GDBID: Fusion gradient distinction binary image descriptor

Jihui Qi*, Changbo Xu
School of Information Engineering, Beijing Institute of Graphic Communication, Beijing, China

ABSTRACT

In recent years, binary descriptors have attracted more and more attention due to their low memory consumption and high speed. It is well known that these representations are worse than higher-dimensional and histogram-based descriptors such as SIFT. Therefore, this paper proposes a fusion gradient distinction binary image descriptor (GDBID). Gradient comparison is added on the basis of the original gray comparison to enrich the information contained in the descriptor. At the same time, the comparison patches of different sizes are obtained by constructing concentric circles to achieve anti-noise. In addition, a threshold is set to filter patches to reduce the dimension of descriptors. Experimental results show that the GDBID has a precision is close to the best algorithm (SIFT), and the time consumption is lower than the fastest ORB in the literature.

Keywords: Binary descriptors, patch size, gray gradient

1. INTRODUCTION

How to represent the local image more effectively is the key to the widespread application of computer vision. These applications include 3D-reconstruction[1], SLAM mapping[2], image retrieval[3] and pose estimation[4]. Local image representation is the most commonly used image representation method, because local characteristics are unique, robust to partial occlusion, viewpoint invariant, and highly effective because they abandon low-information regions. In order to get a local image representation, it is important to extract a set of significant image structures and provide descriptions of each structure. For various structures such as corners, segments, lines and regions[5], the real valued or binary descriptors is used. In addition, the binary method is the fastest method for extraction and matching. This paper addresses the question of optimal binary descriptors.

The SIFT has been proposed for 20 years[6], but it is still recognized as the best technology. However, HPatches benchmarks suggest that improvements can still be made. The deep model-based descriptors have improved the mean Average Precision (mAP) of different tasks among different tasks[7], but at the cost of a sharp increase in computing consuming. This make them unusable on devices with limited hardware and batteries, such as smartphones. With the introduction of a large number of novel descriptors, real-time performance can be achieved on devices with limited resources. The cost is significantly lower accuracy than SIFT.

In order to obtain high quality and low computational descriptors, a variety of binary algorithms have been proposed. In 2010, Calonder proposed the BRIEF[8], which has been widely used in mobile devices due to its significant real-time advantage. Although BRIEF is 100 times faster than SIFT and 10 times faster than SURF, its lack of anti-interference ability makes it difficult to be applied to high-quality matching. Besides, several methods have been proposed such as rBRIEF[9], BRISK[10], FREAK[11], LDB[12], M-LDB[13] and LATCH[14]. These algorithms improve the quality of descriptors from the aspects of denoising, comparison selection mode, visual mechanism, gradient information, image filtering method and comparison quantity. Because these descriptors use overly simplified message, that is, binary comparison of the original gray values of the subset of pixels in the image patch, the discrimination is lower. When matching with large databases, the lack of differences will lead to a large number of false matches. Expensive post-validation methods (such as RANSAC[15] or PROSAC[16]) are often needed to verify the matching results, which increases the running time of the whole process.

This paper introduces a new binary descriptor called fusion gradient distinction binary image descriptor (GDBID). It has similar robustness and speed to the most advanced binary descriptors, but offers a greater distinction than them. The GDBID is implemented through three schemes. First, GDBID uses both gauss average intensity (I_{Gauss}) and first-order

[*] 1807919037@qq.com

Third International Conference on Computer Science and Communication Technology (ICCSCT 2022)
edited by Yingfa Lu, Changbo Cheng, Proc. of SPIE Vol. 12506, 125063W
© 2022 SPIE · 0277-786X · doi: 10.1117/12.2661827

gradients (d_x and d_y) of grid cells in the image patch. Secondly, the comparison cell with different sizes are constructed on the circular neighborhood. To reduce the dimension of the descriptor, a distance threshold is set to filter short distance subsets for test. Third, because the neighborhood is symmetric, the similar position has the same meaning. GDBID takes the four comparison pairs as a tuple, and loops through I_{Gauss}, d_x and d_y to generate string.

2. FUSION GRADIENT DISTINCTION BINARY IMAGE DESCRIPTOR

2.1 Comparison pairs

The BRIEF is the most widely used binary descriptor. It only considers the grayscale relationship of pixel pairs and ignores the grayscale distinction information between adjacent pixels[17]. a lot of useful information is ignored. Therefore, if we can make full use of the gray size and gray distinction information of pixel pairs, a descriptor of higher quality can be obtained. As we all know, gradient is more robust than gray value when dealing with brightness changes. And gradient can sensitively capture gray changes and objectively reflect the changes of images in a certain direction. In addition, the gradient can be calculated using a box filter[18] and accelerated with an integral image. After comprehensive consideration, this paper introduces the first-order gradients as an index to evaluate the difference of gray value. This paper defines three variables for the t pixel pair $(p_{t,1}, p_{t,2})$.

$$I_{Gauss}(t) = Gauss(p_{t,1}) - Gauss(p_{t,2}) \tag{1}$$

$$d_x(t) = Gradient_x(p_{t,1}) - Gradient_x(p_{t,2}) \tag{2}$$

$$d_y(t) = Gradient_y(p_{t,1}) - Gradient_y(p_{t,2}) \tag{3}$$

where $Gradient_x$ and $Gradient_y$ are the regional gradient of pixel pair t in the x and y directions, respectively. $Gauss$ represents the gray value of the corresponding pixel cell after Gaussian filtering.

