A Deformable Local Image Descriptor

Hong Cheng[1 3] Zicheng Liu[2] Nanning Zheng[1] Jie Yang[3]

[1]Xi'an Jiaotong University
Xi'an, China

[2]Microsoft Research
Redmond, USA

[3]Carnegie Mellon University
Pittsburgh, USA

Abstract

This paper presents a novel local image descriptor that is robust to general image deformations. A limitation with traditional image descriptors is that they use a single support region for each interest point. For general image deformations, the amount of deformation for each location varies and is unpredictable such that it is difficult to choose the best scale of the support region. To overcome this difficulty, we propose to use multiple support regions of different sizes surrounding an interest point. A feature vector is computed for each support region, and the concatenation of these feature vectors forms the descriptor for this interest point. Furthermore, we propose a new similarity measure model, Local-to-Global Similarity (LGS) model, for point matching that takes advantage of the multi-size support regions. Each support region acts as a 'weak' classifier and the weights of these classifiers are learned in an unsupervised manner. The proposed approach is evaluated on a number of images with real and synthetic deformations. The experiment results show that our method outperforms existing techniques under different deformations.

1. Introduction

Local image descriptors computed for interest regions have been successfully applied to different areas such as image matching, object detection/recognition, and information retrieval, which makes researchers in computer vision community enthuse over constructing invariant image descriptors [26, 19, 16, 15, 18, 12, 5, 4]. For object recognition and image matching, most existing approaches often look for image descriptors which are invariant to rotation and scaling, such as SIFT (Scale-Invariant Feature Transform) [16] and its variants (PCA-SIFT [12] and SIFT-GC [20]), Geodesic Intensity Histogram (GIH) [15], GLOH (Gradient Location and Orientation Histogram) [19], Spin Images [11], shape context [3], and steerable filter [6], geometric filter [5], etc. Recently, image descriptors learned from training samples have been proposed [14, 10, 26, 2]. Some comparative studies on local image descriptors can be found in [19, 13].

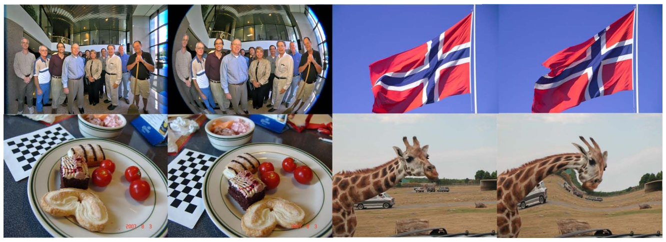


Figure 1. Examples of image pairs with various deformations in our experiments: The left image pair of the first row images are with fisheye lens deformation, the right image pair of the first row images with nonrigid deformation, the left image pair of the second row images with affine deformation, and the right image pair of the second row images with synthetic deformation.

The procedure of using image descriptors for point matching usually consists of three steps. The first step is interest point detection. The second step is to compute feature values of a support region surrounding an interest point. The final step ranks the similarity between a query point and candidate points. While most of the existing work on local image descriptors has been focused on the second step, this research attempts to improve both step 2 and 3 to better handle general deformations, such as fisheye lens deformation caused by fisheye lens distortion, nonrigid deformation, affine deformation, and other synthetic deformations. Figure 1 shows examples of these types of deformations.

The SIFT approach and its variations are most commonly used local image descriptors due to their invariance to scaling and rotations. However, the SIFT feature is not invariant to general deformations [12, 15]. Consequently, it does not work very well for some detection and recognition tasks [27, 21, 15, 23]. There are two factors that affect the performance of SIFT features. First, its performance relies on the optimal scale selection. Sometimes it is difficult to find the appropriate scale for a given interest point thus resulting in a false matching. Therefore, some scale selection strategies are proposed in object recognition [24, 23]. Fig-

978-1-4244-2243-2/08/$25.00 ©2008 IEEE

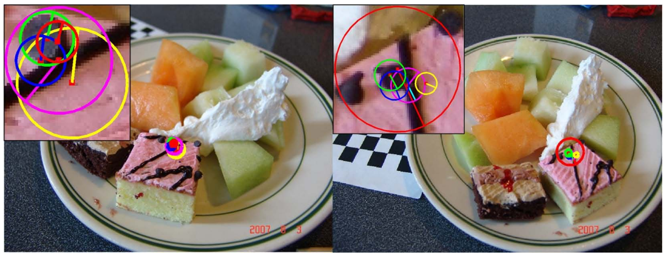


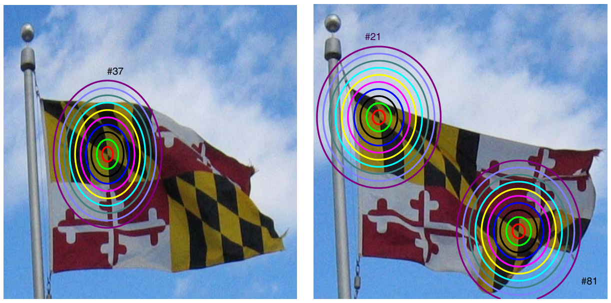
Figure 2. An illustration of the failure of the SIFT approach with the optimal scales of several interest points. The radiuses of support regions represent scales of different DoG points, and the same colors of the left and right sub-figures denote the same interest points.

