

A Comparative Study of Local Image Feature Description

Lifen Zhou^{1, a *} and Yanlin Tao^{2, b}

^{1, 2} School of Information Engineering, Qujing Normal University, Qujing 655011, Yunnan, China

^a 872053339@qq.com, ^b 401004120@qq.com

Keywords: Local feature; Feature detection; Feature description; SIFT

Abstract. In view of the local image feature description algorithm of recent year, we compare them with SIFT, which is a classic algorithm. We found the gap of every kind of description algorithm with some image transformations, and this can be a reference when feature description algorithm used in some application of computer vision. Our results show that MROGH and MRRID has the best performance, and then is LIOP, DAISY and HRI-CSLTP are better than SIFT in some image transformation.

Introduction

Image local invariant feature refers to the image features of the image is invariant to image rotation, scale transformation, affine transformation, perspective transformation, image blur, image compression etc. Because of its excellent properties have been widely used in various applications of computer vision and pattern recognition, such as target recognition [1], wide baseline matching [2], image retrieval [3], panoramic image mosaic [4].

Image local feature description is an important stage of local feature extraction, it refers to the image feature area to detect the feature description into regional description vector, so the more convenient for the following use. The simplest feature description method is to transform the pixels in the feature region directly into the description vector, but it is not widely used because of its poor performance. A description of local image features, which most attention is put forward by Lowe SIFT (Scale invariant feature transform) [1]. SIFT algorithm because of its superiority has been widely used, and many scholars have made improvements, such as PCA-SIFT[5], GLOH[6], SURF[7] etc. In recent years, there are many new characteristics description methods have emerged, such as the CS-LBP[8] center symmetric local binary value model LBP (local binary pattern structure descriptors, HRI-CSLTP[9] will CS-LBP expansion into three value coding, the descriptor has a stronger ability to distinguish. DAISY[10] is similar to SIFT and GLOH descriptors, but the calculation is faster than SIFT, and effectively used in wide baseline stereo matching. According to the SIFT algorithm is required to estimate the direction of the main shortcomings, Bing fan, Wang Zhenhua etc. [11-13] studied a new descriptor design framework, it does not need to estimate the main direction, using this framework MROGH (Multisupport region Order-Based gradient histogram) MRRID (Multisupport Region Rotation and Intensity Monotonic Invariant Descriptor) and LIOP (Local Intensity Order Pattern) algorithm is proposed.

Based on several classic good performance description algorithm are compared and analyzed. By comparing the performance gap between the various descriptors, when these feature description algorithm was applied to various applications in computer vision can be used as reference.

Local Image Feature

Image local invariant feature extraction includes two steps: feature detection and feature description. Feature detection is to detect the characteristics of location, characteristics of general feature detection as feature points or feature regions. Feature description is the representation of the feature. There are many classic algorithms, the classical algorithm reference [14] and [15] respectively for feature detection and feature description are discussed and compared. [15] The new algorithm does not contain some of the good performance in recent years, so this paper a novel algorithm for recent

comparative study, including: DAISY, HRI-CSLTP, LIOP, MROGH and MRRID, these five kinds algorithm and SIFT algorithm were compared.

Local Image Feature Description Algorithm

DAISY Algorithm. DAISY [10] descriptors are Gaussian kernel convolution and weighted based on obtained. For an image, the image is first calculated in each point of the H gradient direction, indicated $G_o(u, v)$ by the dot (u, v) in the direction o of the gradient, the gradient map is expressed as:

$$G_o = \left(\frac{\partial I}{\partial o} \right)^+ \quad (1)$$

where I represents the image, $(\bullet)^+$ represents $(a)^+ = \max(a, 0)$.

Then the gradient map and Σ different values of the Gauss kernel convolution, so as to realize the multi-scale space, we obtain different scale convolution direction map:

$$G_o^\Sigma = G_\Sigma * \left(\frac{\partial I}{\partial o} \right)^+ \quad (2)$$

For a point in the image gradient direction histogram:

$$h_\Sigma(u, v) = [G_1^\Sigma(u, v), G_2^\Sigma(u, v), \dots, G_H^\Sigma(u, v),]^T \quad (3)$$

where $G_1^\Sigma, G_2^\Sigma, \dots, G_H^\Sigma$ represents Σ In different directions of the gradient map, then the normalized processing is obtained $\tilde{h}_\Sigma(u, v)$.

