A LOW DIMENSIONAL INTENSITY-BASED FEATURE DESCRIPTOR FOR FAST IMAGE MATCHING

Zhi Hu[1], Yibo Fan[1], and Xiaoyang Zeng[1]*
[1]State Key Lab of ASIC & System, Fudan University, Shanghai, China
*Corresponding Author's Email: fanyibo@fudan.edu.cn

ABSTRACT

In this paper, we propose a novel image feature descriptor based on intensity information. In comparison with the widely-used SIFT algorithm, it can be implemented easily, computed efficiently, while at the same time demonstrating high performance under various conditions including the transformation of rotation, scale, viewpoint, illumination and JPEG compression. Additionally, this descriptor's low dimensionality leads to lower memory usage.

INTRODUCTION

Image matching is recognized as a crucial step in many computer vision tasks including image registration, object recognition and stereo robot vision. Presently there are mainly two feasible approaches to deal with the task, one is to align images directly based on the difference of intensity information, and the other is to generate the descriptors on the basis of the information around a number of selected keypoints.

The former (direct matching) can often be implemented with simple algorithm (to run a exhaustive search with an error metric function) , however, due to its high computational cost and adaptability to only simple transformation models, it is usually applied to patch matching and fast image stitching among several images with only translational transformation.

The latter (feature-based matching) has gained much popularity in the recent years, especially with the emergence of SIFT [1] algorithm which has shown greater robustness and more accurate matching results under various conditions. The major steps of the algorithms of this type can be briefly described as follows: 1. Search and select keypoints over the whole image. 2. Generate a descriptor for each keypoints on the basis of neighbouring pixels. 3. Match keypoints of two images with their corresponding descriptors. Employing a relatively small number of information-dense descriptors can reduce the calculation as well as enhance the robustness of the algorithm. Moreover, complex transformation model can be estimated with matching result by RANSAC algorithm [2]. Now it has been widely used in panorama generation, object recognition and image retrieval owing to its good performance.

A detailed performance evaluation of feature descriptors conducted by Mikolajczyk et al. [3] suggests that SIFT and GLOH [3] are the most competitive ones. SIFT is a robust and highly discriminant descriptor invariant to scale, rotation and illumination changes. For a give image, it detects the keypoints and for each one generates a descriptor which is a 128-dimensional descriptor. Because it's computationally-intensive, a series of algorithms focusing on the optimization has been put forward. PCA-SIFT [4] proposed by Y. Ke and R. Sukthankar uses the principal component analysis to reduce the dimensionality of the vector in order to accelerate matching process. Fast Approximated SIFT [5] presented by Grabner et al. introduces integral image to speed up the keypoint selection and descriptor calculation process. These two approaches both suffers the loss of accuracy while reducing the execution time. GLOH alters the region for descriptor generation producing a high-dimensional vector and then uses PCA to reduce the dimensionality. According to the evaluation report [3], GLOH performs slightly better than SIFT but has higher computational cost.

It can be concluded from the existing literature that many feature descriptors have some aspects in common. Firstly, they seldom directly use intensity information, for a number of researchers believe that gradient or other information is more robust and contains more valuable information. Secondly, they have high dimensionality (except for the vectors processed by PCA, but extra offline computation is required). In fact, if the intensity information is adopted appropriately, developing a high performance descriptor with low dimensionality is still possible.

This paper presents a robust, discriminative feature descriptor directly based on intensity information which is invariant to scale, rotation, illumination, small viewpoint change and JPEG compression. Meanwhile, its low dimensionality accelerates the computation but it is not a hindrance to good feature matching quality, for experiment results show that the descriptor performs almost as well as SIFT in most cases.

The remainder of the paper is organized as follows. Section 2 briefly introduce the keypoint detection procedure of SIFT, for the descriptor generation is based on those selected keypoints. In Section 3 we detail the descriptor and analyze the choice of parameters. Section 4 will give the experimental results of the descriptor in the respect of quality, execution time and application in image stitching. Section 5 concludes the paper.

AN OVERVIEW OF KEYPOINT EXTRACTION PROCEDURE IN SIFT

Keypoint extraction is comprised of 3 steps, which will be elaborated below.

