# A Multi Support Rotationally Invariant Descriptors Based on Gradient Histogram

**Ashutosh Singh Kushwaha, Vaibhav Purwar**

*Abstract*— **Local image is crucial in many computer vision tasks description. A good local image descriptor is expected to have high discriminative ability, so that the described point can be easily distinguished from then point. This paper presents a novel method for calculating local pool feature based on their gradients orders in multiple support region.**

**This paper analyze the effects on the rotationally descriptors based on gradient histogram. Basically pooling by intensity orders is not only invariant to rotation and monotonic intensity changes but also encodes ordinal information in to a descriptor. On local features on gradients descriptor obtained, the Multi-support region order based gradient histogram (MROGH). We determine the experimental results on image matching with different local descriptors, such as Scale invariant feature transform (SIFT), Gradient location and orientation histogram (GLOH) and Principle component analysis (PCA).**

*Index Terms*— **Local image descriptor, rotation invariance, MROGH, image matching, SIFT, GLOH, PCA.**

## I. Introduction

Computer vision was seen as more of a mathematical problem, which was additionally limited by the computational resources of the time. During that period, many of the tools and basic approaches we still use today were developed. Only in the last decade has the technology advanced sufficiently to tackle lofty computer vision problems and obtain reliable results that are making it into real world applications everywhere.

A good local image descriptor is expected to have high discriminative ability so that the described point can be easily distinguished from other points. Meanwhile, it should also be robust to a variety of possible image transformations, such as scale, rotation, blur, illumination, and viewpoint changes, so that the corresponding points can be easily matched across images which are captured under different imaging conditions. Improving distinctiveness while maintaining robustness is the main concern in the design of local image descriptors. In this thesis, we focus on designing local image descriptors for interest regions.

This works proposes a novel method for descriptor construction. By using gradient-based local features, local image descriptors are obtained, Multi-support Region Order-Based Gradient Histogram (MROGH) and Gradient location and orientation histogram. They are rotation invariant without relying on a reference orientation and still have high discriminative ability. Therefore, they possess the advantages of the existing two kinds of descriptors, but effectively avoid their disadvantages.

**Ashutosh Singh Kushwaha**, Electronics and Communication Engineering, M.Tech Student, Kanpur Institute of Technology, Kanpur, India.

**Mr. Vaibhav Purwar**, Electronics and Communication Engineering, Head of Department, Kanpur Institute of Technology, Kanpur, India.

## II. Related Work

There are three main steps in matching points by local image descriptors. The first step is to detect points in images. The detected points should be detectable and matchable across images which are captured under different imaging conditions. Interest point detectors and descriptors have become popular for obtaining image to image correspondence for 3D reconstruction [1] searching databases of photographs [2] and as a first stage in object or place recognition [3]. In a typical scenario, an interest point detector is used to select match able points in an image and a descriptor is used to characterize the region around each interest point. The output of a descriptor algorithm is a short vector of numbers which is invariant to common image transformations and can be compared with other descriptors in a database to obtain matches according to some distance metric.

Local image descriptors have received a lot of attention in the computer vision community. Many local descriptors have been developed since the 1990s [4], [5], [6], [7], [8], [9]. Perhaps one of the most famous and popular descriptors is Scale Invariant Feature Transform (SIFT) [4]. According to the comparative study of Mikolajczyk and Schmid [10] have systematically compared the performance of ten recent descriptors and they advocate their GLOH descriptor which was found to outperform other candidates.

Besides low-level features (e.g., histogram of gradient orientation in SIFT) which are used for descriptor construction, choosing an optimal support region size is also critical for feature description. Some researchers reported that a single support region is not enough to distinguish incorrect matches from correct ones. Mortensen et al. [11] proposed combining SIFT with global context computed from curvilinear shape information in a much larger neighborhood to improve the performance of SIFT, especially for matching images with repeated textures. Harada et al. proposed a framework of embedding both local and global spatial information to improve the performance of local descriptors for scene classification and object recognition. Cheng et al. [12] proposed using multiple support regions of different sizes to construct a feature descriptor that is robust to general image deformations. In their work, a SIFT descriptor is computed for each support region, then they are concatenated together to form their descriptor. Moreover, they further proposed a similarity measure model, Local-to-Global Similarity model, to match points described by their descriptors.

Our work is fundamentally different from the previous ones. Many previous methods are not strictly rotation invariant since they need to assign a reference orientation for each interest point, such as SIFT, GLOH, and PCA. Since our proposed descriptors do not rely on a reference orientation, they should potentially be morerobust. In fact, the need for an orientation for reference is also a drawback and bottleneck of the previous methods which utilize multiple support regions, hence largely differentiating our method from them. Although some local descriptors such as spin image and Rotation Invariant Feature Transform (RIFT) [13] achieve rotation invariance without requiring a reference orientation, they are less distinctive since some spatial information is lost due to their feature pooling schemes. In our work, local features are pooled together according to their intensity orders. Such a feature pooling scheme is inherently rotation invariant, and also invariant to monotonic intensity changes.

