BSIFT: Boosting SIFT Using Principal Component Analysis

Mehran Fotouhi, Student Member, IEEE
Department of Computer Engineering
Sharif University of Technology
Tehran, Iran
fotouhi@ce.sharif.edu

Seyyed Ehsan Mirsadeghi, Student Member, IEEE
Electrical Engineering Department
Amirkabir University of Technology
Tehran, Iran
e.mirsadeghi@aut.ac.ir

Shohreh Kasaei, Senior Member, IEEE
Department of Computer Engineering
Sharif University of Technology
Tehran, Iran
skasaei@sharif.edu

Karim Faez
Electrical Engineering Department
Amirkabir University of Technology
Tehran, Iran
kfaez@aut.ac.ir

***Abstract*—Feature descriptors usually have high dimensionality to efficiently represent key points. Finding matches between large sets of descriptors is a basic step in many applications in computer vision and image processing. When the number of descriptors is large, detection of corresponding points can be extremely time-consuming. The goal of this paper is reducing the computational cost in the matching stage especially for SIFT descriptor. We apply the *principal components analysis* (PCA) on two sets of SIFT features of images and find a coarse matching between points. Then, the *Kullback-Leibler* (KL) divergence similarity score is used to improve the matching accuracy. Experimental results show that our proposed technique can reduce the dimension of SIFT and the related matching cost with approximately the same average precision compared to the conventional approach.**

Keywords— corresponding points; key points; SIFT descriptor; principal components analysis; dimension reduction; KL similarity score.

I. INTRODUCTION

Finding Corresponding points in images has been widely used in many computer vision and image processing applications; such as 3D reconstruction, target tracking, image retrieval, image categorization, human re-identification, human activity recognition, and so forth. The correspondence problem is determining which point in one image corresponds to which point in another. For solving this problem, two stages are needed. At the first stage, one of the key point extraction algorithms is applied on each image to detect and describe key points. As such, the correspondence problem is reduced to computing distances between sets of features. In the second stage, based on one of the similarity measures, corresponding feature vectors are recognized.

In recent years, different local descriptors have been developed [1]. The SIFT descriptor, introduced by Lowe in 1999 [2], is one of the most popular descriptors that have been widely used in many applications. The SIFT algorithm consists of three stages: i) a scale-covariant extractor of interest points based on the *difference-of-Gaussian* (DoG) filter, ii) an invariant and stable representation of interest points based on local weighted histogram of gradients, and iii) the *nearest neighbor* (NN) ratio and the *best bin first* (BBF) algorithm to find correspondences on the extracted 128-dimension feature vectors. Therefore, the SIFT utilizes a complicated algorithm with a low matching efficiency, which does not meet the requirements of real-time applications [3]. Thus, it is reasonable to expect that some optimizations and modifications on each stage may lead to superior performance compared to the conventional technique. Usually, a modification in detection of silent point is needed to present a new keypoint descriptor. Description of interest points is the distinctive stage that can affect the matching strategy and therefore the overall performance of the algorithm.

In this paper, an efficient method for finding the matches between two feature point vectors of the images is proposed. First, the PCA is applied on SIFT descriptors of images to achieve initial matches in lower dimension by applying the nearest neighbor and BBF algorithms. Then, the original feature vectors of candidate points are normalized to compute the KL divergence distance on each corresponding points. Final matches are selected based on applying a threshold on the KL divergence distance. Our key idea is to use redundancies between descriptors to detect candidate matches and then refine them. The proposed method determines more matches and also finds more true corresponding points. Experimental results show that using the KL divergence measure instead of the nearest neighbor distance reduces the imprecision.

The main contributions of this work are as follows: i) a significant reduction in run-time, almost without loss of precision when computing distances between sets of SIFTs and finding matches, ii) applying PCA on two images' SIFT descriptors without needing to use any training data, and iii) finding corresponding points in multi steps. The speed-up is due to two reasons. First, we consider initial matches in lower dimension in the PCA space rather than 128-dimensions feature point vectors. As such, finding matches are extremely fast. Second, the KL divergence measure is calculated just on candidate corresponding points.

The rest of this paper is organized as follow. In Section II some related work are discussed. Proposed method is presented in Section III. The effectiveness and performance of the proposed method are shown in Section IV. Finally, Section V concludes the paper.