ure 2 shows an example that the SIFT approach performs poorly because the Difference of Gaussian (DoG) points of the query image have different scales from those of the target image. Second, even if the optimal scale is found, a single support region may not be enough to determine the correct matching. If a support region is too small, the information can be less discriminative thus resulting in false matching, whereas if the region is too large, two support regions surrounding the same point on two images (e.g., before and after deformation) may have very different image statistics due to deformations thus resulting in two feature vectors not close to each other. For example, in Figure 3, the 3 smaller support regions of the first interest point (Figure 3(a)) have almost the same local similarity as those of the second (Figure 3(b)) and third (Figure 3(c)) interest points, but the three points are different points. The same observation is made in [20], of which authors proposed to augment the SIFT with a global context vector that adds curvilinear shape information from a much larger neighborhood to reduce mismatches of similar local descriptors. However, it is difficult to predict changes of a support region and define the global context vector under general deformations.

Many real world images are deformed either physically (e.g., by wide angle lens and fisheye lens) or digitally (e.g., by some image processing tools). To address general deformations, Ling proposed a deformation invariant image descriptor GIH [15]. One drawback is that it assumes the deformation along different directions to be isotropic. This assumption is usually not true in practice. The second drawback is that the deformation invariance comes at the cost of discriminative power. The obtained feature basically loses all the orientation information in the support region.

The performance of feature-point based image matching is critically dependent on similarity measure in the nearest neighbor search [1, 17]. In paper [2], point matching is considered as a classification problem, and a supervised boosting algorithm is used to select features from a feature pool. In [22], the distance measure is learned from examples for specified tasks. In visual category recognition, local distance functions that are globally consistent are learned by a supervised learning approach [7]. Mahamud and Hebert [17] proposed to derive the optimal distance measure by minimizing the nearest neighbor misclassification risk. However, most of the existing algorithms use supervised distance measure, which, unfortunately, requires expensive labelled training samples.

This paper presents a novel local image descriptor that is robust against general image deformations. We make two contributions to the existing local image descriptors in enhancing their ability of handling general deformations of images. First, we propose to use multiple support regions of different sizes surrounding an interest point. A feature vector is computed for each support region, and the concatenation of these feature vectors forms the descriptor for this interest point. Second, we propose a new similarity measure model, Local-to-Global Similarity (LGS) model, for point matching with the new descriptor that takes advantage of the multi-size support regions. Each support region acts as a 'weak' classifier. These 'weak' classifiers are then combined with an unsupervised boosting strategy.



(a) (b)

Figure 3. An example of point matching with a single support region: (a) A query point #37; (b) The first and second best matched points (#21 and #81) with #37, when the blue circle is used as the support region.

2. An Overview of the Proposed Approach

Instead of using a single support region as existing image descriptors, we use multiple support regions of different sizes for any given interest points. One advantage is that we bypass the problem of choosing the optimal support region size which is particularly difficult when there exist general deformations, and it is easy to see this point from Figure 2 where the corresponding interest points on two images have different support regions due to incorrect optimal scales. The second advantage is that features computed from a single support region may not contain enough information to determine the correct matching. For example, the query point 37 is shown in Figure 3(a). When we used the blue circle as the support region, the top two best matches on the right target image are the point 21 and 81

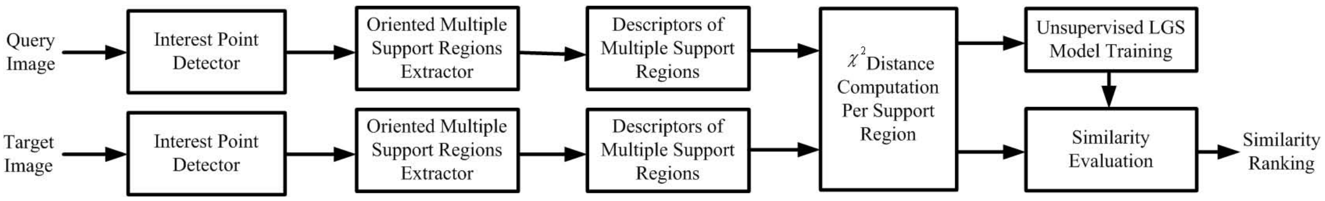

Query Image
Interest Point Detector
Oriented Multiple Support Regions Extractor
Descriptors of Multiple Support Regions
Target Image
Interest Point Detector
Oriented Multiple Support Regions Extractor
Descriptors of Multiple Support Regions
χ^2 Distance Computation Per Support Region
Unsupervised LGS Model Training
Similarity Evaluation
Similarity Ranking

Figure 4. The framework of our deformable local image descriptor.

in Figure 3(b). Both are incorrect. In fact, we tried each of the 10 support regions as shown in the figure, none of them gave the correct matching. In comparison, our LGS model was able to find the correct matching because it effectively combines multiple support regions of different sizes.