Extrema Detection in DoG Scale Space

In order to achieve scale invariance, searching for keypoints at different scales is indispensable, and thus the scale space is required. In SIFT it is a Gaussian pyramid consisting of images of incremental scales. Then the difference of each pair of nearby image in the pyramid is computed to convert the Gaussian pyramid into the difference-of-Gaussian pyramid (DoG). This difference is an approximation of Laplacian of Gaussian, and the extrema of Laplacian of Gaussian represent the stable feature points, hence the extrema detection process in DoG, which produces the candidates of the keypoints.

Keypoint Selection and Subpixel Refinement

With the detected extrema, the position of each pixel is recalculated the at sub-pixel level accuracy by Taylor expansion. Pixel will be discarded if the displacement between the sub-pixel extremum and itself is greater than a predetermined threshold. After the process, there are still pixels whose location is susceptible to transformation lying on the edge, and the ratio $Tr^2(H)/Det(H)$ (Here H represents the Hessian matrix of the pixel) is adopted as the criterion to judge whether a point is on the edge and hence needs to be discarded.

Orientation Determination

For each selected keypoint, a dominant orientation should be calculated to achieve rotational invariance. In SIFT, a 36-bin histogram (10°for each bin which represents a certain orientation) is built and accumulated by the gradient orientation (weighted by the product of Gaussian function and the norm of the gradient) of every keypoint. Orientation is computed by the interpolation of the orientation of the highest bin and its neighbouring bins.

THE PROPOSED KEYPOINT FEATURE DESCRIPTOR

Steps of Descriptor Generation

The previous procedure has determined the coordinates of the keypoints along with their scale and orientation, so the next stage is to generate a descriptor representing the unique feature of each keypoint. The generation process chiefly consists of following steps.

1) SET THE REGION FOR DESCRIPTOR GENERATING

For an input keypoint, a region is set around it which is the source of information of neighbouring pixels. It is a 3σ×3σ square (here σ stands for the scale of the current keypoint) with 6×6=36 subregions. To achieve the rotational invariance, the square will be rotated aligning to the main orientation of the keypoint (as shown in Fig. 1).

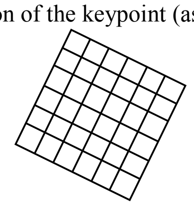


Figure 1: Region for descriptor generating.

2) OBTAIN INTENSITIES FROM THE REGION

The step after set the region is calculating the mean intensity value of each subregion. In the calculation, the intensity value of every pixel within the corresponding subregion is weighted by a Gaussian function. This is because peripheral pixels are more likely to be misregistered, whether they are in or out of the keypoint's region has more uncertainty, so these pixels should be emphasized less. The less the mean intensity values of peripheral subregions are, the less the probability that the differences of these values in 2 descriptors will influence the difference of the 2 entire descriptors while matching a pair of keypoints. This calculation procedure can be expressed as

$$m_i = \frac{\sum I(x,y) \cdot G(x,y,1.5\sigma)}{N_i} \tag{1}$$

where N_i stands for the number of pixels in the i-th subregion.

Furthermore, this step can be accelerated if some assumptions are given. If two images are taken from the same horizontal level, i.e. without rotational transformation, e.g. robot vision, integral image can be used for faster mean value calculation. As for the weighting function, we may apply the Gaussian function to the whole sub-region (i.e. every pixel in the sub-region shares the same weight) rather than every single pixel (i.e. the weight of every pixel in the same sub-region are calculated separately) to get an approximate result. When there is a small-degree in-plane rotation, we may consider the approach presented by [5], i.e. only the central point of each sub-region is rotated and the calculation of its mean intensity is done without rotate the sub-region to a certain orientation. In this way we can apply the efficient integral image method to calculate the mean intensity in O(1) time.

3) GENERATE THE DESCRIPTOR

It is not infeasible to concatenate the 36 m_i values as the descriptor; nevertheless, to maintain stability under illumination changes, some improvement should be made to meet with the demand. There are chiefly two problems ought to be considered.

Firstly, the bias and gain change of the image, that is to say, the intensity of the pixel is changed linearly

$$\widetilde{I} = (1+a)I + b \qquad (2)$$

where a stands for gain and b stands for bias.