II. RELATED WORK

The SIFT algorithm has a good robustness to illumination change, geometric distortion, resolution difference, image scaling, and rotation [4]. The goal of SIFT feature vector improvement is to enhance one or more of these factors: the matching efficiency, robustness, and run-time of matching.

In the related literature, several methods have been proposed to deal with the curse-of-dimensionality and speeding up the SIFT procedure. The PCA-SIFT presented in [5], utilizes only the SIFT extraction stage and then applies its own description stage. For each extracted interest point, the PCA-SIFT applies PCA to the gradient patches extracted around local features. Therefore, a compact feature representation is obtained. Because of PCA projection, the PCA-SIFT requires an offline stage for training and then estimating its covariance matrix. Therefore, the PCA-SIFT is significantly slower than SIFT. Different performances are reported on various test data [4, 5]. The PCA-SIFT is much faster in the matching stage because of its low dimensional space. In [6], PCA-SIFT descriptor and hierarchical quantization technique is used to produce the bag-of-feature for object categorization. Then, PCA-SIFT is performed separately on each object category. The proposed method shows that this technique can reduce the dimension of SIFT up to 80% for object categorization. In [7], the idea in PCA-SIFT is applied on SIFT and SURF algorithms to generate descriptors. But, two strategies are utilized for matching: the *nearest neighbor distance ratio* (NNDR) strategy like SIFT and the nearest neighbor strategy like PCA-SIFT. Different kernels are trained for SURF and SIFT using 40000 feature vectors extracted from sample images.

In [8], the *vector quantization* (VQ) histogram is proposed as an alternate representation for local image descriptor; instead of SIFT's weighted orientation histograms. VQ-SIFT is tested for image retrieval and it is claimed that VQ-based local descriptors are more robust to rotation, projective transforms, and illumination change. And, therefore it is more suitable for image retrieval than the standard SIFT algorithm.

In [9], instead of using the SIFT description stage, kernel projection techniques are applied on orientation gradient patches. The new descriptor is named *kernel projection-based SIFT* (KPB-SIFT) and is more compact (36-dimension) than SIFT descriptor. It does not require the pre-training step needed by PCA-based descriptors. The Walsh-Hadamard kernel is chosen as a basic kernel because of its good performance in discrimination and its computational efficiency. Although PCA-SIFT and KPB-SIFT have reduced the dimensionality of SIFT feature vector, but in both approaches the overhead of changing descriptor reduces the overall performance.

The description vectors are mapped into hamming space in [10] and the hamming metric is used to compare the resulting representations. As such, the size of descriptors is reduced by representing them as short binary strings.

The most similar approach to our proposed method can be found in [3] where P. Cao *et al.* combined the SIFT and the *locality preserving projection* (LPP) as the LPP-SIFT algorithm. LPP is a linear version of manifold learning that can exploit the most discriminative features to reduce the dimensionality of datasets. Based on their approach, the SIFT algorithm is used for extracting and describing interest points. Then, the dimension reduction is accomplished by applying the linear LPP on concatenated feature vectors of two images. The Euclidean distance is used to achieve the coarse matching feature point vectors. To overcome the mismatch problem, the Gaussian neighborhood weighted average which makes use of the gray level information of the pixel is performed on each pair of matched points. Compared to our approach, LPP transform and Gaussian neighborhood weighted average are much slower than PCA transform and KL-divergence similarity measure.

III. PROPOSED METHOD

The main objective of this work is reducing the dimensionality of SIFT feature vectors to improve the matching run-time and the robustness to viewpoint changing. The main steps of the proposed methodology are illustrated in Fig. 1. Each stage is described in the following sections.

