A NOVEL LOCAL FEATURE DESCRIPTOR FOR IMAGE MATCHING

Heng Yang, Qing Wang

School of Computer Science and Engineering
Northwestern Polytechnical University
Xi’an 710072, P. R. China

ABSTRACT

Image matching is a fundamental task of many problems in computer vision. This paper presents a novel local feature descriptor based on the gradient distance and orientation histogram (GDOH), which can be used for reliably matching between different views of a scene for wide baseline. The proposed descriptor is invariant to image scale, rotation, illumination and partial viewpoint changes. At present, the SIFT descriptor is generally considered as the most appealing descriptor for practical uses, but the high dimensionality is a drawback of SIFT in the feature matching step. The purpose of GDOH is to reduce the dimensional size of the descriptor, yet still maintain distinctness and robustness as much as SIFT. The experimental results show that the proposed descriptor can result in effectiveness and efficiency in image matching and image retrieval application.

***Index Terms*—**local feature descriptor, invariance, image matching

1. INTRODUCTION

Local invariant features have been widely used in image matching and other computer vision applications, such as image retrieval, recognition of object categories, panoramas building, texture recognition and scene reconstruction, etc [1-7, 17-18]. The features are invariant to image rotation, scale, illumination changes and even affine distortion. Therefore, they are distinctive, robust to partial occlusion, resistant to nearby clutter and noise. Generally, there are two concerns for extracting the local features. The first is to detect keypoints including assigning the localization, scale and dominant orientation for each keypoint. The assigned parameters are used to describe a local image region. Typically, the keypoints are identified by searching the local peaks of the images over scale-space and selected to preserve only those that are likely more stable under transformations. The second is to compute a descriptor for the detected regions, which is our focus in this paper. The descriptor is ideally required to be highly distinctive and as invariant as possible over transformations caused by changes in camera pose and lighting.

There are a number of previous works on identifying representations for local image regions. Schimid and Mohr [8] employed a rotationally invariant descriptor which made the features matched under arbitrary orientation change between the two images. Johnson and Hebert [11] introduced an expressive descriptor spin image which was generated using a histogram of the relative position of neighborhood points to the interesting point in 3D space. These images are invariant to rigid transformations of points. Matthew Brown et al. [5] sampled an 8×8 patch of pixels around the sub-pixel location of the interest point, using a spacing of 5 pixels between samples, and then the descriptor vector was normalized so that the features are made invariant to affine changes in intensity. Finally, the Harr wavelet transform was performed on the 8×8 descriptor patch and the first 3 non-zero wavelet coefficients were used to index the features, which can accelerate the matching process. Van Gool [12] employed the Generalized Color Moments to describe the multi-spectral nature of the data. The moments characterized the shape and the intensities of different color channels in a local region. Florack et al. [13] derived differential invariants to obtain rotation invariance. Schaffalitzky and Zisserman [1] presented the complex filters which are orthonormal, therefore the Euclidean distance can be used to compute the similar score. Lowe [9,10] proposed the SIFT descriptor, which is based on the image gradients in each interest point’s local region. The descriptor is represented by a 3D histogram of gradient locations and orientations and is created by storing the bins in a 128-dimensional vector (8 orientation bins for each of the 4×4 location bins). The contribution to the location and orientation bins is weighted by the gradient magnitude. Finally, the descriptor vector is normalized to unit length to eliminate the effects of linear illumination change. Various refinements based on this scheme have been proposed. Yan Ke and Rahul Sukthankar [14] proposed PCA-SIFT descriptor to reduce the dimensionality of SIFT descriptor. Like SIFT, their descriptors firstly encoded the salient aspects of the image gradient in the feature point’s neighborhood. Then instead of using SIFT’s smoothed weighted histograms, they applied Principal Components Analysis (PCA) on the

978-1-4244-2571-6/08/$25.00 ©2008 IEEE

gradient image and yielded a 20-dimensional descriptor, which resulted in significant space benefit and speed gains. Mikolajczyk and Schmid [15] evaluated several different descriptors and identified the SIFT descriptor as the most resistant to common image deformations. At the same time, they also proposed a new descriptor called GLOH. GLOH is an extension of the SIFT descriptor. It divides local circular region into 17 location bins and the gradient orientations are quantized in 16 bins, which results in 272 bin histogram. The size of the descriptor is reduced with PCA and the 128 largest eigenvectors are used for description. However, GLOH is computationally more expensive and need extra offline computation of patch eigenspace like PCA-SIFT does.

The SIFT descriptor still seems to be the most appealing descriptor for practical applications nowadays, in particular for on-line applications, due to its distinctiveness and relative fast speed. However, the high dimensionality of the descriptor is a drawback of SIFT in the matching phase. To address this issue, this paper presents a novel local scale invariant descriptor called GDOH (gradient distance and orientation histogram) to represent the local image regions. Experimental results show that the dimensional size of our descriptor is much lower than SIFT descriptor, yet still distinctive and robust as SIFT to image rotation, scale, illumination and partial viewpoint changes.