## III. THE PROPOSED METHOD

The key idea of our method is to pool rotation invariant local features based on intensity orders. Instead of assigning a reference orientation to each interest point to make the computation of local features rotation invariant, we calculate local features in a locally rotation invariant coordinate system. Thus, they are inherently rotation invariant. Meanwhile, sample points are adaptively partitioned into several groups based on their intensity orders. Then, the rotation invariant local features of sample points in these groups are pooled together separately to construct a descriptor. Since the intensity orders of sample points are rotation invariant, such a feature pooling scheme is also rotation invariant.

On local features on gradients descriptor obtained, the Multi-support region order based gradient histogram (MROGH).

### *A. Affine Normalized Regions*

The detected regions for calculating descriptors are either circular or elliptical regions of different sizes based on the used region detectors, this work is to design a local descriptor of the normalized region, which is a circular region of radius 20.5 pixels. Thus, the minimal patch that contains the normalized region is in size of 41 x 41 pixels. A similar patch size is also used in [10].

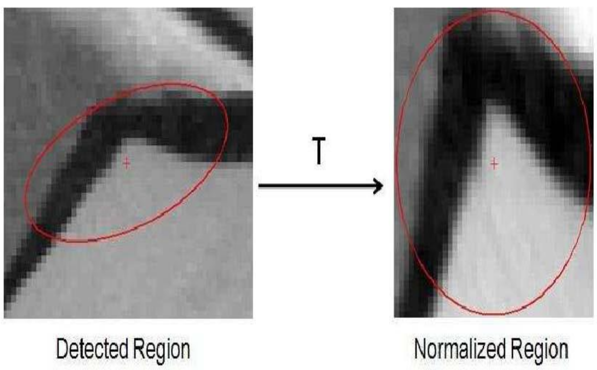



Figure 1 Affine normalization of a detected region to the canonical circular region

If the detected region is larger than the normalized region, the image of the detected region is smoothed by a Gaussian kernel before region normalization.

### *B. Descriptors*

We present the implementation details for the descriptors used in our experimental evaluation. one single support region is not enough to distinguish incorrect matches from correct ones in general. Two non corresponding interest points may accidently have similar appearances in a certain local region. However, it is less likely that two non corresponding interest points have similar appearances in several local regions of different sizes. In contrast, two corresponding interest points should have similar appearances in a local region of any size, although some small differences may exist due to localization error of interest point and region detection. As shown in Fig. 2, we choose support regions as the N nested regions centered at the interest point with an equal increment of radius. In each support region, the rotation invariant local features of all the sample points are then pooled by their intensity orders.

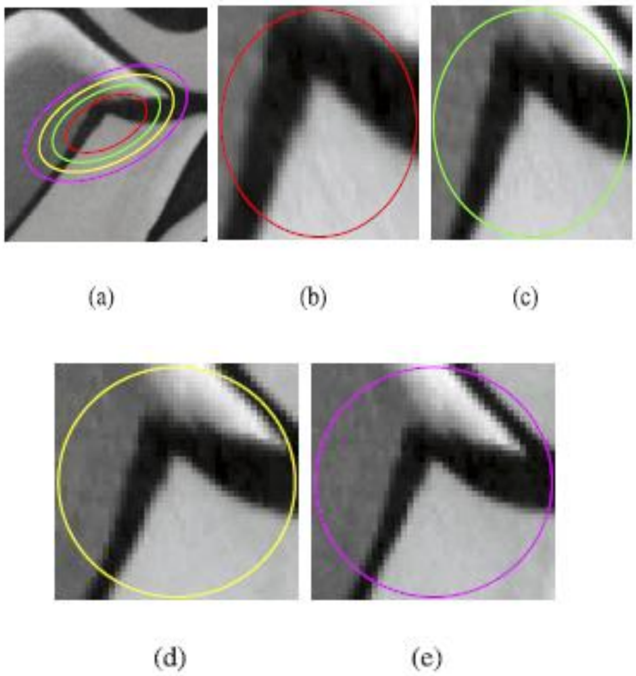



Figure 2 Selection of four support regions and their normalization. (a) The selected support regions and (b)-(e) the normalized regions.