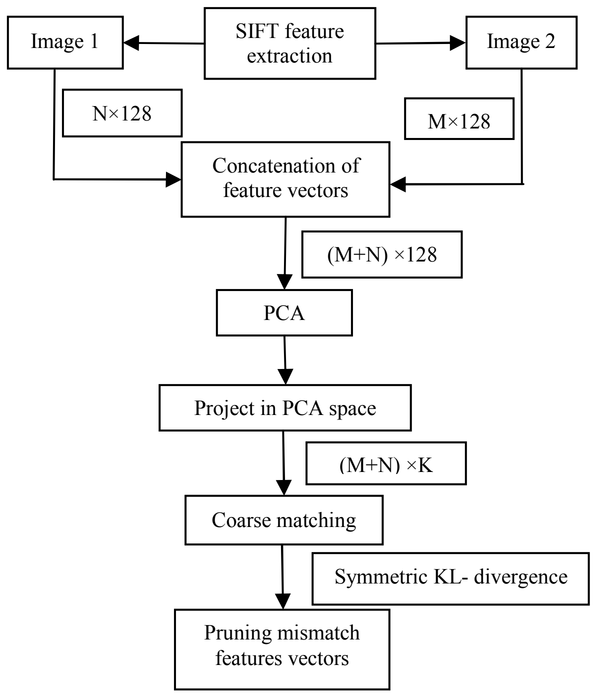

Fig. 1. Overall structure of proposed method.

A. *Interest Point Extraction and Description*

To take advantage of SIFT robustness and capability, the interest point extraction and description stages are similar to the SIFT algorithm [2]. These consist of four stages: i) scale-space peak selection in DOG images, ii) keypoint localization,

iii) identifying dominant orientation for each keypoint based on its local patch, and iv) producing descriptor for each interest point based on image gradients in its local neighborhood.

B. Dimension Reduction and Coarse Matching

The feature set extracted by SIFT is X={X_1, X_2, …, X_N}, where $X_i \in R_{128}$. Our algorithm reduces the feature vectors by projecting the data in low dimensional space R^K, where K<<128. The PCA transform of the feature vectors is used; since SIFT descriptors are highly redundant. PCA is an orthogonal linear transform that uncorrelates data by less (or equal) number of original variables. Principal components are guaranteed to be independent if the data set is jointly normally distributed. But. PCA is sensitive to outliers when dimension is reduced [11]. In fact, PCA is chosen for dimension reduction, because: i) it is one of the fastest linear transforms that is suitable for this purpose, ii) it has a slight robustness against noise in data, and iii) there are several works to make PCA faster and more robust against outliers [12].

Like PCA-SIFT, we assume that SIFT descriptor has a Gaussian distribution [5]. It is reasonable to consider that SIFT can extract enough number of correct corresponding keypoints from each image. To reinforce this assumption and increase the correlation among data, feature vectors of images are concatenated. Therefore, by applying PCA for dimension reduction, representation of data in lower dimensional space is possible; the distance between each pair of corresponding feature point vectors decreases whereas its distance to other feature vectors may increase.

In dimension reduction, original data is mapped to PCA space based on K principal components, where K is less than original data dimension, as

$$Y = W_K^T X . \tag{1}$$

The number of principal components is chosen based on the percentage of their energy compared to the total energy. Therefore, the dimension of feature vectors is reduced up to K, which K can be determined by

$$\frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{128} \lambda_i} > \theta \tag{2}$$

where λ_i are diagonal elements of covariance matrix. Threshold θ is chosen in the way that 80% of PCA energy in the original space is transformed into reduced dimension space. In the first matching step, initial matches are achieved on low dimensional feature vectors by applying the nearest neighbor and BBF algorithms; like SIFT matching stage. In our experiments rougher threshold than SIFT is chosen for nearest neighbor distance because in the dimension reduction process, some information is lost.

C. Pruning Mismatesches

To refine the matched interest points which are detected in the previous step, symmetric KL-divergence distance is used on original feature vectors. KL-divergence measure is defined as

$$D_{KL}(P \| Q) = \sum_i \ln(\frac{P(i)}{Q(i)})P(i) . \tag{3}$$

KL-divergence is not a symmetric measure and thus it is not a true distance metric. Therefore, a symmetric version of KL-divergence is used as a distance measure of two feature point vectors [13]

$$D_{KL}(P,Q) = \frac{D_{KL}(P \| Q) + D_{KL}(Q \| P)}{2} . \tag{4}$$

To select a threshold for KL-divergence, the result of KL-divergence distances is mapped to range 0-1, by

$$S_{KL}(P,Q) = \exp(-\frac{D_{KL}(P,Q)}{2\sigma^2}) . \tag{5}$$

It uses as s similarity score. As suggested in [14], applying similarity scores is much more effective than distance measure. To compute the KL-divergence distance, each pair of matched feature vectors is normalized. Then, based on the KL-divergence similarity score, corresponding points are selected. The effect of this step can be considered as a pruning mismatched points.