Figure 4 is an overview of our interest point matching framework with the proposed image descriptor. Given a query image and a target image, we first detect interest points on two images. Any existing interest point detectors can be used for this step. For each interest point on two images, we extract multiple support regions of different sizes. A feature vector is then computed for each support region. For each pair of interest points and each pair of corresponding support regions which surround the two interest points respectively, we compute a χ^2 distance. For each interest point in the query image, we train an LGS model in an unsupervised fashion. The LGS model is then used to evaluate the similarities between the interest point in the query image and all the interest points in the target image.

3. Multiple-Size Support Regions

We use a circular window to extract a support region given a scale σ. Let (x_0, y_0) denote an interest point in an image I, and $s(x_0, y_0)$ a support region surrounding (x_0, y_0). For each pixel (x, y) in the support region, we compute its gradient magnitude and orientation. Then we divide a support region s into L_1 subregions. For each subregion, we compute the histogram of the gradient directions where the number of orientation bins is denoted as L_2. Thus we obtain a feature vector of dimension $L = L_1 \times L_2$.

Our descriptor for a given support region is similar to the SIFT descriptor with two exceptions. First, the SIFT approach use a Gaussian smoothed image to compute gradient magnitude and orientation with the assumption that an optimal scale is already obtained, while we directly compute gradient magnitude and orientation in the original image. Secondly, we use Harris matrix [9] to compute the principal orientation of the support region instead of gradient histogram. The reason why we use regional Harris matrix instead of the point Harris matrix is because the average orientation of pixels within a support region is more stable than pixel principal orientation.

Given a support region s, its Harris matrix is given by

$$\mathbf{H}_s = \begin{bmatrix} \sum_w D_x^2 & \sum_w D_x D_y \\ \sum_w D_x D_y & \sum_w D_y^2 \end{bmatrix}, \quad (1)$$

where w ranges over the pixels in s, and D_x and D_y are obtained by the convolution between I_x, I_y and a Gaussian function G_{σ_0} with variance σ_0.

We take the eigenvector V_1 that corresponds to the largest eigenvalue, λ_1, of the Harris matrix $\mathbf{H}_s$ as the principal orientation

$$\theta(s) = \text{atan}(V_1(1)/V_1(2)). \quad (2)$$

Given an interest point, we extract multiple support regions of different sizes and compute the oriented gradient histogram for each support region. The sizes of the multiple support regions are given by

$$\sigma(s) = s \cdot \sigma_0, \ s = 0, \cdots, 2N. \quad (3)$$

where σ_0 is a base level of sizes. The nested support regions centered at an interest point are denoted as

$$\mathbb{S} = \{s\}, \ s = 0, \cdots, 2N. \quad (4)$$

For each support region, we compute a feature vector as described in the beginning of this section. Therefore, for any given interest points, we obtain a set of feature vectors corresponding to its multiple support regions. The feature vectors in smaller support regions contain more local information around the interest point while those in the larger support regions contain more global information.

4. Point Matching Using an LGS Model

4.1. Support Region Alignment

Let q denote a query point, let p_1 ,..., p_M denote the set of candidate points. Let h_0, ..., h_{2N} denote the histograms computed at the multi-size support regions centered at q. For each candidate point p_c, $c = 1, ..., M$, let $g_{c,0}$, ..., $g_{c,2N}$ denote the histograms computed at the multi-size support regions centered at p_c. When there is a scaling transformation between the query image and the target image, a support region h_s, $s = 0, ..., 2N$, of the query point may not correspond to a support region of the same size $g_{c,s}$. Instead, h_s may correspond to $g_{c,s-k^*}$, where k^* is constant

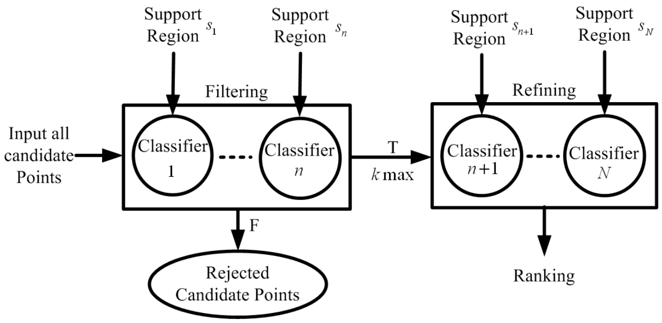


Figure 5. An illustration of the Local-to-Global Similarity model.

(could be negative) for all s which depends on the scaling factor. Since the scaling factor is not known a prior, we estimate the shift k^* as follows.

