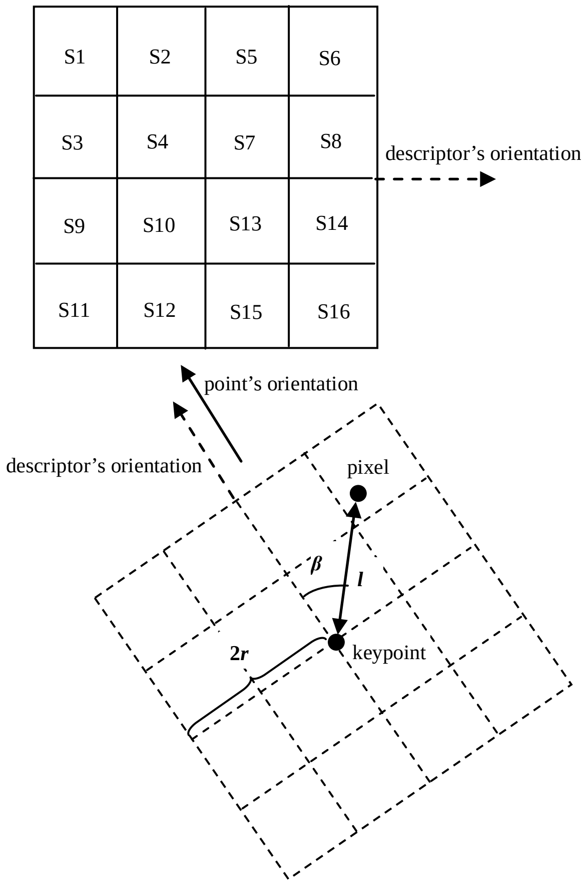


Figure 2 Square structure.

The feature vector computation process of the descriptor with concentric circle structure can be divided into five steps:

Step 1: decrease the influence of image noise, extract the R, G, B color components of original image and make Gaussian convolutions with the R, G, B color values respectively;

Step 2: place the descriptor's center on the location of keypoint, rotate the coordinates of the descriptor to point to the keypoint's orientation, and set the radiuses of the concentric circles k times longer than that of the point's scale factor;

Step 3: build the external square A of the concentric circle descriptor, and set the square's center on the location of keypoint, as shown in Fig.3;

Step 4: select each pixel in square A in turn, and compute the distance l from the pixel to the point and the included angle β from the connection line of the pixel and the point to the descriptor's orientation. According to distance l and angle β, judge which subregion the pixel belongs to;

Step 5: compute the vector elements' values in each subregion to form the feature vector, and normalize the vector elements to unit length.

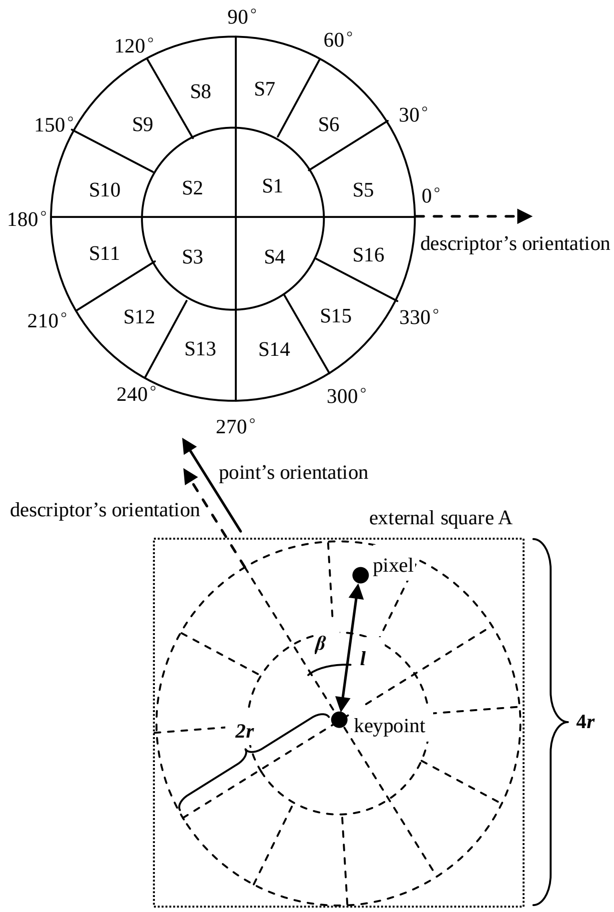


Figure 3 Concentric circle structure.