IV. Experimental Results

Performance comparison of the proposed method and the original version of D. Lowe's SIFT method according to accuracy and computational cost in the matching stage is presented here. The popular *Mikolajczyk* dataset [15] (explained next) is used for this purpose. SIFT code is used from *VLFeat* open source library [16]. Keypoint extraction parameters are chosen according to Table I.

In our experiments, different thresholds are applied in the matching stage. As mentioned in [7], SIFT works better with the nearest neighbor distance ratio strategy. The threshold for NNDR is set to 1.5 for SIFT and 1 for our proposed method. The threshold of 1.5 is fair for SIFT, because an acceptable result between correct ratio and correct number of matches is achieved by this value. Lower threshold is used in the first stage of our method to increase the chance of finding more correct matches while obtaining a higher correct ratio than SIFT. In the second stage, usage of the KL-divergence similarity score guarantees to have acceptable correct ratio with almost more correct matches than SIFT. Another threshold which is used for pruning mismatched points step is set to 0.075. This value is chosen experimentally.

A. Mikolajzyk Dataset

This dataset contains eight different groups of images in PPM format which are taken in real-world situations. They differ in viewpoint, illumination, blurriness, zoom, and rotation. More information about this dataset is given in Table II. The homographies between image pairs are available as ground truth. The famous sample called Graffiti is shown in the Fig 2.

TABLE I. SELECTED PARAMETERS FOR SIFT KEYPOINT EXTRACTION.

Parameters	***Number of Octaves***	***Variance of Gaussian Window***	***Magnification Factor***	***min L2-norm*** [a]	***Edge Threshold***	***Peak Threshold***	***Levels per Octave***
Value	Max	2	3	Infinity	40	0	3

a. This value related to descriptors before normalization

TABLE II. MIKOLAJZYK DATASET INFORMATION.

Group	***Bikes***	***Trees***	***Graffiti***	***Wall***	***Bark***	***Boat***	***Leuven***	***UBC***
Type of Deformation	Blur	Blur	Viewpoint	Viewpoint	Zoom+Rotation	Zoom+Rotation	Light	JPEG Compression
Image Size	1000×700	1000×700	800×640	1000×700	765×512	800×640	921×614	800×640

In total, 48 images of this dataset are used to evaluate the quality and speed of the proposed method. At first, the effect of changing the percentage of energy (which is used to reduce the dimension of descriptors) is examined.

The 1-Precision or imprecision, shows the ratio between the numbers of incorrect retrieved matches over the total matched keypoints. This defined by

$$1 - Precision = \frac{Incorrect\ Matches\ Retrived}{Total\ Matches}\ . \quad (6)$$

In Fig. 3, the 1-Precision vs. different images is plotted as a performance measure. In Fig. 4, the speedup ratio is presented. Because the imprecision cannot express the total performance of algorithms, the number of correct matches for Graffiti are also presented in Table 3. For the rest samples, the correct matches with fixed PCA energy level (80%) are listed in Tables 4-10. In these tables, the highlighted rows with gray color show the number of correctly matched keypoints.

During our experiments, we observed that for different images in the same group the same energy level of PCA results in reducing the length of feature vector with approximately the same dimension. As such, a same dimension length is used for all images in same groups.

TABLE III. TOTAL AND CORRECT MATCHES FOR GRAFFITI SAMPLE.

Method	**PCA Energy**	**Equivalent Dimension**	**Index of Images**				
			$I_1 \rightarrow I_2$	$I_1 \rightarrow I_3$	$I_1 \rightarrow I_4$	$I_1 \rightarrow I_5$	$I_1 \rightarrow I_6$
BSIFT	*50 %*	*8*	1384	725	380	298	197
			957	305	50	9	0
	60 %	*13*	1473	882	475	366	293
			986	367	74	10	**3**
	70 %	*19*	1533	995	572	441	350
			1005	**421**	94	12	**3**
	80%	*30*	1554	1039	609	468	390
			1010	**441**	**105**	**15**	**3**
	90%	*51*	1568	1050	633	491	414
			1014	**448**	**111**	**18**	2
SIFT	*100%*	*128*	1505	1001	612	501	438
			997	417	98	14	2

TABLE IV. WALL GROUP – DIMENSION OF FEATURE VECTOR IS 36.