If the number of different circles Q , DAISY is the descriptor:

$$\begin{aligned} D(u_0, v_0) = & [\tilde{h}_{\Sigma_1}^T(u_0, v_0), \tilde{h}_{\Sigma_1}^T(I_1(u_0, v_0, R_1)), \dots, \tilde{h}_{\Sigma_1}^T(I_T(u_0, v_0, R_1)), \\ & \tilde{h}_{\Sigma_2}^T(I_1(u_0, v_0, R_2)), \dots, \tilde{h}_{\Sigma_2}^T(I_T(u_0, v_0, R_2)), \dots \\ & \tilde{h}_{\Sigma_Q}^T(I_1(u_0, v_0, R_Q)), \dots, \tilde{h}_{\Sigma_Q}^T(I_T(u_0, v_0, R_Q)),] \end{aligned} \quad (4)$$

where $I_j(u, v, R)$ represents the position of the point (u, v) in the direction j from which it is located R .

HRI-CSLTP Algorithm. HRI-CSLTP[9] descriptor is obtained by CS-LBP descriptor is extended to three value encoding, CS-LBP descriptor is describing the algorithm local two value model based on LBP, to a certain point (x, y) , the local two value model is defined as LBP:

$$LBP_{R,N}(x, y) = \sum_{i=0}^{N-1} s(n_i - n_c) 2^i \quad (5)$$

where $s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & \text{otherwise} \end{cases}$, n_c As the point (x, y) pixel values, n_i as the center point,

R as the uniform sampling radius of the spot value, N as a total of sampling points.

If the local two value model is defined as the descriptor, when the sampling point is much larger dimension of the descriptor, in this case, Heikkila proposed CS-LBP two value model of center symmetry:

$$CS-LBP_{R,N,T}(x,y) = \sum_{i=0}^{N/2} s(n_i - n_{i+(N/2)}) 2^i \quad (6)$$

where $s(x) = \begin{cases} 1, & x > T \\ 0, & \text{otherwise} \end{cases}$, T Is the selection of threshold, Two points on the point

(x, y) of symmetry of the pixel values are compared, if the difference is greater than the threshold value, the corresponding value is 1, otherwise 0.

In order to enhance the performance of Gupta descriptors, CS-LBP will be extended to the three value model, obtained the CS-LTP descriptors:

$$CS-LTP(p,D,T) = f(n_0 - n_4) + f(n_2 - n_6) \times 3 \quad (7)$$

where $f(x) = \begin{cases} 0 & x < -T \\ 2 & x > T \\ 1 & \text{else} \end{cases}$, p as the center of the pixel, D as the radius of the

neighborhood, T as the selected threshold.

According to pixels in the regional characteristics of the relevant order histogram HRI (histogram of relative Intensities) block, for each local block calculated descriptor CT-LTP descriptor HRI-CSLTP.

LIOP Algorithm. LIOP[13] descriptor is the character of the detected (standardized round) in accordance with the pixel ordering of sub blocks, each point of calculating the description vector of the local coordinate system for a pixel point x , If sampling N pixels in the neighborhood, the corresponding pixel value: $P(x) = \{I(x_1), I(x_2), \dots, I(x_N)\}$, for the N pixel value, in accordance with the order of ranking from small to large, the numbers form a bit N sequence π :

$$\pi = \gamma(P) = (i_1, i_2, \dots, i_N), I(x_{i_1}) < I(x_{i_2}) < \dots < I(x_{i_N}) \quad (8)$$

Then an index table, the first column index table is possible in order, the second column is the corresponding index value, There are sort of $N!$ situation The corresponding index value $Ind(\pi)$ found by the serial number π , will be in accordance with the type $N!$ is converted to a bit vector:

$$\phi(\pi) = V_{N!}^{Ind(\pi)}, \pi \in \Pi \quad (9)$$

where $V_{N!}^{Ind(\pi)}$ is only the $Ind(\pi)$ number is 1, the rest are 0 about the $N!$ dimensional vector, LIOP descriptors such as shown in formula (10).

$$\begin{aligned} LIOP(x) &= \phi(\gamma(P(x))) = V_{N!}^{Ind(\gamma(P(x)))} \\ &= (0, \dots, 0, \underset{(Ind(\gamma(P(x))))}{1}, 0, \dots, 0) \end{aligned} \quad (10)$$

Then each sub region in the calculation of descriptors are accumulated by the sub region descriptors, finally are sequentially connected by sub region descriptors obtained the final descriptor, given by (11):

$$\begin{aligned} LIOP \text{ descriptor} &= (des_1, des_2, \dots, des_B) \\ des_i &= \sum_{x \in bin_i} LIOP(x) \end{aligned} \quad (11)$$

where bin_i express the sub region, des_i express sub region descriptor, B as the number of regional characteristics of divided regions.