2.2 Sampling pattern

The size of the pixel cell affects the robustness and uniqueness of the descriptor. On the one hand, small size is more sensitive to gradient changes, and descriptors can capture more detailed changes with higher resolution. On the other hand, large size contains more grayscale information and descriptors are more robust, but not sensitive to detail changes. As shown in Figure 1, we obtain the set of comparison pairs by constructing concentric circles. For the sampling points on the same circle, the same radius is selected to construct the circular neighborhood. It is the pixel cell, and the neighborhood radius represents the standard deviation of Gaussian blur. Obviously, this sampling pattern contains pixel cell of different sizes. Therefore, in the process of constructing the comparison pair set, that is, comparing cell of the same size and different sizes, it can give better consideration to both the whole and the local.

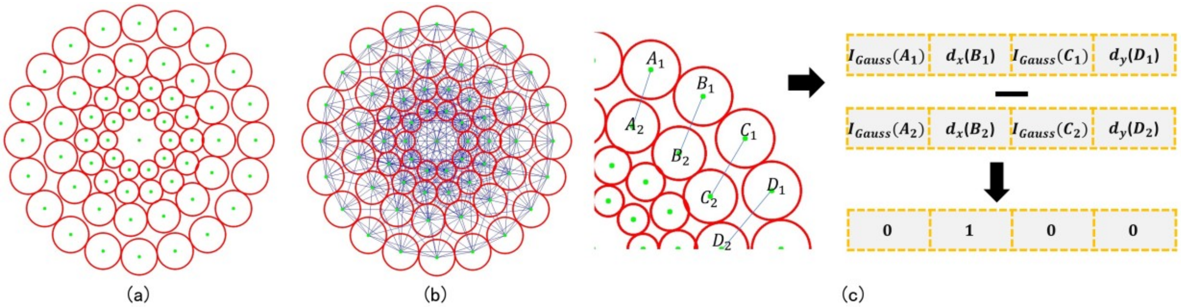

Figure 1. Algorithm description: (a) Sampling pattern when N = 60, green dot represents sampling point, red circle represents pixel cell size; (b) Set of comparison pairs, 512 pairs in total; (c) The four adjacent comparison pairs are grouped into a group, and the gray values (I_{Gauss}) and the gradients (d_x and d_y) in the x and y directions are respectively compared to generate 4-bit descriptors.

Because there are many sampling points in the neighborhood of each keypoint. If the pair is arbitrarily combined, the set $G=\{(p_{t,1},p_{t,2})\}_{t=1...N}$,the descriptors have the problem of too high dimension. It is not conducive to practical application. In addition, not all comparisons are valuable to the generated descriptors, some are ineffective in describing keypoints, and some may even interfere. In order to facilitate calculation, we defines the distance threshold ϑ to filter the comparison pair set S:

$$S=\{(p_{t,1},p_{t,2})\in G\,|\,\|p_{t,1}-p_{t,2}\|<\vartheta\} \tag{4}$$

2.3 Building the Descriptor

The descriptor dimension plays an significant role. On the one hand, the higher dimension contains more information, which is more beneficial to image matching, but will increase the time consumption. On the other hand, although low-dimensional descriptors are faster, they are not conducive to late matching. Ideally, we want descriptors to be both rich in image information and low in dimension. The sampling pattern in this paper has a large number of comparison pairs. If each comparison pair computes three aspects of information (gray value, x - and y-direction gradients), the dimension will be high. As shown in Figure 1b, pixel cell size and comparison relationship are centrally symmetric in space. So the comparison of symmetric positions has the same meaning. Based on this, we uses a grouping comparison method to generate descriptors. Each group contains four comparison pairs, which are compared in the order of gray value, gradient in X direction, gray value and gradient in Y direction to achieve the purpose of dimensionality reduction.