Given any shift k, where $-N \leq k \leq N$, we define the alignment error between q and p_c to be

$$E(q, p_c, k) = \sum_{s=max(0,k)}^{max(0,k)+N} d(h_s, g_{c,s-k}), \tag{5}$$

where

$$d(h_s, g_{c,s-k}) = \frac{1}{2} \sum_{l=1}^{L} \frac{[h_s(l) - g_{c,s-k}(l)]^2}{h_s(l) + g_{c,s-k}(l)} \tag{6}$$

is the χ^2 distance between the two histograms.

We choose k^* to minimize the alignment errors between q and all the candidate points p_c. That is,

$$k^* = \mathop{\textbf{arg min}}_{-N \leq k \leq N} \left\{ \min_{1 \leq c \leq M} E(q, p_c, k) \right\}. \tag{7}$$

Denote $k_0 = max(0, k^*)$. For subsequent point matching, we choose the histograms of the N support regions for query point q: h_{k_0}, ..., h_{k_0+N-1}. For each candidate point p_c, its histograms of the corresponding support regions are $g_{c,k_0-k^*}, \cdots, g_{c,k_0+N-1-k^*}$. To simplify descriptions, we will omit the shift k^* in the rest of the paper, and assume there are N support regions for a query point q and also N support regions for each candidate point p_c.

4.2. A Local-to-Global Similarity Model

Let h_1, ..., h_N denote the histograms computed at the N multi-size support regions centered at a query point q. Let $g_{c,1}$, ..., $g_{c,N}$ denote the histograms computed at the multi-size support regions centered at candidate points p_c, $c = 1, ..., M$. To take advantage of multi-size support regions, we have developed an LGS model to find the matching point of q among candidate points p_c, $c = 1, ..., M$. As shown in Figure 5, an LGS model contains a cascaded list of classifiers where each support region acts as a 'weak' classifier. The classifiers are divided into two modules: a filtering module and a refining module. The filtering module rejects those candidate points which are unlikely to match the query point. The refining module refines the ranking of the remaining candidate points.

The classifier at each stage is selected based on its proximity structure which measures the similarity of the classifier to the rest of the classifiers. Intuitively speaking, the first classifier has the best overall similarity to all the classifiers so that it is relatively safe to reject those candidate points whose histograms in the corresponding support region are not close to that of the query point. Similarly, the second classifier has the best overall similarity to the rest of the classifiers, etc. The proximity structure is learned by an unsupervised learning strategy as described in the next section.

Like the Adaboost algorithm in object detection [25], based on the 'weak' classifier assumption of multiple support regions, an LGS model can boost point matching performance by combining multi-size support regions as long as the performance of each support region is slightly better than random.

4.3. Proximity Structure Learning

For each support region, $s = 1, ..., N$, and candidate point p_c, $c = 1, ..., M$, let $d_s(q, p_c)$ denote the χ^2 distance between h_s and $g_{c,s}$, that is

$$d_s(q, p_c) = \frac{1}{2} \sum_{l=1}^{L} \frac{[h_s(l) - g_{c,s}(l)]^2}{h_s(l) + g_{c,s}(l)}. \tag{8}$$

For any two candidate points p_{c_1} and p_{c_2}, we use $P_s(q, p_{c_1}, p_{c_2})$ to denote the proximity order of the two candidate points relative to q:

$$P_s(q, p_{c_1}, p_{c_2}) = \begin{cases} 1, & d_s(q, p_{c_1}) < d_s(q, p_{c_2}) \\ 0, & d_s(q, p_{c_1}) = d_s(q, p_{c_2}) \\ -1, & d_s(q, p_{c_1}) > d_s(q, p_{c_2}) \end{cases}, \tag{9}$$

where $s = 1, \cdots, N$.

Now we define a proximity matrix between two different support regions s and l as

$$\mathbf{D}_{sl}(q)_{c_1 \times c_2} = 1 - |P_s(q, p_{c_1}, p_{c_2}) - P_l(q, p_{c_1}, p_{c_2})|/2, \tag{10}$$

where $s, l \in \{1, \cdots, N\}$.

Given a query point q, we compute its proximity property of the support region s with respect to the rest of support regions by

$$F_s(q) = \sum_{l=1, l \neq s}^{N} \|\mathbf{D}_{sl}(q)\|_{\mathrm{F}}, \tag{11}$$

where $\| \cdot \|_{\mathrm{F}}$ denotes the Frobenius matrix norm.

By abuse of notation, we still use $s_1, \cdots, s_N$ to denote support regions obtained from proximity structure learning.