This effect can simply be eliminated by 2 steps. 1). Subtract the mean value of m_i ($1 \le i \le 36$) from each m_i , i.e. $m_i' = m_i - \overline{m}$, to form the vector $h = (m_1', m_2', \ldots, m_{36}')$. 2). Normalize the vector h, i.e. $\widetilde{h} = \frac{h}{\|h\|}$.

Secondly, the nonlinear illumination changes. In some cases, the slight change of viewpoint will cause the intensity of some pixels being excessively low or high, which may degrade the robustness. In this paper, this effect is mitigated by thresholding the values in $\widetilde{h}$ to be no larger than 0.30 and no smaller than -0.30. After this, re-normalize the vector forming the final 36-dimension descriptor.

The matching process is conducted by calculating the Euclidean distance of each pair of descriptor vectors as [1] suggests. A pair of keypoints is matched only if the distance ratio of a point to its nearest point and its second nearest one is larger than a matching threshold.

Parameter Determination

1) SIZE OF THE N×N REGION

Based on a test image database, we have conducted the following experiment. Let n=4,5,6,7,8, and calculate the average matching repeatability of several pairs of images with the same matching threshold (we use exhaustive search to do the matching here). The results are shown in Table I.

TABLE I. AVERAGE MATCHING REPEATABILITY WHEN N TAKES DIFFERENT VALUES.

n	4	5	6	7	8
Repeatability	0.4765	0.5391	0.6734	0.6376	0.6555

Therefore, n=6 is suitable for this descriptor.

2) THRESHOLD FOR REDUCING THE NONLINEAR ILLUMINATION EFFECT.

Similar test as the previous one is done at different thresholds and the only difference is that we test precision here because this technique mainly rejects false matches. The results are shown in Table II.

TABLE II. AVERAGE MATCHING PRECISION WHEN THRESHOLD TAKES DIFFERENT VALUES.

Threshold	without	0.26	0.28	0.30	0.32	0.34
Precision	0.849	0.874	0.878	0.884	0.875	0.867

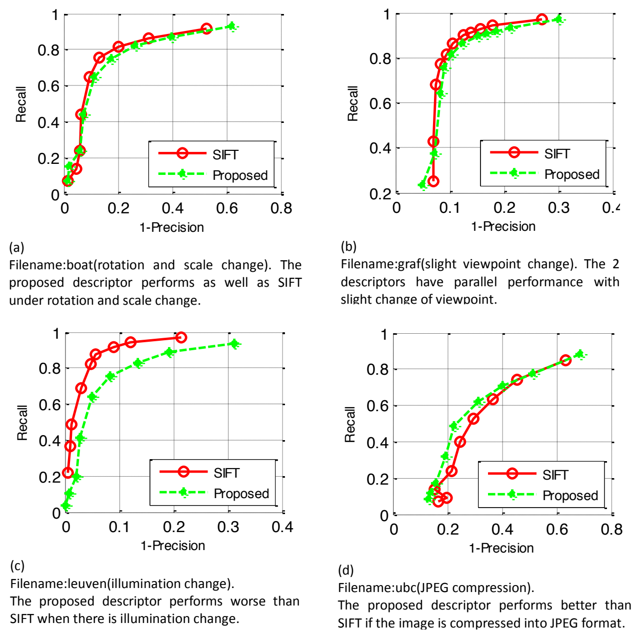
So we choose 0.30 to be the threshold.

EXPERIMENTAL RESULTS

The experiment consists of 3 parts. Part A is the test for quality of the descriptor in comparison with the most stable and widely-used SIFT descriptor. Part B is the test for execution time of the descriptor generation and image matching. Part C is the test of descriptor's application in image stitching. All the tests are performed with the MATLAB implementation of both algorithms.

Test for Matching Quality

In this paper, ROC curve is adopted as the evaluation criterion of the descriptors. Descriptors are tested under various conditions. Fig. 2 shows the test results (exhaustive search is used for feature matching) under rotation and scale change, slight viewpoint change, illumination change and JPEG compression. The test images are obtained from the Internet [6].



(a)
Filename:boat(rotation and scale change). The proposed descriptor performs as well as SIFT under rotation and scale change.

(b)
Filename:graf(slight viewpoint change). The 2 descriptors have parallel performance with slight change of viewpoint.

(c)
Filename:leuven(illumination change).
The proposed descriptor performs worse than SIFT when there is illumination change.