The remainder of this paper is organized as follows. Section 2 details the proposed descriptor GDOH and section 3 briefly describes the feature matching method. The experimental results and related analysis are drawn in section 4. Finally, the conclusion and perspective are summarized in section 5.

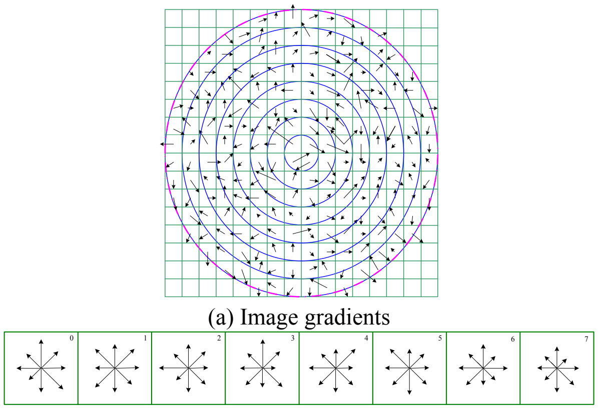
2. LOCAL FEATURE DESCRIPTOR

As described in the introduction, local feature detection and description are generally the two key stages for extracting the local features. This paper uses the same detection scheme as SIFT does, which means we assign the localization, scale and the dominant orientation for each keypoint using the SIFT method. We only focus on the approach to the second stage – the local image descriptor construction.

The task of this stage is to compute a feature vector of the local image structure that will support reliable and efficient matching of features across images over transformations. Our new descriptor is motivated by the SIFT descriptor which is based on the image gradient in each keypoint's local region. The proposed scheme is designed to decrease the dimensional size of the feature vector and maintain the comparable distinctiveness and robustness to the standard SIFT descriptor at the same time.

We named our descriptor as GDOH (gradient distance and orientation histogram) and Figure 1 illustrates the computation of our descriptor. Firstly, the image gradient magnitudes and orientations are sampled around the keypoint location, using the scale of the keypoint to select the level of Gaussian blur for the image [10]. Then, a Gaussian weighting function with σ equal to half the width of the sample region is employed to assign a weight to the magnitude of each point. The purpose of this Gaussian window is to reduce the emphasis on the points that are far from the center of the descriptor (see Figure 1(a)). Next, the gradient orientations are rotated relative to the keypoint dominant direction to achieve rotation invariance and the distance of each gradient point to the descriptor center is calculated too. Thereafter we build the histogram based on the gradient distance and orientation with 8(distance bins)×8(orientation bins) = 64 bins. The descriptor is shown as Figure 1(b). To avoid boundary affects in which the descriptor change sharply as a sample shifts smoothly from one distance to another or from one orientation to another, a bilinear interpolation is employed to distribute the value of each gradient into adjacent histogram bins. The descriptor is formed containing the values of the histogram entries, which is represent by the length of the arrows in Figure 1(b). Our descriptor is a 64-element feature vector for each keypoint and we normalize the vector to unit length to eliminate the effects of illumination change.

Furthermore, we can use different number of the gradient distance bins and orientation bins to form descriptors with different dimensional sizes. Figure 2 shows the performance comparison of GDOH with different dimensionality using recall-(1-precision) graph (see section 4.1). The GDOH48 is created with 6 distance bins and 8 orientation bins (denoted as 6d×8o). Similarly, GDOH64, GDOH96 and GDOH128 are created with 8d×8o, 6d×16o and 8d×16o, respectively. We can find that the best results can be achieved by GDOH64. Therefore, we choose 64-dimension descriptor for each feature point in our subsequent experiments.



(a) Image gradients

(b) Keypoint descriptor

Figure 1. Keypoint descriptor is created by first computing the gradient magnitude, orientation and distance to the center, as shown by (a). The magnitudes are weighted by a Gaussian window, indicated by the red dashed circle. Then these samples are accumulated into distance-orientation histograms, as shown by (b).

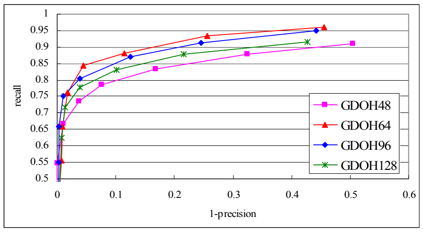


Figure 2 GDOH performance as GDOH dimensionality is varied.