From Equ. (11), we can use the following rule to obtain the classifier s_1 in the cascaded structure by

$$s_1(q) = \mathop{\textbf{arg max}}_{s\in\{1,\cdots,N\}} F_s(q). \tag{12}$$

Similarly, we can obtain the subsequent classifiers by

$$s_k(q) = \mathop{\textbf{arg max}}_{s\in\{1,\cdots,N\}, s\neq s_1,\ldots,s_{k-1}} F_s(q). \tag{13}$$

After obtaining the cascaded list of support regions, we use the first support region s_1 to reject μM candidate points based on $d_{s_1}(q, p_c)$, where μ is an user-specified parameter within the range $(0, 1)$. Subsequently, we use the second support region s_2 to reject $\mu(1-\mu)M$ points, etc. The number of candidates that remain after going through the filtering module is then $k_{max} = (1-\mu)^n M$.

After the filtering step, the refining module determines the ranking of the k_{max} candidate points by making use of the rest of the support regions: $\{s_{n+1}, \cdots, s_N\}$. The ranking score of candidate point p_c is defined as

$$r_q(p_c) = -\sum_{s=1}^{N} \alpha_s d_s(q, p_c), \tag{14}$$

where α_s, $s = 1, \cdots, N$, is the weight of support region s and is proportional to its proximity property obtained from Equ. (11):

$$\alpha_s = F_s / \sum_{i=1}^{N} F_i, \; s = 1, \cdots, N. \tag{15}$$

Note that our distance similarity ranking algorithm is different from BoostMap algorithm [1] in that BoostMap uses a supervised learner to boost 1D embeddings while our method uses an online unsupervised learner which does not require labelled samples.

5. Experimental Results and Analysis

In this section, we present six sets of experiments. Section 5.2 validates the effectiveness of boosting matching performance. In Section 5.3, we investigate the influence of the number of support regions on the matching performance. Section 5.4 compares performances of the LGS and GIH approaches on rotation invariance. In section 5.5, we evaluate the performance of three approaches, LGS, SIFT, and GIH, on fisheye lens deformation images. Section 5.6 validates the robustness of the LGS approach to affine deformation. Finally, we evaluate point matching performance on synthetic images in section 5.7.

5.1. Experiment Setup

Data set: We collected four categories of image pairs (examples shown in Figure 1): fisheye lens deformation, nonrigid deformation, affine deformation, and synthetic deformation, and evaluated the performance of our approach using both our image data set and the data set from [15]. Images with fisheye lens deformation are captured by a camera with a fisheye lens. For each image, we perform fisheye lens deformation correction using Fisheye-Hemi [1]. Nonrigid deformation images are produced by moving nonrigid objects, for example, flags, clothes, etc. The image pairs with an affine transformation are produced by taking pictures of the same scene at the different viewpoints. Synthetic deformation images are produced by applying pre-defined image warping to the original images that are obtained from the Caltech-256 object category data set [8].

Interest Points: Our approach does not require a specific interest point detector. In our experiments, we chose to use two categories of interest points, Harris interest points and DoG interest points.

Evaluation Criterion: Both Receiver Operating Characteristic (ROC) and Recall-Precision are popular in performance evaluation criterions of classifiers and detectors; however, ROC is better for evaluating classifiers while Recall-Precision is better for detectors [12]. Therefore, similar to [12, 19, 15], we use recall-precision to evaluate the performance of interest point matching. The recall-precision is defined as

$$recall = \frac{\#\ correct\ matches}{\#\,possible\ candidate\ matches}. \tag{16}$$

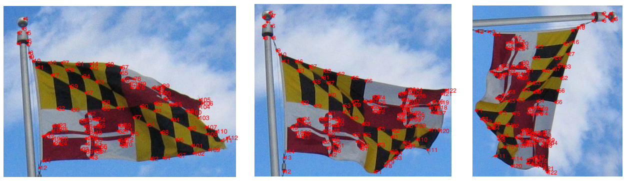
In the following experiments, we choose $n = N/2$, $kmax = 20$, and the angle resolution of 10°. For comparison purpose, similar to the experiments in [15], we remove the points of the query image which do not have matches.

5.2. LGS Model vs. Single Support Regions

This experiment is to show the effectiveness of boosting matching performance. We compare our LGS model to point matching using only one single support region of different sizes. For each interest point of the query image Figure 6(a) and the target image Figure 6(b), we obtain N support regions of different sizes, and thus each support region has a feature vector. We then perform point matches only using one support region s each time, and obtain detection rate and accumulated detection rate in the same way. In Figure 7, the left sub-figure shows the correct matching rate of each support region and the LGS approach that combines all of the multi-size support regions when the rank $R = 1$, and the right sub-figure shows accumulated detection rate which is the detection rate among the top R matches as R varies on the nonrigid image pair. Here, $s1, \cdots, s10$ denote the corresponding accumulated detection rates of the support regions $s = 1, \cdots, 10$. From the Figure 7, we can

[1] http://imagetrendsinc.com/gallery/gallery_hemi.asp

see that the correct detection rate of the LGS model outperforms the best detection rate of each single support region about 5%, and the LGS model is significantly better than most of the single support regions. That is, we can't obtain the better performance than the LGS approach even if we are lucky to find the best size of a support region given an interest point. The horizontal axis indicates ranks of the distance measure.