3.1.3 Comparison of Descriptor Structures

Digital images use the pixel array model to store image information. Because the descriptor would be rotated, the square structure needs to make sure the corresponding relation between sample pixels and descriptor subregion through coordinate transformation. The coordinate transform calculation not only increases the computation, but also brings computation error. According to the distance and angle from sample pixel to keypoint, the descriptor with concentric circle structure judges which subregion the pixel belongs to, which needs no coordinate transform calculation and simplifies the feature vector computation.

3.2 Feature Vector

The gray feature vector is the common feature vector and a new color feature vector is proposed in the paper.

3.2.1 Gray Feature Vector

According to the gray image information, SIFT descriptor computes the 8-bin orientation histogram, which is shown in Fig.4, to obtain the feature vector. Each bin's value of orientation histogram equals to the addition of gradient magnitudes of pixels which fall into this bin. The gradient magnitude and orientation are computed using pixel differences:

$$M(x,y)=\sqrt{\left(I(x+1,y)-I(x-1,y)\right)^2+\left(I(x,y+1)-I(x,y-1)\right)^2} \quad (3)$$

$$\theta(x,y)=\arctan 2\left(D(x,y+1)-D(x,y-1),D(x+1,y)-D(x-1,y)\right) \quad (4)$$

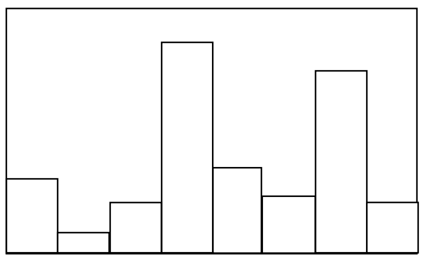


Figure 4 8-bin gradient orientation histogram.

To differentiate the proportion in orientation histogram of sample pixels with different distance to keypoint, the gradient magnitude should be weighted by a Gaussian coefficient. To avoid the boundary affects of neighbouring bins of gradient orientation histogram, the trilinear interpolation is also used to distribute the gradient magnitude of each sample pixel into adjacent histogram bins.

For the 4X4 square descriptor structure which is shown in Fig.2, the 8-bin orientation histograms are computed in each subregion, and we can get 4X4X8=128 dimension feature vectors. For the concentric circle descriptor structure, we can also compute the orientation histograms in 16 subregions and get 128 dimension feature vectors

3.2.2 Color Feature Vector

The feature vector based on the image color information is proposed in the paper. The feature vector selects the mean values of R, G, B color components in each subregion as the vector's elements. The color feature vector has 48 dimension elements for the descriptor with 16 subregions. Compared with the 128 dimension SIFT descriptor, the color feature vector reduces the vector's dimension and simplifies the matching calculation of feature vectors.

To increase the robustness for illumination changes, the color feature vector's elements need to be normalized. However, the normalization process is different from that of gray feature vector which normalizes total elements to unit length. In color feature vector, the vector's elements based on the same color component need to be normalized to unit length respectively, which can enhance the invariance for color spectrum changes in illumination.

The well-known color space models include RGB, YUV, rgb, HSI and so on. According to the real application field, we can choose different color models. However, the color models all have some limitations. In computer vision, to make the analysis tractable, a typical assumption that the object's surfaces are the Lambertian surfaces and only the body reflections are considered. Under this assumption, the pixel's RGB values have no relationship to the viewing direction and the illumination direction. The objects in the real world always can not meet the assumption, but the normal circumstances can meet the assumption approximately. Furthermore, the RGB model can reflect the image information better. Therefore, the color feature vector selects the RGB color model firstly.