MROGH and MRRID Algorithm. MROGH [11, 12] and MRRID descriptor is local descriptor based on multi support, which is calculated in different regions of multiple descriptors on the size . The establishment of LIOP and the same local coordinates.

MROGH is a description method of gradient based feature points according to the calculation X_i of the amplitude and phase of the formula:

$$m(X_i) = \sqrt{D_x(X_i)^2 + D_y(X_i)^2} \theta(X_i) = \tan^{-1}(D_y(X_i)/D_x(X_i)) \quad (12)$$

where $D_x(X_i) = I(X_i^1) - I(X_i^5)$, $D_y(X_i) = I(X_i^3) - I(X_i^7)$, $I(X_i^j)$ representation of pixel values, The $[0, 2\pi)$ angle of

the range is divided into a d column, $\text{dir}_i = (2\pi/d) \times (i - 1)$, $i = 1, 2, \dots, d$, so that the feature vector of the point X_i is:

$$F_G(X_i) = (f_1^G, f_2^G, \dots, f_d^G) \quad (13)$$

where $f_j^G = \begin{cases} m(X_i) \frac{(2\pi/d - \alpha(\theta(X_i), \text{dir}_j))}{2\pi/d}, & \text{if } \alpha(\theta(X_i), \text{dir}_j) < 2\pi/d \\ 0, & \text{otherwise} \end{cases}$, $\alpha(\theta(X_i), \text{dir}_j)$ which

represents the angular difference between $\theta(X_i)$ and dir_j .

MRRID is described based on the feature of strength .Assuming the sampling $2m$ sampling points in the points X_i around: X_i^j , $j = 1, 2, \dots, 2m$, by pixel comparison of point X_i symmetry point value, a dimension vector m is obtained: $(\text{sign}(I(X_i^{m+1}) - I(X_i^1)), \text{sign}(I(X_i^{m+2}) - I(X_i^2)), \dots, \text{sign}(I(X_i^{m+m}) - I(X_i^m)))$, the point X_i of the feature vector is:

$$F_I(X_i) = (f_1^I, f_2^I, \dots, f_{2m}^I) \quad (14)$$

where

$$f_j^I = \begin{cases} 1, & \text{if } \sum_{k=1}^m \text{sign}(I(X_i^{k+m}) - I(X_i^k)) \times 2^{k-1} = (j-1) \\ 0, & \text{otherwise} \end{cases} \quad \text{sign}(x) = \begin{cases} 1, & x > 0 \\ 0, & \text{otherwise} \end{cases}$$

Comparison and Analysis of Experiment

Experimental Data and Evaluation Standard. The standard image database in [16] to evaluate the performance of the algorithm, select the database contains the following five kinds of representative geometry and illumination transform image: perspective changes, scale changes, image rotation, image blur, illumination changes, in which each includes 6 images.

To evaluate the use of descriptors in the [6] literature evaluation standard, which is based on the number and the number of errors, Recall-Precision criteria for evaluation of correct matching of two images. The matching strategy is to use the Euclidean distance as the similarity measure, using the nearest and second nearest neighbor distance ratio (Nearest neighbor distance ratio, NNDR) [1] Calculate the recall rate (recall) and precision (precision) according to the formula (15).

$$\text{recall} = \frac{\#correct \ matches}{\#correspondences} \quad 1 - \text{precision} = \frac{\#false \ matches}{\#all \ matches} \quad (15)$$

where $\#correspondences$ is right, the number of real, calculated in accordance with the method in literature[17].

Experimental Results. In order to compare the performance of all descriptors, this paper uses affine invariant feature detection method to extract the features of the test image, then according to various descriptors will feature regional transformation as the feature vector, by comparing the performance of RP curve to compare to describe all kinds of performance, 5 in this paper, the descriptor algorithm and the classical SIFT algorithm is compared and experimental results as shown in Figure 1-8. RP curve horizontal coordinates: $1 - precision$, Vertical coordinate: $recall$. The general case is a gradually increasing with the increase of the curve, a perfect descriptor will get in any precision recall rate is 1 of the curve, more close to this curve, the corresponding performance is more outstanding.

Figure1-8 shows the experimental results including blurred images, image scaling, rotation, perspective transformation and illumination changes, each transformation shows the transformation degree is smaller and larger two amplitude image results, as shown in Figure bikes1-2 said bikes image group in the first picture image and the second image performance curve, the bikes image group belongs to the image fuzzy, fuzzy degree gradually increased. From Figure 1 and 2 the results can be seen in the image fuzzy transformation, five kinds of descriptors are reviewed in this paper the performance of SIFT algorithm is better than the classic, including LIOP, MROGH and MRRID showed a performance more prominent.