(a) A Query Image

(b) Target Image 1 without Rotation

(c) Target Image 2 with Rotation

Figure 6. The Nonrigid Image Pairs and its rotated images from the Data Set from [15]. The interest points are generated by Harris corner detectors.

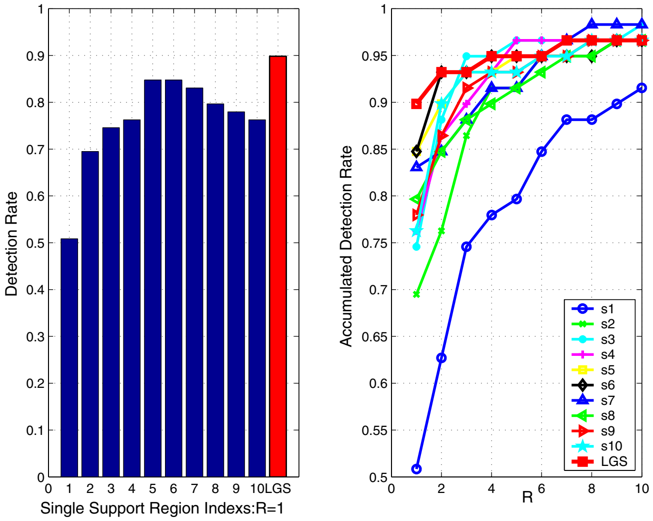


Figure 7. A performance comparison between each single support region and the LGS approach.

5.3. Performances vs. Support Regions

In this section, we study the LGS approach performance for different number of support regions. We should be able to have more than 10 different sizes of support regions but it would require more computational power. We use $N = 2, 4, 6, 8, 10$ for the flag images in the data set from [15] to evaluate the performance of the LGS approach. Figure 8 shows the performance of point matching varies with the number of support regions. The accumulated detection rate of different number of support regions, 2,4,6,8, and 10, are shown in Figure 8. It is obvious that the detection rate improves as support regions increase, and eventually 90% query points find the correct matching points from the first most similarity candidate point when all 10 support regions are utilized.

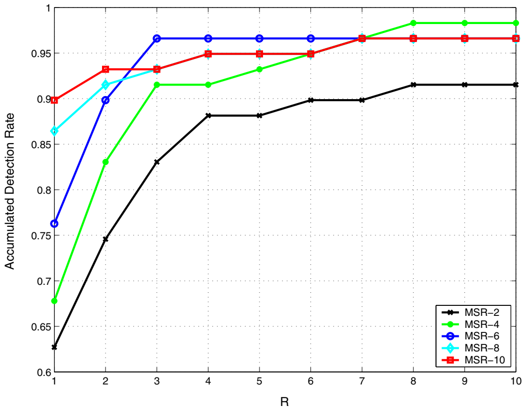


Figure 8. The influence of the number of support regions on the detection rate.

5.4. Rotation Invariance

In this experiment we evaluate rotation invariance property of the LGS and the GIH approaches. We implemented the GIH approach using the online code provided by [15]. The parameters are set to: $\alpha = 0.98$, $K = 13$, $M = 8$. Figure 6(a) is a query image and Figure 6(b) and Figure 6(c) are used as target images. Note that Figure 6(c) is obtained just by rotating Figure 6(b).

In the experiment, we take the flag image pairs from the date set [15], a query image (Figure 6(a)) and a target image 6(b), and rotate the target image with 90^0 as the second target image(Figure 6(c)). For these three images, we use the proposed LGS and GIH approaches to match Harris interest points between the query image and target images. Consequently, four recall-precision curves are obtained, shown in Figre 9. In the figure, 'LGS' and 'LGS-R' curves indicate performances of the LGS approach on the point matching of the original query image to the target and rotated target images, respectively, while 'GIH' and 'GIH-R' curves represent performances of the GIH approach in the same way. The Figure 9 shows a comparison of detection rate and accumulated detection rate between the LGS and GIH approaches on two different target images. The left figure illustrates the detection rate of the top rank and the right one shows accumulated detection rate over the top ten ranks. We can see that (1) the LGS approach detects about 90% points as the first choice; (2) the LGS approach outperforms the GIH approach on both query images; and (3) image rotation has little effect on the performance of the LGS approach, while the GIH approach has a significant performance drop on the rotated image.

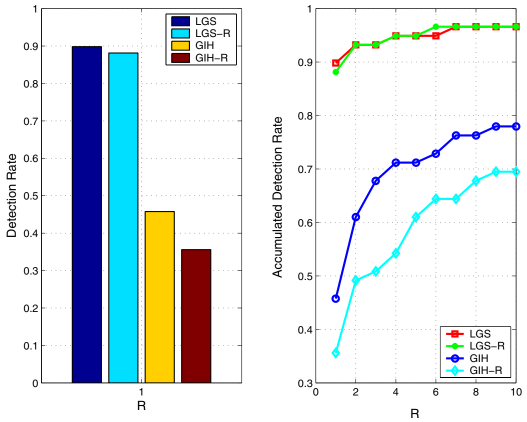


Figure 9. A performance comparison between the LGS and GIH on the image pairs before and after rotation shown in Figure 6.