Method	**Index of Images**				
	$I_1 \rightarrow I_2$	$I_1 \rightarrow I_3$	$I_1 \rightarrow I_4$	$I_1 \rightarrow I_5$	$I_1 \rightarrow I_6$
BSIFT	1963	1756	1459	1149	857
	1403	**1254**	**671**	**285**	**25**
SIFT	1758	1454	1057	602	277
	1418	1246	647	249	16

TABLE V. BIKES GROUP - DIMENSION OF FEATURE VECTOR IS 30.

Method	**Index of Images**				
	$I_1 \rightarrow I_2$	$I_1 \rightarrow I_3$	$I_1 \rightarrow I_4$	$I_1 \rightarrow I_5$	$I_1 \rightarrow I_6$
BSIFT	2542	2445	1898	1522	1172
	1905	**1739**	**1164**	**822**	418
SIFT	2284	2159	1678	1360	1134
	1882	1716	1154	814	418

TABLE VI. TREES GROUP - DIMENSION OF FEATURE VECTOR IS 35.

Method	**Index of Images**				
	$I_1 \rightarrow I_2$	$I_1 \rightarrow I_3$	$I_1 \rightarrow I_4$	$I_1 \rightarrow I_5$	$I_1 \rightarrow I_6$
BSIFT	2255	2214	2122	1730	1237
	1139	**907**	**542**	**448**	**188**
SIFT	1522	1478	1213	993	662
	1085	866	480	411	165

TABLE VII. BARK GROUP - DIMENSION OF FEATURE VECTOR IS 35.

Method	**Index of Images**				
	$I_1 \rightarrow I_2$	$I_1 \rightarrow I_3$	$I_1 \rightarrow I_4$	$I_1 \rightarrow I_5$	$I_1 \rightarrow I_6$
BSIFT	1175	672	547	480	369
	713	30	52	200	59
SIFT	1158	648	536	427	312
	717	**34**	**54**	**214**	**63**

TABLE VIII. BOAT GROUP - DIMENSION OF FEATURE VECTOR IS 31.

Method	**Index of Images**				
	$I_1 \rightarrow I_2$	$I_1 \rightarrow I_3$	$I_1 \rightarrow I_4$	$I_1 \rightarrow I_5$	$I_1 \rightarrow I_6$
BSIFT	1366	1200	724	576	424
	1051	872	**309**	208	**32**
SIFT	1330	1166	660	544	372
	1040	872	308	**211**	30

TABLE IX. LEUVEN GROUP - DIMENSION OF FEATURE VECTOR IS 27.

Method	Index of Images				
	$I_1 \rightarrow I_2$	$I_1 \rightarrow I_3$	$I_1 \rightarrow I_4$	$I_1 \rightarrow I_5$	$I_1 \rightarrow I_6$
BSIFT	1702	1600	1511	1455	1363
	1339	**1233**	**1105**	**1001**	**888**
SIFT	1584	1520	1405	1357	1234
	1313	1207	1094	992	875

TABLE X. UBC GROUP - DIMENSION OF FEATURE VECTOR IS 34.

Method	Index of Images				
	$I_1 \rightarrow I_2$	$I_1 \rightarrow I_3$	$I_1 \rightarrow I_4$	$I_1 \rightarrow I_5$	$I_1 \rightarrow I_6$
BSIFT	1630	1446	1252	972	595
	1434	**1228**	**967**	**594**	**324**
SIFT	1557	1382	1173	877	615
	1434	1224	953	585	309

Fig. 2. Graffiti samples of Mikolajczyk dataset. Samples have varying viewpoint changes.