Let $I'(X')$ be rescaled from image $I(X)$ by a constant factor f, and then $I'(fX)=I(X)$, $X=(x, y)$. If a γ-normalized homogeneous differential expression assumes a local extremum at $(X_a; \sigma_a)$ in the scale space representation of $I(X)$, the corresponding local extremum in the scale space representation of $I'(X')$ will be located at $(X'_a; \sigma'_a)$, then $X'_a= fX_a$ and $\sigma'_a = f\sigma_a$ [16]. There is a region W, whose area is D_w, in $I(X)$. In the region W, x is in the interval $[x_1, x_2]$ and y is in the interval $[\varphi_1(x), \varphi_2(x)]$. The corresponding region of W in $I'(X')$ is W', where $y' = fy$ and $x' = fx$. In the region W', x' is in the interval $[x'_1, x'_2]$, y' is in the interval $[\varphi'_1(x'), \varphi'_2(x')]$ and the area of W' is $D_{w'}$. If $C(x, y)$ and $C'(x', y')$ denote the same color component (one of R, G, B) value in $I(X)$ and $I'(X')$ respectively, then $C'(x', y') = C'(fx, fy) = C(x, y)$, and we can get:

$$\begin{aligned}
&\left(\int_{x_1'}^{x_2'}\int_{\varphi_1'(x')}^{\varphi_2'(x')} C'(x',y')dx'dy'\right)/D_{w'} \\
&=\left(\int_{x_1'}^{x_2'}\int_{\varphi_1'(x')}^{\varphi_2'(x')} C'(x',y')dx'dy'\right)/\left(\int_{x_1'}^{x_2'}\int_{\varphi_1'(x')}^{\varphi_2'(x')} dx'dy'\right) \\
&=\left(f^2\int_{x_1}^{x_2}\int_{\varphi_1(x)}^{\varphi_2(x)} C(x,y)dxdy\right)/\left(f^2\int_{x_1}^{x_2}\int_{\varphi_1(x)}^{\varphi_2(x)} dxdy\right) \\
&=\left(\int_{x_1}^{x_2}\int_{\varphi_1(x)}^{\varphi_2(x)} C(x,y)dxdy\right)/D_w \\
&\left(x'=fx, y'=fy, C'(x',y')=C(x,y)\right) \qquad (5)
\end{aligned}$$

Therefore, the mean values of color component of the subregion W' in $I'(X')$ are equal to that of the corresponding subregion W in $I(X)$.

3.2.3 Comparison of Feature Vectors

The gray feature vector calculates all 8-bin orientation histograms in each subregion to form higher dimensions vector. It has strong invariance for many image transformations on one hand, increases the computation complexity of vectors matching on the other hand. Furthermore, the gradient orientation is utilized to form feature vector, therefore the gray feature vector is more sensitive to image rotation. The color feature vector calculates the mean values of R, G, B color components in each subregion to form lower dimension vector. It not only decreases the computation complexity of matching, but also influences the robustness to image transformations. The mean values of color components wouldn't decrease the invariance for image rotation, but need to compute on color images.

3.3 Keypoints Matching

After detecting the keypoints and computing their feature vectors in two images, we can use the corresponding point pairs to match the two images. For each keypoint in one image, its matching point in other image has the most similar feature vector. Therefore, the matching of points is to find the most similar feature vector in scale space of two images.

The Euclidean distance comparison method presented in SIFT descriptor is a better method to measure the similarity of feature vectors. The measure method calculates the distance ratio of the closest neighbor and the second-closest neighbor to match point pairs. If the distance ratio is above the threshold, the feature vector is matched with the closest neighbor. If the distance ratio is below the threshold, the closest neighbor is close to the second-closest neighbor and the matching can not be ensured.

After obtaining a set of initial matching pairs, we also need to use the RANSAC (Random Sample Consensus) method to reject the outliers.