(d)
Filename:ubc(JPEG compression).
The proposed descriptor performs better than SIFT if the image is compressed into JPEG format.

Figure 2: ROC curves under various conditions.

The results suggest that the proposed low dimensional descriptor has good performance as SIFT does in most cases. However it doesn't do very well when there is salient illumination change. A probable explanation is that SIFT calculates the gradient magnitude of all pixels around the keypoint, which is more robust to the local bias change than the mean intensity value of the region.

Test for Execution Time

First of all, we test the average descriptor generation time of SIFT and the proposed descriptor with a same set of images. The execution time has incorporated the step of keypoint detection and both descriptors share the same code during this process. Results are shown in Table III.

TABLE III. THE DESCRIPTOR GENERATION EXECUTION TIME TEST

Filename	Execution Time of SIFT/s	Execution Time of Proposed Descriptor/s
boat (1 image)	14.27	7.24
graf (1 image)	8.11	4.39
leuven (1 image)	8.38	4.54
ubc (1 image)	10.58	5.44
total (1 image)	41.34	21.61

The execution time test results show the proposed algorithm only needs 52% time to generate the descriptors by comparing with SIFT.

Next we test the average matching time (exhaustive search is used) with a same set of descriptors. The results are shown in Table IV.

TABLE IV. THE DESCRIPTOR MATCHING EXECUTION TIME TEST

Filename	Execution Time of Descriptor Matching/s	
	SIFT	Proposed Descriptor
boat (2 images)	27.61	7.07
graf (2 images)	3.32	0.95
leuven (2 images)	2.81	0.69
ubc (2 images)	8.54	2.09
total (2 images)	42.28	10.80

The test shows that the proposed descriptor has a clear advantage in matching process owing to its low dimensionality.

To sum up, the execution time for descriptor generation and the feature matching has been substantially reduced.

Application: Image Stitching

Since matched keypoints can be employed to find the homography between 2 images, we the effect of our descriptor in image stitching in which homography is crucial. Fig. 3 and 4 shows the results of stitching 2 pictures taken from the same scene using the proposed descriptor.

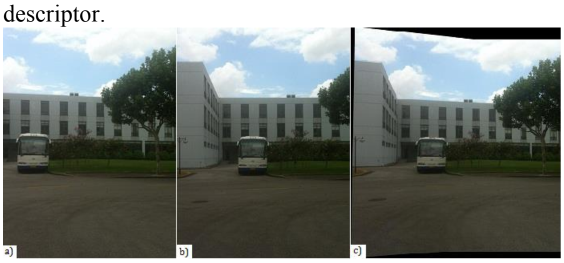


Figure 3: Image stitching test I. (c) is the result of stitching (a) and (b).

Clearly, the proposed descriptor can be a good integral part in image stitching tasks.

CONCLUSION

This paper proposes a novel image feature descriptor directly based on intensity information which has low dimensionality and low computational cost. Its robustness, distinctiveness, speed have been demonstrated in the experiment by comparison with SIFT. Further experiments show that the proposed descriptor can be applied perfectly to image stitching.

Figure 4: Image stitching test II. (c) is the result of stitching (a) and (b).

ACKNOWLEDGEMENT

This paper was supported by National Natural Science Foundation of China (61306023),Specialized Research Fund for the Doctoral Program of Higher Education (SRFDP, 20120071120021), STCSM (13511503400), National High Technology Research and Development Program (863, 2012AA012001)

REFERENCES

[1] D. Lowe, Distinctive image features from scale invariant keypoints, IJCV, 60(2), 91-110,2004.

[2] Fischler, M., Bolles, R.: Random sample consensus: a paradigm for model fitting with application to image analysis and automated cartography, Commun. ACM 24, 381–395, 1981.

[3] K. Mikolajczyk and C. Schmid, A Performance Evaluation of Local Descriptors, IEEE Transactions on Pattern Analysis and Machine Intelligence, 27 (10), pp. 1615-1630, 2005.

[4] Y. Ke and R. Sukthankar, PCA-SIFT: A more distinctive representation for local image descriptors, CVPR (2), pages 506-513, 2004.

[5] M. Grabner, H. Grabner, and H. Bischof, Fast approximated SIFT, Proc, ACCV, pages 918–927, 2006.

[6] http://www.robots.ox.ac.uk/~vgg/research/affine/