5.5. Fisheye Deformation

Fisheye lens deformation is a type of unique and common deformation in practice, where a wide-angle lens causes distorted appearance of regular objects. We downloaded 6 pairs of images with fisheye lens deformations from Internet to evaluate our approach. We studied performance of three different approaches, SIFT, GIH, and LGS, on fisheye deformation. In this and the following experiments, we choose the SIFT approach for comparison because it is widely used, and it is an excellent representative of generic local image descriptors. In this experiment, similar to the previous rotation invariant experiment, we first detect Harris interest points, and then implement point matching with the three approaches.

Note that in this experiment, we fixed the scale of the SIFT descriptor for both the query image and the target image to a given scale because it is hard to obtain correct scales which are consistent with the two images for a customized interest point (indicated in Figure 2). We used $s = 4$ in Equ. (3), which gives the SIFT approach the best point matching performance among single support regions. A comparison of the accumulated detection rates for three approaches is shown in Figure 10. The horizontal axis indicates ranks of the distance measure. From the figure, we can see that the LGS approach outperforms both the SIFT approach and the GIH approach.

5.6. Affine Deformation

In the real world, taking pictures of the same scene at different viewpoints often causes affine deformation, which makes most of existing local descriptors fail because this kind of deformation brings occlusion, variation of intensity layout, etc. The proposed LGS approach is robust to affine deformation because it uses multi-size support regions to compute distance measure between a query point and candidate points.

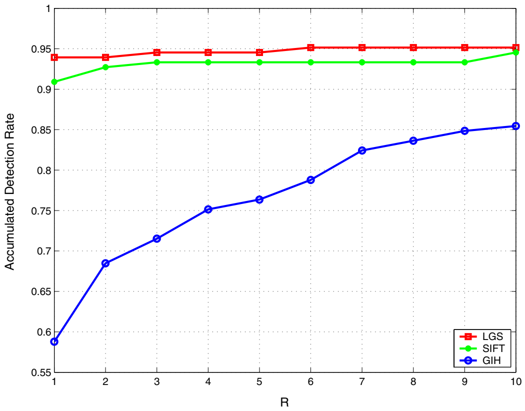


Figure 10. A comparison among three approaches, LGS, SIFT, and GIH, on fisheye deformation images.

We first tried the full SIFT approach, but its outputs only less 1% interest point matching. Furthermore, only a half of the reported point matching are correct. Therefore, we choose the SIFT descriptor with user-supplied scales in Figure 11. In this experiment, we randomly select DoG point of the SIFT approach as interest points. The test image pairs with affine deformation and some occlusion are really hard to point matching, and even so our LGS approach still remarkably outperforms the SIFT approach. Figure 11 shows a comparison between the SIFT and LGS approaches on test images with affine deformations.

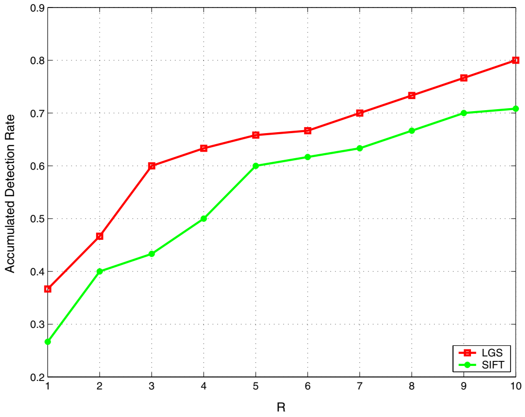


Figure 11. A comparison between two approaches, LGS and SIFT, on our data set with affine deformation.

5.7. Synthetic Deformation

In this section, we evaluate the three approaches, LGS, SIFT, and GIH, on images with synthetic deformations using our general deformation database as shown in Figure 1. Ten image pairs are used to evaluate our approach, and most of which are produced by applying pre-defined image warping function to the original images from [8]. Figure 12

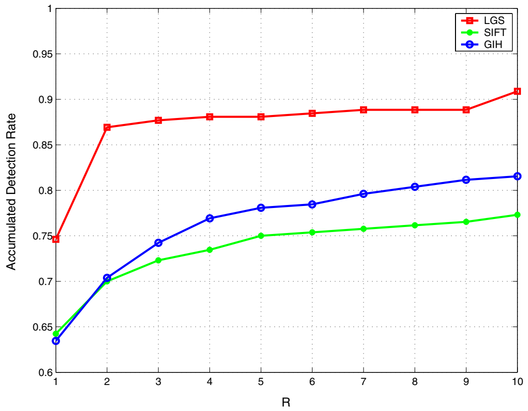


Figure 12. Experimental results on images with synthetic deformations.

shows the accumulated detection rate of the LGS, SIFT, and GIH approaches on images with synthetic deformations.