4 Experimental Results and Analysis

In the comparison with descriptors' performance, the recall-precision criterion is often used to estimate their performance. RP denotes the number of corresponding keypoint-to-keypoint pairs in two images I_1 and I_2. MP denotes the number of matching point-to-point pairs in two images. TM and FM denote the number of correct matches and false matches respectively. Then, the recall ratio and the 1-precision ratio in recall-precision criterion are as follows:

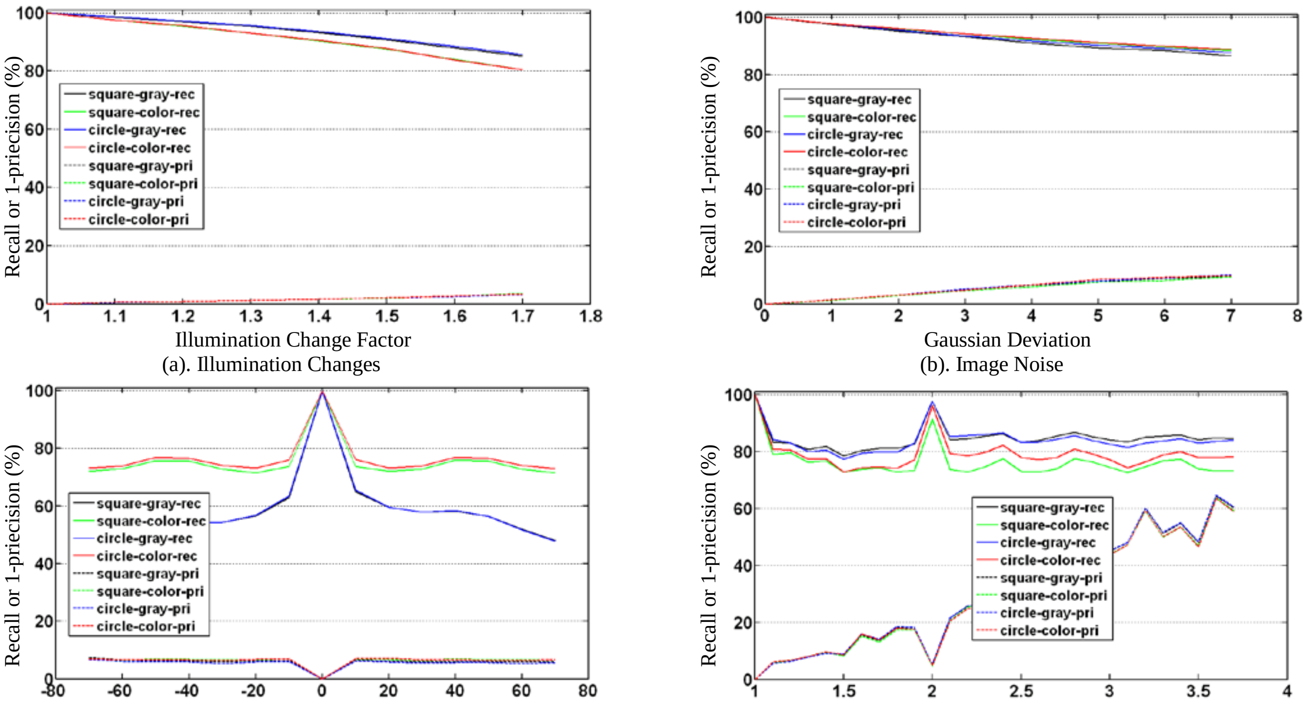
$$recall = \frac{TM}{RP} \quad 1 - precision = \frac{FM}{MP} = \frac{FM}{TM + FM} \qquad (6)$$

The recall ratio is computed as a ratio between the number of correct matches and the number of corresponding point pairs. The 1-precision ratio represents the ratio between the number of false matches and the total number of matches.

When the errors are in location, scale and orientation do not exceed the thresholds, two points are regarded as the corresponding point pair. In our experiment, the location error threshold is 1.5×2^p (p is the octave number in SIFT pyramid) pixels, the scale error threshold is $2^{1/2}\sigma$ (σ is the corresponding scale factor), and the orientation error threshold is 15°.

The experimental results are the statistical data collected from 100 real color scene images and their transformations. We compare and analyze four descriptors which are the combinations of different structures and different feature vectors. The four descriptors include square-gray, square-color, circle-gray and circle-color.