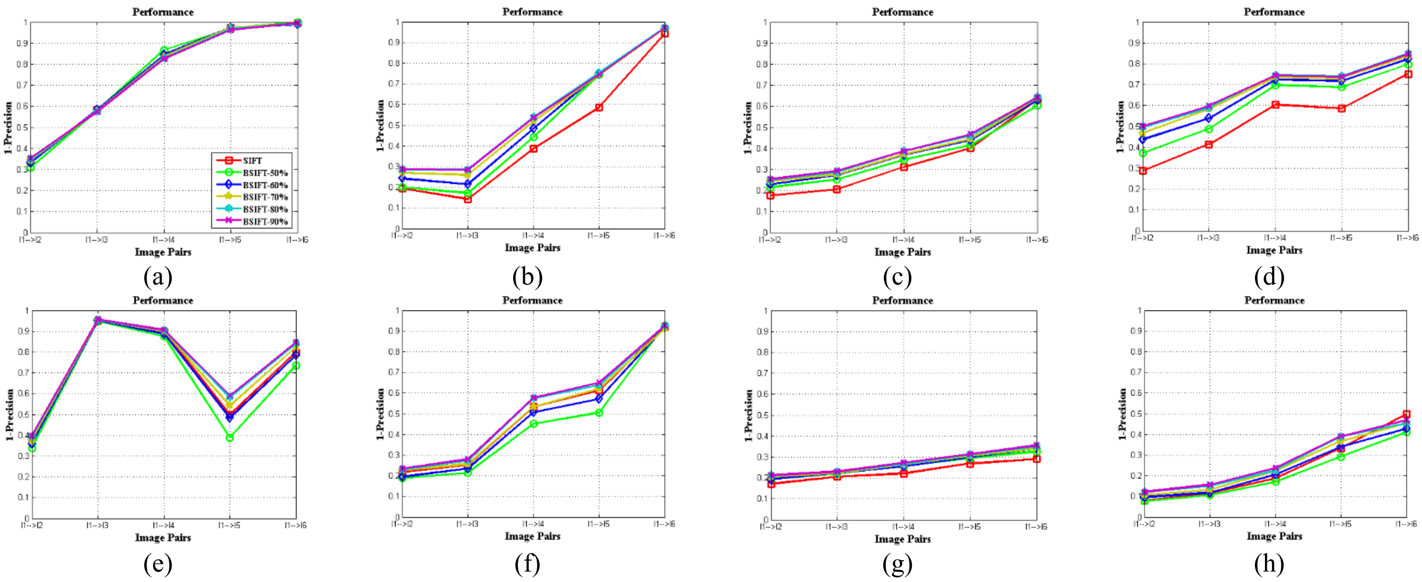

Fig. 3. 1-precision measure of 5 different energy levels of PCA in contrast with SIFT for Mikolajzyk dataset. (a-h) Graffiti, Wall, Bikes, Trees, Bark, Boat, Leuven, and UBC, respectively.

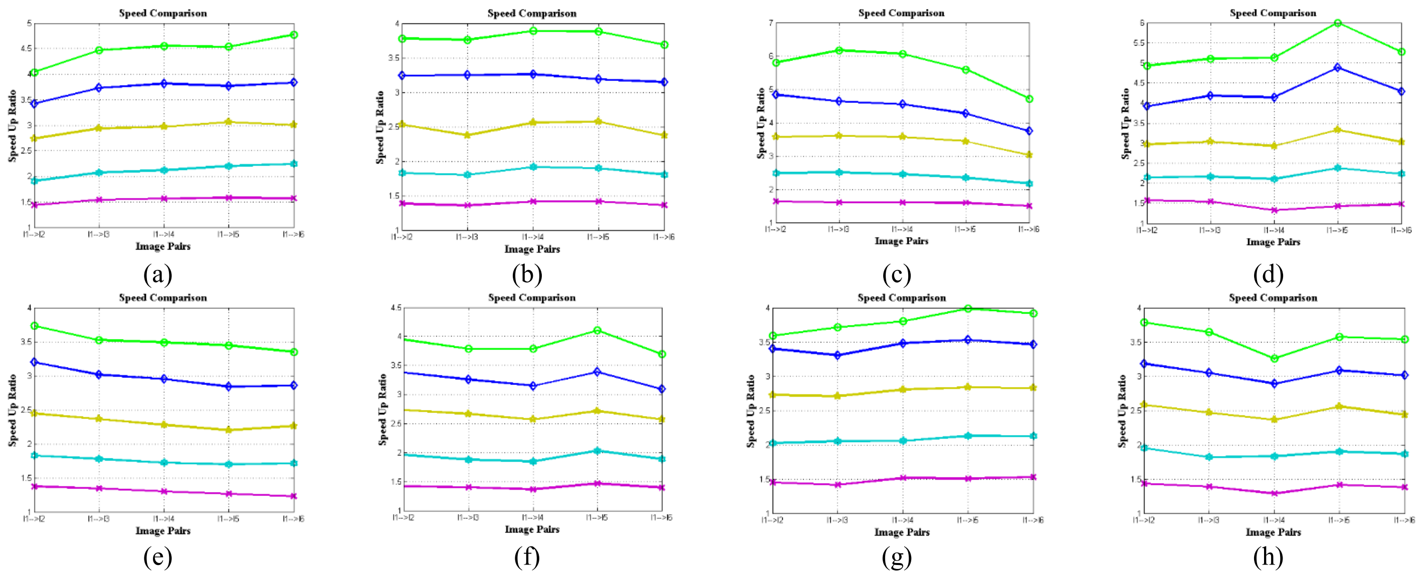

Fig. 4. Speedup ratio of our proposed method for Fig.3.