The set $S=\{s_t\}=\{(p_{t,1},p_{t,2})\}_{t=1...T}$ is divided into three subsets, gauss grayscale subset ($G=\{s_t\in S|t=1,3,5...T-1\}$), x-direction gradient subset ($X=\{s_t\in S|t=2,6,10...T\text{-}2\}$), and y-direction gradient subset ($Y=\{s_t\in S|t=4,8,12...T\text{-}4\}$). As shown in Figure 1c, each of the four comparison pairs generates a 4-bit descriptor as a $Tuple(i)=\{E_{i,1},E_{i,2},E_{i,3},E_{i,4}\}=\{I_{Gauss}(G(i)), d_x(X(i)), I_{Gauss}(G(i+1)), d_y(Y(i))\}$. Getting the unit descriptor:

$$\mathrm{b}(Tuple(i),n)=\begin{cases}1,0<E_{i,n}\\0,otherwise\end{cases} \tag{5}$$

3. EXPERIMENTS AND RESULTS

The experiment platform is a 64-bit Windows10 computer, using Oxford optical image dataset[19] for testing. This dataset contains five types of images, and tests the our algorithm from five different aspects, including image blur, light, JPG compression, rotation scale conversion, and viewpoint. Each class contains multiple groups of images. Each group contains six images. The first is the original image, and the other five are the result of gradually strengthening the transformation of the original image according to its category. In the experiment, SIFT, SURF, ORB, BRISK and our algorithm (GDBID) are selected for comparison. The mainstream RANSAC was used for matching. Experiments show that the descriptor is effective.

3.1 Evaluation metrics

(1) Time of generating unit string

The calculation formula is equation (6). N is the total number of keypoints and T is the time required to generate all descriptors.

$$t=\frac{T}{N} \tag{6}$$

(2) Matching precision

The calculation formula is equation (7). $\mathrm{Num}_{\mathrm{correct}}$ matches refers to the number of correct matches while $\mathrm{Num}_{\mathrm{false}}$ matches refers to the number of false matched points.

$$Precision=\frac{Num_{correct}}{Num_{correct}+Num_{false}}\times 100\% \tag{7}$$

3.2 Results

(1) Time consuming

In order to compare the time consuming by different algorithms to generate unit string, 100 images in the dataset were selected for experiment. Firstly, the total number of descriptors generated by each algorithm and the total time were consumed. And then the average value was calculated. As shown in Table 1, the GDBID takes significantly less time to generate unit string than other algorithms. Comprehensive analysis shows that GDBID has good real-time performance.

Table 1. Time consuming by different algorithms to generate unit string.

Algorithm	**SIFT**	**SURF**	**BRISK**	**ORB**	**GDBID**
t/us	534	323	146	129	110

(2) Matching precision

In order to calculate the matching precision of each algorithm in different scenarios, we set each group in the dataset into five matching groups, namely 1/2, 1/3, 1/4, 1/5 and 1/6. First, the matching precision of the five algorithms in each matching group was calculated, and then the mean value was calculated.

As shown in Figure 2, the GDBID performs well in the blur, light, and JPG compression. Compared with the best algorithm, the matching precision difference is less than 1%. It still has high precision in the large difference matching groups (1/5 group and 1/6 group). It is even the best performing algorithm in some matching groups. In the rotation scale conversion and viewpoint, the GDBID is similar to others. Therefore, the matching effect of GDBID is good.

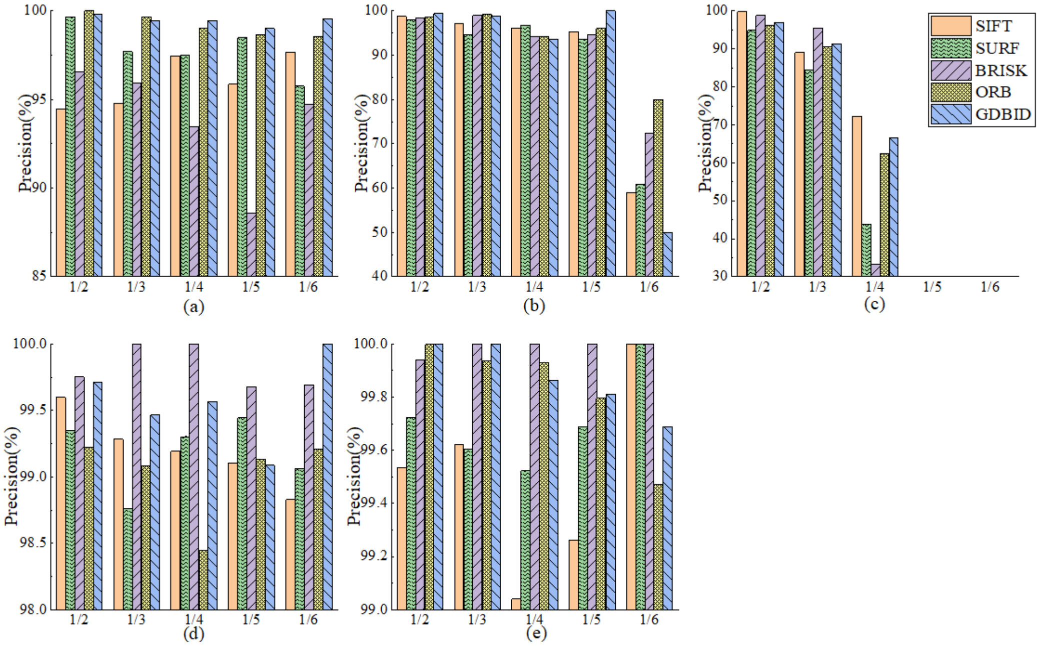

Figure 2. The precision of each matching group under six algorithms: (a) blur, (b) rotation scale conversion (c)viewpoint (d) light, (e) JPG compression.