Fig.5(a)-(d) presents the recall and 1-precision of four different descriptors. The square-gray descriptor is shown by the black line, the square-color descriptor by the green line, the circle-gray descriptor by the blue line and the circle-color descriptor by the red line. The solid lines are used to show the recall results, while the dashed lines are used to show the 1-precision results. The image transformations include illumination changes, image noise, image rotation and scale changes. The Euclidean distance is used to measure the similarity, and the distance ratio threshold is set to 0.7.



(a). Illumination Changes

(b). Image Noise

(c). Image Rotation

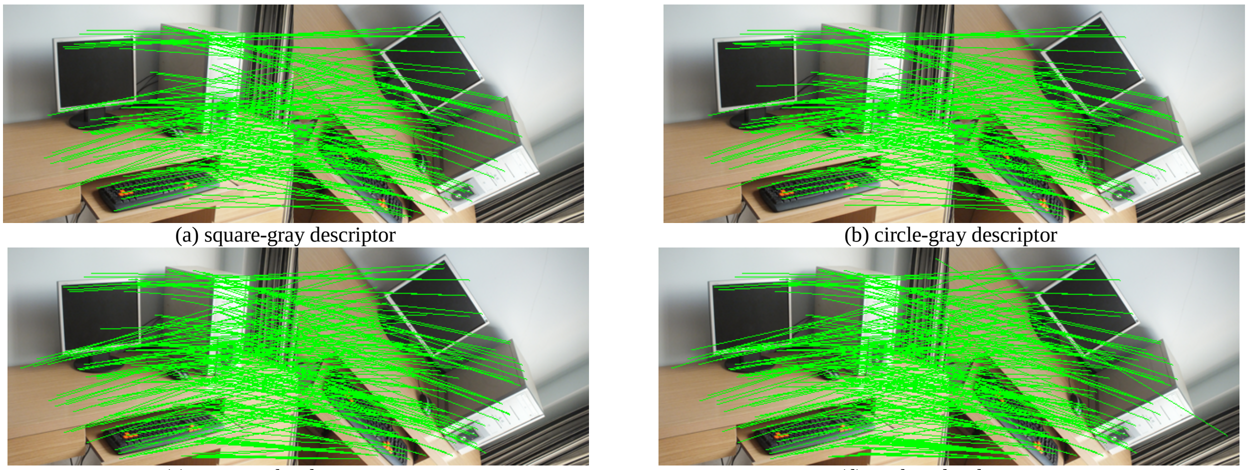
(d). Scale Changes

Figure 5 The recall-precision results of four descriptors.

From the experimental results, we can see that the four combined descriptors have the similar 1-precision results, but different recall results. As to illumination changes, the recall ratios of square-gray descriptor and circle-gray descriptor are basically the same, those of square-color descriptor and circle-color descriptor are also closed, but the recall ratios of two gray feature vectors are greater than those of the two color feature vectors. The descriptor's structure wouldn't influence the robustness for illumination changes. However, the gradient orientation histogram has stronger invariance for illumination changes than that of the mean values of color components. As to image noise, the recall ratios of the four descriptors are similar, but the color feature vector has slightly better recall ratio than that of the gray feature vector. On one hand, the descriptor's structure has less influence on the robustness for image noise. On the other hand, the color feature vector decreases the influence of image noise because of the use of the mean values. As to image rotation, the gradient orientation error caused by image rotation is greater than that of the mean values of color components, therefore the color feature vector has clearly better recall than that of the gray feature vector in terms of image rotation. Compared with the square structure, the concentric circle structure also has slightly better invariance for image rotation. As to scale changes, the recall ratio of the gray feature vector is better than that of the color feature vector, and the image scaling influences the mean values more.

The descriptor's structure has slight influence on the robustness for image transformations, but the calculation of feature vector with concentric circle structure is simpler and more convenient. The gray feature vector has better performance in terms of illumination changes and scale changes, but the color feature vector has obviously stronger robustness for image rotation and slight better invariance for image noise. Compared with 128 dimension gray feature vector, the color feature vector only has 48 dimensions, which can evidently reduce the computation complexity. Furthermore, because of the influence of the re-sampling operation, the result lines in the scale change experiment are flexural.