V. CONCLUSION

A combination of PCA and symmetric KL-divergence for speeding up the matching stage of D. Lowe's SIFT while increasing the number of correct matches between images was presented. The obtained results show that for two related images that have considerable corresponding keypoints, computing the PCA and finding the coarse matching are much faster than available matching algorithms. Because of noise and outliers, the pruning mismatches stage is utilized to improve the accuracy of the coarse matching stage. The proposed method was evaluated on a publicly available dataset which includes 48 real-world distorted images with different deformations; such as varying illumination, zoom, viewpoint, rotation, and blurring level. Comparisons to the state-of-the-art keypoint matching methods showed that the proposed method achieves acceptable matching accuracy in a much faster matching process.

VI. REFERENCE

[1] T. Tuytelaars and K. Mikolajczyk, "Local invariant feature detectors: A survey," in Foundation and Trends in Computer Graphics and Vision, vol. 3, No. 3, pp. 177–280, 2008.

[2] D. G. Lowe, "Object recognition from local scale invariant features," In Proceeding of the Seventh International Conference on Computer Vision, Washington D.C, vol. 2, pp. 1150-1157, September 1999.

[3] P. Cao, R. Ting, Z. Jin-lin and Y. Zhou, "An improved SIFT matching algorithm based on locality preserving projection LPP," In Proceedings of the 4th International Conference on Internet Multimedia Computing and Service, ACM, pp. 192-19, 2012.

[4] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," Pattern Analysis and Machine Intelligence, PAMI, vol. 27, no. 10, pp. 1615–1630, 2005.

[5] Y. Ke and R. Sukthankar, "PCA-SIFT: a more distinctive representation for local image descriptors," In Computer Vision and Pattern Recognition, CVPR 2004, vol. 2, pp. II-506, 2004.

[6] N. Watcharapinchai, S. Aramvith, S. Siddhichai, and S. Marukatat, "Dimensionality reduction of SIFT using PCA for object categorization," International Symposium on IEEE of Intelligent Signal Processing and Communications Systems, ISPACS 2008, pp. 1-4, 2009.

[7] R. Valenzuela, E. González, W. R. Schwartz, and H. Pedrini, "Dimensionality Reduction Through PCA over SIFT and SURF Descriptors," In 11th IEEE Conference on Cybernetic Intelligent Systems, CIS'2012, vol. 1, pp. 58-63, 2012.

[8] Q. Chen, F. Lee, K. Kotani, and T. Ohmi, "Local image descriptor using VQ-SIFT for image retrieval," Journal of world academy of sicense engeenering and technology, waset, vol. 59, pp. 414-419, 2011.

[9] G. Zhao, L. Chen, G. Chen, and J. Yuan, "KPB-SIFT: a compact local feature descriptor," In Proceedings of the international conference on Multimedia, pp. 1175-1178, 2010.

[10] C. Strecha, A. M. Bronstein, M. M. Bronstein and P. Fua, "LDAHash: Improved matching with smaller descriptors," Pattern Analysis and Machine Intelligence, IEEE PAMI Transactions, vol. 34, no. 1, pp. 66-78, 2012.

[11] J. Shlens, "A tutorial on principal component analysis," Systems Neurobiology Laboratory, University of California at San Diego, 2005.

[12] A. E. Abdel-Hakim and M. El-Saban "FRPCA: fast robust principal component analysis," 21st international conference on pattern recognition , ICPR, pp. 413-416, Novamber 2012.

[13] D. H. Johnson and S. Sinanovic, "Symmetrizing the kullback-leibler distance," IEEE Transactions on Information Theory, vol. 1, no. 1, pp. 1-10, 2001.

[14] K. Ma and J.Ben-Arie, "vector array based multi-view face detection ith compound exemplars," In Computer Vision and Pattern Recognition, CVPR 2012.

[15] http://www.robots.ox.ac.uk/~vgg/research/affine/index.html

[16] http://www.vlfeat.org/