6. Conclusions

In this paper, we have proposed a novel local region descriptor using multi-size support regions centered at an interest point. We have developed an LGS model for similarity measure that takes advantage of the multiple support regions. The approach has been evaluated on images with variety of deformations including fisheye lens deformation, nonrigid deformation, affine deformation, and other various synthetic deformations. The experiments show that our deformable local image descriptor is robust to general image deformations, and it outperforms existing techniques for point matching.

Acknowledgments

This research was partially supported by the gift grants from the General Motors, Microsoft; grants from NSF(No. IIS-0534625), NIH(No. 1U01 HL091736 01), the National Basic Research Program of China (No. 2007CB311000), and the NSFC (No. 60705024). The first author was also partially supported by the scholarship from China Scholarship Council. The work was performed when the first author was at Carnegie Mellon University. We also thank the anonymous reviewers for their valuable suggestions.

References

[1] V. Athitsos, J. Alon, S. Sclaroff, and G. Kollios. Boostmap: An embedding method for efficient nearest neighbor retrieval. In *Proc. CVPR*, pages 268–275, 2004. 2, 5

[2] B. Babenko, P. Dollar, and S. Belongie. Task specific local region matching. In *Proc. ICCV*, 2007. 1, 2

[3] S. Belongie and M. Jitendra. Shape matching and object recognition using shape contexts. *PAMI*, 24(4):509–522, 2002. 1

[4] A. C. Berg and J. Malik. Geometric blur for template matching. In *Proc. CVPR*, 2001. 1

[5] G. Carneiro and A. D. Jepson. Flexible spatial configuration of local image features. *PAMI*, 29(12):2089–2104, 2007. 1

[6] W. T. Freeman and E. H. Adelson. The desgin and use of steerable filter. *PAMI*, 13(9):891–906, 1991. 1

[7] A. Frome, Y. Singer, F. Sha, and J. Malik. Learning globally-consistent local distance functions for shape-based image retrieval and classification. In *Proc. ICCV*, 2007. 2

[8] G. Griffin, A. D. Holub, and P. Perona. Caltech-256 object category dataset. In *Caltech Technical Report*, 2007. 5, 7

[9] C. Harris and M. Stephens. A combined corner and edge detector. In *Proceedings of The Fourth Alvey Vision Conference*, pages 147–151, 1988. 3

[10] G. Hua, M. Brown, and S. Winder. Discriminant embedding for local image descriptors. In *Proc. ICCV*, 2007. 1

[11] A. E. Johnson and M. Hebert. Using spin images for efficient object recognition in cluttered 3d scenes. *PAMI*, 21(5):433–449, 1999. 1

[12] Y. Ke and R. Sukthankar. PCA-SHIFT: A more distinctive representation for local image descriptors. In *Proc. CVPR*, volume II, pages 506–513, 2004. 1, 5

[13] H. Lejsek, F. H. Asmundsson, and B. T. Jonsson. Scalability of local image descriptors: A comparative study. In *ACM Multimedia*, pages 589–598, 2006. 1

[14] V. Lepetit and P. Fua. Keypoint recognition using randomized trees. *PAMI*, 28(9):1465–1479, 2006. 1

[15] H. Lin and D. W. Jacobs. Deformation invariant image matching. In *Proc. ICCV*, pages 1466–1473, 2005. 1, 2, 5, 6

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

[17] S. Mahamud and M. Hebert. The optimal distance measure for object detection. In *Proc.CVPR*, pages 248–255, 2003. 2

[18] K. Mikolajczyk and C. Schmid. Scale and affine invariant interest point detectors. *IJCV*, 60(1):63–86, 2004. 1

[19] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *PAMI*, 27(10):1615–1630, 2005. 1, 5

[20] E. N. Mortensen, H. Deng, and L. Shapiro. A SIFT descriptor with global context. In *Proc. CVPR*, 2005. 1, 2

[21] A. Opelt, M. Fussenegger, A. Pinz, and P. Auer. Weak hypotheses and boosting for generic object detection and recognition. In *Proc. ECCV*, pages 71–84, 2004. 1

[22] G. Shakhnarovich. Learning task-specific similarity, 2006. MIT PhD thesis. 2

[23] T. Tuytelaars and C. Schmid. Vector quantizing feature space with a regular lattice. In *Proc. ICCV*, 2007. 1

[24] A. Vedaldi and S. Soatto. Local features, all grown up. In *Proc. CVPR*, 2006. 1

[25] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Proc. CVPR*, pages 511–518, 2001. 4

[26] S. A. Winder and M. Brown. Learning local image descriptors. In *Proc. CVPR*, pages 1–8, 2007. 1

[27] L. Yang, P. Meer, and D. J. Foran. Multiple class segmentation using a unified framework over mean-shift patches. In *Proc. CVPR*, pages 1–8, 2007. 1