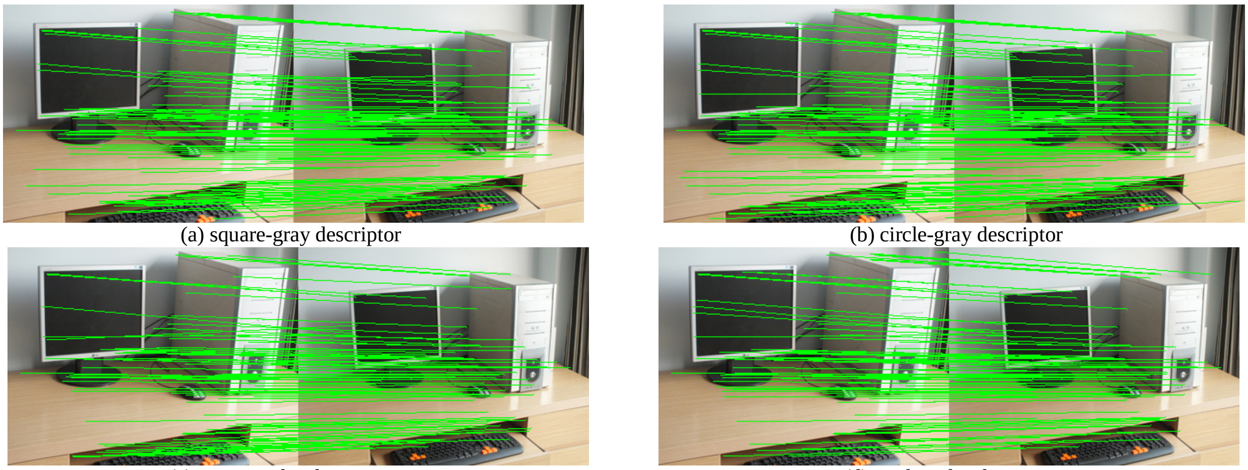
Fig.6 shows the keypoints matching results for image rotation of the four combined descriptors. We use the SIFT detector to gain the SIFT points in two images, and then utilize the four descriptors to match the SIFT points. As to the square-gray descriptor, the number of right matching pairs is 140 and 3 inconsistent matching pairs are rejected by the RANSAC method. As to the circle-gray descriptor, there are 142 right matching pairs and 2 inconsistent matching pairs. The square-color descriptor has 179 right matching pairs and 2 inconsistent matching pairs. The circle-color descriptor has 184 right matching pairs and 3 inconsistent matching pairs. According to the experimental results, we can see that the color feature vector has better robustness for image rotation than that of the gray feature vector.



(a) square-gray descriptor (b) circle-gray descriptor

(c) square-color descriptor (d) circle-color descriptor

Figure 6 The keypoints matching results for image rotation.



(a) square-gray descriptor (b) circle-gray descriptor

(c) square-color descriptor (d) circle-color descriptor

Figure 7 The keypoints matching results for scale change.

Fig.7 shows the keypoints matching results for image scaling of the four combined descriptors. As to the square-gray descriptor, the number of right matching pairs is 131 and 6 inconsistent matching pairs are rejected by the RANSAC method. The circle-gray descriptor has 128 right matching pairs and 7 inconsistent matching pairs. The square-color descriptor has 99 right matching pairs and 11 inconsistent matching pairs. The circle-color descriptor has 102 right matching pairs and 6 inconsistent matching pairs. According to the experimental results, we can see that the robustness for scale change of the gray feature vector is better than that of the color feature vector.

Table 1 shows the keypoints matching time for image rotation and image scaling in Fig.6 and Fig.7 respectively. The matching time of the square-gray descriptor and the circle-gray descriptor are similar. Because of the same dimension of their feature vectors, the matching time of the square-color descriptor and the circle-color descriptor are same too. However, because the color feature vector's dimension is less than the SIFT feature vector's dimension, the matching time of the new color feature vector is less than it of the SIFT feature vector obviously.