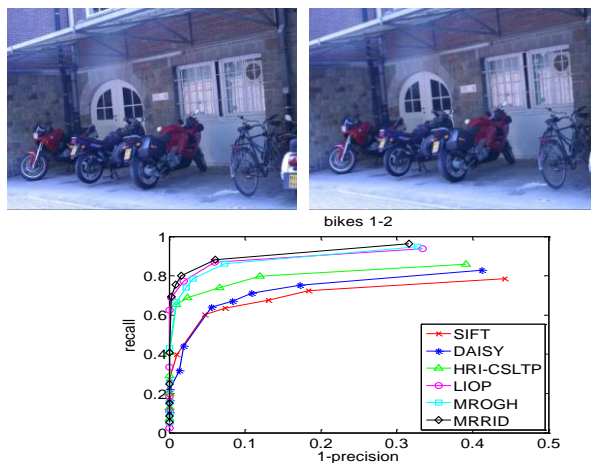


Figure 1. bikes1-2

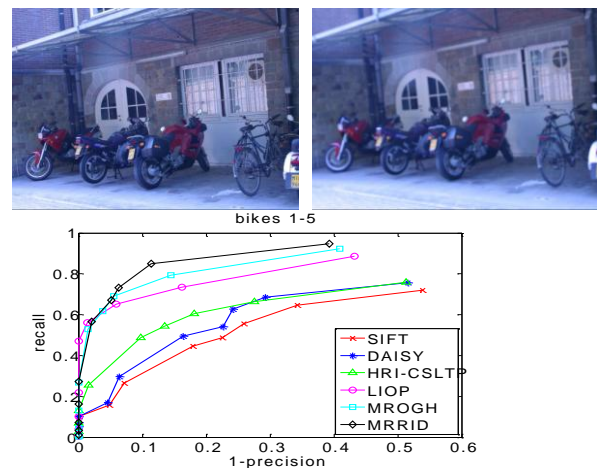


Figure 2. bikes1-5

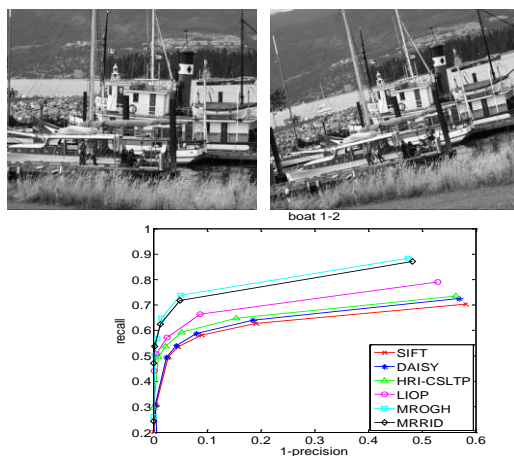


Figure 3. boat1-2

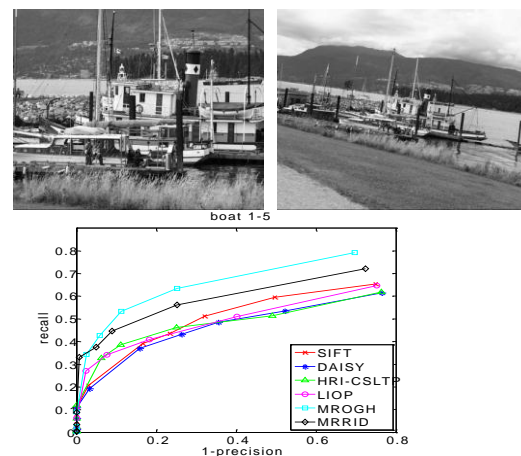


Figure 4. boat1-5



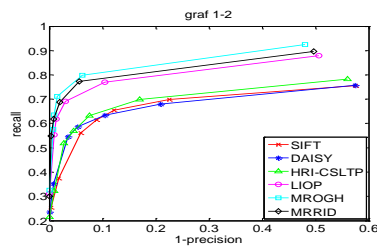


Figure 5.graf1-2

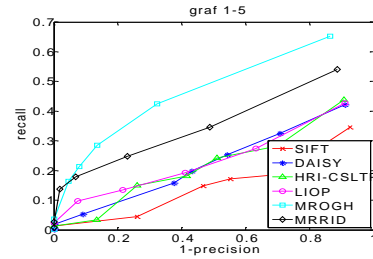


Figure 6.graf1-5

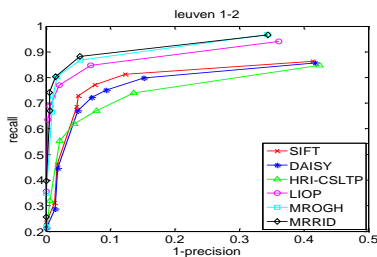


Figure 7. leuven1-2

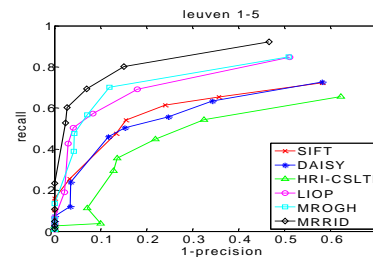


Figure 8. leuven1-5

From all the experimental results can be seen in all of the MROGH image transform and MRRID show the performance of the most prominent, the second is LIOP DAISY in scale transformation, rotation transformation and perspective transformation in SIFT and HRI-CSLTP with similar performance, in the light of the poor performance of HRI-CSLTP transformation.

Conclusions

In recent years, a novel five descriptors have been reviewed in this paper, they will be with classical SIFT algorithm were compared image transform RP performance curve can be seen MROGH and MRRID performance of the algorithm is best, next is the descriptor liop, image transform Daisy and HRI-CSLTP than sift the performance better. In the dimension of the descriptor, HRI-CSLTP, MROGH and MRRID of large dimension, subsequent use slower calculating speed. The time required for the calculation of a single descriptor, because the method of MROGH and MRRID using multi computing descriptors, the computing time is longer. Because of the characteristics of the time required for various descriptors in performance, dimension and structure of the descriptor, for different applications can choose the most appropriate description algorithm.

Acknowledgements

We acknowledge partial funding by yunnan Provincial Department of education project (2013C013) and Qujing Normal University research project (2012QN024).

References

- [1] LOWE D G. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 2004, 60(2): 91- 110.

- [2] TUYTELAARS T, VAN GOOL L. Matching widely separated views based on affine invariant regions[J]. *International Journal of Computer Vision*, 2004, 59(1): 61-85.
- [3] NISTER D, STEWENIUS H. Scalable recognition with a vocabulary tree[A]. *IEEE Conference on Computer Vision and Pattern Recognition*[C]. New York, NY, USA, 2006. 2161-2168.