4. CONCLUSIONS

In order to get more robust descriptors faster, this paper proposes the fusion gradient distinction binary image descriptor (GDBID). Improving the performance of descriptors requires enriching the information contained in strings as much as possible. Firstly, concentric circles were constructed to obtain comparison pairs of different sizes, which makes the descriptor. Then, in order to reduce the string dimension, the distance threshold is defined to preserve valid comparison pairs. Finally, the gray value and gradient are used to enrich the descriptor information. In order to better verify the advancement of GDBID, the time of generating unit string and matching precision is taken as measurement tools in this paper. Experimental results show that the proposed algorithm has higher matching precision and speed.

REFERENCES

[1] Schonberger, J. L., Frahm, J. M., "Structure-from-motion revisited," Proc. Conference on Computer Vision and Pattern Recognition, 4104-4113(2016).

[2] Mur-Artal, R., Montiel, J. M. M. and Tardos, J. D., "ORB-SLAM: A versatile and accurate monocular SLAM system," IEEE Transactions on Robotics, 31, 1147-1163(2015).

[3] Nister, D. and Stewenius, H., "Scalable recognition with a vocabulary tree," Proc. Conference on Computer Vision and Pattern Recognition, 2161-2168(2006).

[4] Wohlhart, P. and Lepetit, V., "Learning descriptors for object recognition and 3D pose estimation," Proc. Conference on Computer Vision and Pattern Recognition, 3109-3118(2015).

[5] Suarez, I., Munoz, E., Buenaposada, J.M. and Baumela, L., "FSG: A statistical approach to line detection via fast segments grouping," Proc. of Int.Conf. on Intell. Robots Systems, 97-102(2018).

[6] Lowe, D.G., "Distinctive image features from scale-invariant keypoints," International Journal of Computer Vision, 60, 91-110(2004).

[7] Balntas, V., Lenc, K., Vedaldi, A. and Mikolajczyk, K., "Hpatches: A benchmark and evaluation of handcrafted and learned local descriptors," Proc. Conference on Computer Vision and Pattern Recognition, 5173-5182(2017).

[8] Calonder, M., Lepetit, V., Strecha, C. and Fua, P., "BRIEF: Binary robust independent elementary features," Proc. European Conference on Computer Vision, Springer, 778-792(2010).

[9] Rublee, E., Rabaud, V., Konolige, K. and Bradski, G., "ORB: An efficient alternative to SIFT or SURF," Proc. IEEE Int'l Conf. Computer Vision (ICCV), (2011).

[10] Leutengger, S., Chli, M. and Siegwart, R.Y., "BRISK: Binary robust invariant scalable keypoints," Proc. IEEE Int'l Conf. Computer Vision (ICCV), (2011).

[11] Alahi, A., Ortiz, R. and Vandergheynst, P., "FREAK: Fast retinal keypoint," Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), (2012).

[12] Yang, X. and Cheng, K. T., "Local difference binary for ultrafast and distinctive feature description," Proc. IEEE Transactions on Pattern Analysis & Machine Intelligence, 36(1), 188-194(2013).

[13] Alcantarilla, P. F., "Fast explicit diffusion for accelerated features in nonlinear scale spaces," Proc. British Machine Vision Conference (BMVC), (2013).

[14] Levi, G. and Hassner T., "LATCH: Learned arrangements of three patch codes," Proc. IEEE Winter Conference on Applications of Computer Vision, (2015).

[15] Fischler, M. A. and Bolles, R. C., "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," Comm. ACM, 24, 381-395(1981).

[16] Chum, O. and Matas, J., "Matching with PROSAC—Progressive sample consensus," Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR'05), vol. 1, pp. 220-226 (2005).

[17] Ma, C., Hu, X., Xiao, J., et al., "Improved ORB algorithm using three-patch method and local gray difference," Sensors (Basel, Switzerland), 20(4), (2020).

[18] Simard, P. Y., Haffner, P. and Lecun, Y., [Boxlets: A Fast Convolution Algorithm for Signal Processing and Neural Networks], MIT Press, (1999).

[19] Mikolajczyk, K. and Schmid, C., "A performance evaluation of local descriptors," Proc. IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(10), 1615-1630(2005).