Table 1 The keypoints matching time of four descriptors.

	square-gray descriptor	circle-gray descriptor	square-color descriptor	circle-color descriptor
image rotation	297ms	297ms	156ms	156ms
image scaling	266ms	265ms	141ms	141ms

5 Conclusions

In this paper, we have developed the concentric circle structure and the color feature vector of scale-invariant descriptor, and compared their performance with the square structure and the gray feature vector. The concentric circle structure is built by two concentric circles which take the keypoint as the center. The radius of outer circle is twice longer than that of inner circle. According to the angle interval of 90 degree, the inner circle is divided into 4 subregions. On the basis of the angle interval of 30 degree, the ring between the outer circle and the inner circle is divided into 12 subregions. Because the concentric circle structure needn't coordinate transform calculation, it decreases the computation complexity of feature matching. The color feature vector selects the mean values of R, G, B color components in each subregion as the vector's elements, has strong invariance for many image transformations, and especially has strong robustness for image rotation. Moreover, the color feature vector reduces the vector's dimension evidently and reduces the computation of points matching.

References:

[1] T. Yamazaki, T. Fujikawa and J. Katto, Improving the performance of SIFT using bilateral filter and its application to generic object recognition, *Proceeding of the 2012 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2012)*, March. 2012, pp. 945-948.

[2] L. Chang, M. M. Duarte, L. E. Sucar and E. F. Morales, A bayesian approach for object classification based on clusters of SIFT local features, *Expert Systems with Applications*, Vol.39, 2012, pp.1679-1686.

[3] Q. Zhang, T. Rui, H. Fang and J. Zhang, Particle filter object tracking based on Harris-SIFT feature matching, *Proceedings of the 2012 Internatinal Workshop on Infornation and Electronics Engineering (IWIEE 2012)*, March. 2012, pp. 924-929.

[4] M. Zhong, J. Zeng and X. Xie, Panorama stitching based on SIFT algorithm and Levenberg-Marquardt optimization, *Proceedings of the 2012 Internatinal Conference on Medical Physics and Biomedical Engineering (ICMPBE 2012)*, Vol.33, 2012, pp. 811-818.

[5] G. Carneiro and A D. Jepson, Multi-scale phase-based local features, *Proceeding of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2003)*, Vol.1, June. 2003, pp. 736-743.

[6] K. Mikolajczyk and C. Schmid, Indexing based on scale invariant interest points, *Proceedings of the 8th IEEE International Conference on Computer Vision (ICCV 01)*, Vol. 1, July. 2001, pp. 525-531.

[7] D G. Lowe, Distinctive image features from scale-invariant keypoints, *International Journal of Computer Vision*, Vol.60, 2004, pp.91-110.

[8] K. Mikolajczyk and C. Schmid, A performance evaluation of local descriptors, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.27, 2005, pp.1615-1630.

[9] G. Yu and J. M. Morel, A fully affine invariant image comparison method, *Proceeding of the 2009 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2009)*, April. 2009, pp1597-1600.

[10] J. M. Morel and G. Yu, ASIFT: A new framework for fully affine invariant image comparison, *SIAM Journal on Imaging Sciences*, Vol.2, 2009, pp438-469.

[11] Y. Ke and R. Sukthankar, PCA-SIFT: a more distinctive representation for local image descriptors, *Proceeding of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2004)*, June. 2004, pp506-513.

[12] J. Li, An image feature point matching algorithm based on fixed scale feature transformation, *Optik*, Vol.124, 2013, pp.1620-1623.

[13] A. E. Abdel-Hakim and A. A. Farag, CSIFT: a SIFT descriptor with color invariant characteristics, *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2006)*, June. 2006, pp1978-1983.

[14] R. J. Alitappeh, K. J. Saravi and F. Mahmoudi, A new illumination invariant feature based on SIFT descriptor in color space, *Proceedings of the Internatinal Symposium on Robotics and Intelligent Sensors 2012 (IRIS 2012)*, September. 2012, pp. 305-311.

[15] A.Verma, C.Liu, and J.Jia, New colour SIFT descriptors for image classification with applications to biometrics, *International Journal of Biometrics*, Vol.3, 2011, pp56-75.

[16] T. Lindeberg, Feature detectionwith automatic scale selection, *International Journal of Computer Vision*, Vol.30, 1998, pp.79-116.