- [4] BROWN M, LOWE D G. Automatic panoramic image stitching using invariant features. *International Journal of Computer Vision*, 2007, 74(1): 59-73.
- [5] KE Y, SUKTHANKAR R. PCA-SIFT: A more distinctive representation for local image descriptors[A]. *IEEE Conference on Computer Vision and Pattern Recognition*[C]. Washington, DC, USA, 2004. 506-513.
- [6] MIKOLAJCZYK K, SCHMID C. A performance evaluation of local descriptors. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 2005,27(10): 1615-1630.
- [7] BAY H, TUYTELAARS T, VAN GOOL L. Surf: Speeded up robust features[M]. Berlin Heidelberg. Springer-Verlag, 2006.
- [8] HEIKKILA M, PIETIKAINEN M, SCHMID C. Description of interest regions with local binary patterns. *Pattern recognition*, 2009, 42(3): 425-436.
- [9] GUPTA R, PATIL H, MITTAL A. Robust order-based methods for feature description[A]. *IEEE Conference on Computer Vision and Pattern Recognition* [C]. San Francisco, CA, USA, 2010. 334-341.
- [10] TOLA E, LEPETIT V, FUA P. Daisy: An efficient dense descriptor applied to wide-baseline stereo[J]. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 2010, 32(5): 815-830.
- [11] Fan B, Wu F, Hu Z. Aggregating gradient distributions into intensity orders: A novel local image descriptor[C]. *IEEE Conference on Computer Vision and Pattern Recognition*, 2011: 2377-2384.
- [12] FAN B, WU F, HU Z. Rotationally invariant descriptors using intensity order pooling. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 2012, 34(10): 2031-2045 .
- [13] WANG Z, FAN B, WU F. Local intensity order pattern for feature description[C]. *IEEE International Conference on Computer Vision*, 2011: 603-610.
- [14] Mikolajczyk K, Schmid C. Scale & affine invariant interest point detectors. *International journal of computer vision*, 2004, 60(1): 63-86.
- [15] Mikolajczyk K, Schmid C. A performance evaluation of local descriptors. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 2005, 27(10): 1615-1630.
- [16] Information on <http://www.robots.ox.ac.uk/~vgg/research/affine/>.
- [17] MIKOLAJCZYK K, TUYTELAARS T, SCHMID C, et al. A comparison of affine region detectors. *International Journal of Computer Vision*, 2005, 65(1-2): 43-72.