3. FEATURE MATCHING

Given keypoint descriptors extracted from a pair of two images, we first find a set of candidate feature matches using Best-Bin-First (BBF) algorithm[16,10], which is an approximate nearest-neighbor searching method in high-dimensional spaces. We only consider the matches in which the distance ratio of nearest neighbor to the second-nearest neighbor is less than a threshold[10]. This measure performs well and effectively, since correct matches should have the closest neighbor significantly closer than the closest incorrect match.

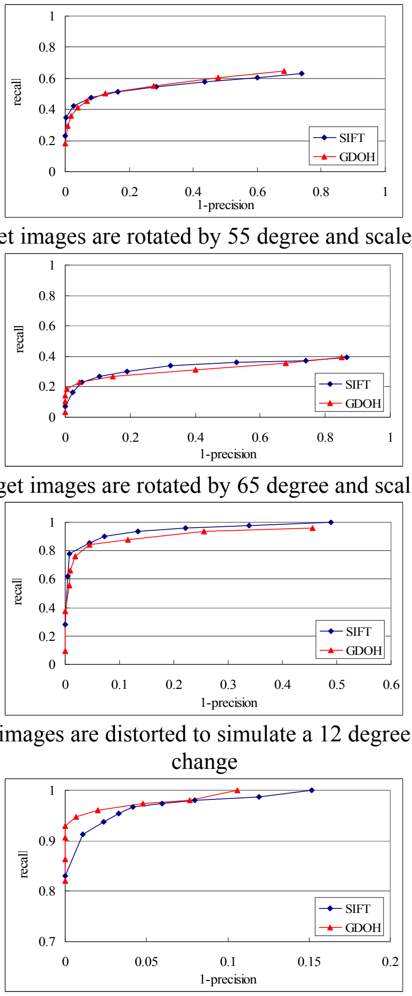
4. EXPERIMENTAL RESULTS AND DISCUSSIONS

We compare the performance of SIFT and GDOH by image matching experiments and an image retrieval application. The dataset for image matching experiments comes from [19], which contains test images of various transformation types. The image retrieval experiment uses a small dataset [20], which includes 30 images of 10 household items. All the experiments are implemented on the PC with Pentium(R)4 2.80GHz CPU and 512M memory.

4.1. Image matching experiments

The descriptor evaluation criterion we used here is recall vs. 1-precision graph [14], which captures the fact that we want to increase the number of correct positives while minimizing the number of false positives. We obtain the curves along with the variation of the ratio threshold (see section 3). Figure 3 presents the results on images with different conditions, where (a) rotation of 55 degree and scale of 1.6; (b) rotation of 65 degree and scale of 4; (c) 12 degree viewpoint change; (d) 20% intensity reduction. From Figure 3(a) and (b), we can see that under the geometric transformation, the performance of GDOH can be a little better than SIFT when the scale factor between images is small, while a little worse than SIFT when the scale factor is relatively larger. Furthermore, under the viewpoint change (Figure 3(c)), GDOH performs a little worse than SIFT, while under the condition of intensity change (Figure 3(d)), GDOH outperforms SIFT slightly. In a word, GDOH can performs comparatively with SIFT over various transformation types of images.

Table 1 lists the comparison result of average matching time of SIFT and GDOH, respectively. The average matching time is recorded in detail according to different dataset group. We can see that GDOH is significantly faster than SIFT in the image matching stage, since GDOH requires about 63% of the time of SIFT to do 65 pairs of image matching.



(a) Target images are rotated by 55 degree and scaled by 1.6

(b) Target images are rotated by 65 degree and scaled by 4

(c) Target images are distorted to simulate a 12 degree viewpoint change

(d) Intensity of target images is reduced 20%

Figure 3 GDOH vs. SIFT on a matching task where the images are under different conditions.

Table 1 Average matching times for SIFT and GDOH

Dataset		Average matching time (ms)	
		SIFT	GDOH
Rotation & Scale	Boat (9 pairs)	4849	2939
	East_park (10 pairs)	2650	1648
	Resid (10 pairs)	1753	1137
Viewpoint	Graff6 (8 pairs)	2363	1451
Illumination	Fruits (7 pairs)	324	216
	Graph (11 pairs)	305	204
Noise	Bike (5 pairs)	1622	1106
	Tree (5 pairs)	6975	4459
Total time cost for 65 pairs		155183	97490

4.2. Image retrieval experiments

We evaluate the performance of SIFT and GDOH in an image retrieval application for real-world scenes taken from different viewpoints. The dataset [20] contains 30 images with 10 groups of different items. Our image retrieval experiment is similar to that conducted by Yan Ke et al.[14]. We first extract the descriptors of each image in the image dataset. Then we find matches between every pair of images. We consider matches if the distance ratio of the nearest neighbor to the second-nearest neighbor is less than a threshold. We regard the number of matched feature vector as a similarity measure between images. For each image, the top 2 images with most matched number are returned. If the returned 2 images are both in the same group of the query image, the algorithm is awarded 2 points. If only one image is in the same group of the query image, it is awarded 1 point. Otherwise, it is given no point. Finally, the scores are divided by 60 which is the total correct match number.

We tune the ratio threshold to obtain the best retrieval results (SIFT ratio threshold: 0.4; GDOH ratio threshold: 0.55), which are shown in Table 2. In addition, the time cost for the whole image retrieval process (29*30=870 pairs of image for matching) is also listed in Table 2. From Table 2, we can see that GDOH's correct retrieval result is even a little better than that of SIFT, which once again demonstrates that the proposed descriptor GDOH is as distinctive and robust as SIFT descriptor. Furthermore, from the time comparison result, it is proved again that GDOH can result in much faster matching (the time cost of GDOH is about 61% of that of SIFT).

Table 2 Performance results of SIFT and GDOH in an image retrieval application (870 pairs of image for matching)

	SIFT	GDOH
Correct retrieval rate	48.3%	55%
Time cost (ms)	593307	364668

5. CONCLUSION

This paper proposes a novel local feature descriptor for image matching. Our new descriptor, named GDOH, is created based on the gradient distance and orientation histogram and can be invariant to image rotation, scale, illumination and partial viewpoint changes. The experimental results show that GDOH is as distinctive and robust as SIFT descriptor. Furthermore, the dimensionality of GDOH is much lower than that of SIFT, which can result in high efficiency in image matching and image retrieval application.

ACKNOWLEDGEMENT

This work is supported by National Hi-Tech Development Programs of China under grant No. 2007AA01Z314.

REFERENCES

[1] F. Schaffalitzky, and A. Zisserman, Multi-view Matching for Unordered Image Sets, or "How Do I Organize My Holiday Snaps?", In: Europe Conference on Computer Vision (ECCV), vol. 1, Copenhagen, Denmark, pp. 414-431, 2002.

[2] V. Ferrari, T. Tuytelaars, and L. Van Gool, Wide-baseline muliple-view correspondences. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), vol. 1, Madison, Wisconsin, USA, pp. 718-725, 2003.

[3] M. Brown and D. G. Lowe, Recognising panoramas. In: International Conference on Computer Vision (ICCV), vol. 3, Nice, France, pp. 1218-1227, 2003.

[4] M. Brown and D.G. Lowe, Unsupervised 3D object recognition and reconstruction in unordered datasets, International Conference on 3-D Digital Imaging and Modeling, Canada, pp. 56-63, 2005.

[5] M. Brown, R. Szeliski, and S. Winder. Multi-image matching using multi-scale oriented patches. In: IEEE Conf. on Computer Vision and Pattern Recognition, Vol. 1, pp. 510-517, 2005.

[6] Noah Snavely, Steven M. Seitz, Richard Szeliski, Photo tourism: Exploring photo collections in 3D. ACM Transactions on Graphics (SIGGRAPH Proceedings), 25(3), pp.835-846, 2006.

[7] S. Lazebnik, C. Schmid, and J. Ponce, Sparse Texture Representation Using Affine-Invariant Neighborhoods, In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 319-324, 2003.

[8] C. Schmid and R. Mohr, Local grayvalue invariants for image retrieval, IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(5), pp. 530-534, 1997.

[9] D.G. Lowe, Object recognition from local scale-invariant features, In: International Conference on Computer Vision (ICCV), Greece, pp. 1150-1157, 1999.

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

[11] A. Johnson and M. Hebert, Object recognition by matching oriented points, In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 684-689, 1997.

[12] L. Van Gool, T. Moons, and D. Ungureanu, Affine photometric invariants for planar intensity patterns. In: Europe Conference on Computer Vision (ECCV), pp. 642-651, 1996.

[13] L. Florack, B. ter Haar Romeny, J. Koenderink, and M. Viergever. General Intensity Transformations and Second Order Invariants, Proc. Seventh Scandinavian Conf. Image Analysis, pp. 338-345, 1991.

[14] Y. Ke and R. Sukthankar, PCA-sift: A more distinctive representation for local image descriptors, In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Washington, DC, USA, pp. 506-513, 2004.

[15] 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.

[16] J. Beis, and D.G. Lowe, Shape indexing using approximate nearest-neighbour search in high-dimensional spaces, In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Puerto Rico, pp. 1000-1006, 1997.

[17] M. Brown and D.G. Lowe. Automatic Panoramic Image Stitching Using Invariant Features, International Journal of Computer Vision, pp. 59-73, 2007.

[18] K. Mikolajczyk, B. Leibe and B. Schiele, Local feature for object class recognition. In: International Conference on Computer Vision (ICCV), Vol.2, pp. 1792-1799, 2005.

[19] http://lear.inrialpes.fr/people/mikolajczyk/

[20] http://www.cs.cmu.edu/~yke/pcasift/