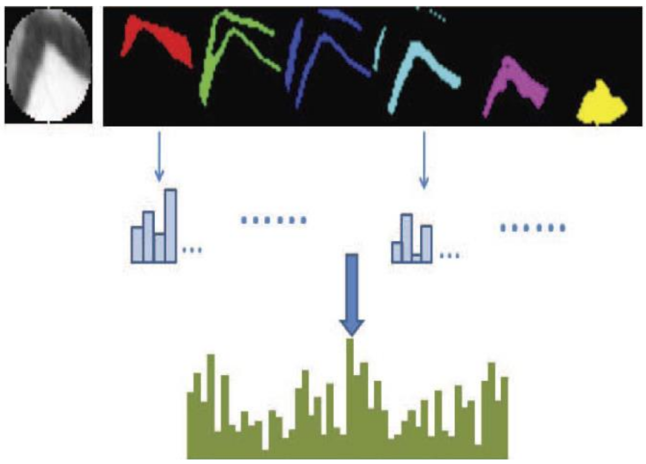


Figure 3 Procedure of pooling local features based on their intensity orders for a support region

As shown in Fig. 3, first they are pooled together to form a vector in each partition which is obtained based on the intensity orders of sample points and then the accumulated vectors of different partitions are concatenated together to represent this support region.

## IV. Performance Evaluation

In order to evaluate their influences on the performance of the proposed descriptors, we conducted image matching experiments on 142 pairs of images[2] with different parameter settings. These 142 image pairs are mainly selected from the data set of zoom and rotation transformations. Note that they do not contain image pairs in the standard Oxford data set [14] because those image pairs are used for the descriptors evaluation in the later stage.

### *A. Data Set*

The proposed descriptors were evaluated on the standard Oxford data set, in which image pairs are under various image transformations, including viewpoint change, scale and rotation changes, image blur, JPEG compression, and illumination change. It can be seen from these results that although the performance of each descriptor varied with different feature detectors (Hessian-Affine or Harris-Affine in this experiment), the relative performance among different descriptors is consistent with different feature detectors. The results show that MROGH and other evaluated local descriptors in all the tested cases.

Figure 4 Example of Bike image (data set)

### *B. Image Matching*

To evaluate the performance of the proposed descriptors, we conducted extensive experiments on image matching. We followed the evaluation procedure proposed by Mikolajczyk and Schmid [10]. The codes for evaluation were downloaded from their website [14]. Three other local descriptors were evaluated in our experiments for comparison: SIFT, GLOH and PCA.

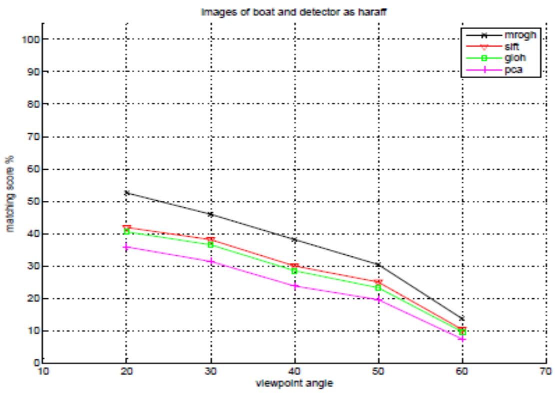



Figure 5 Matching score images of boat and detector as haraff

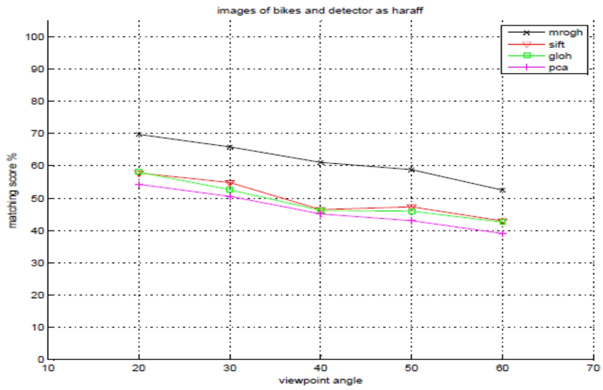
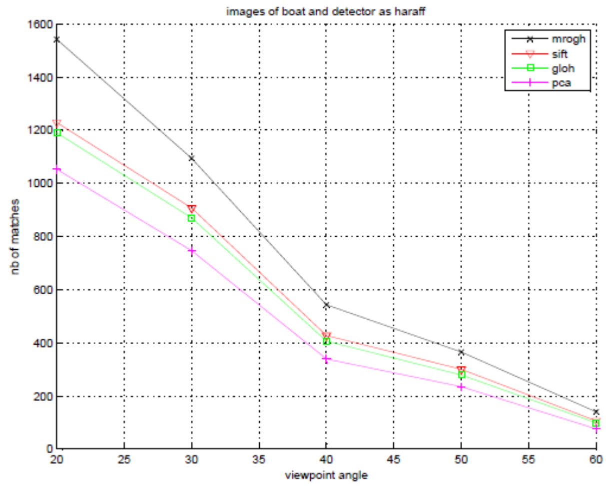


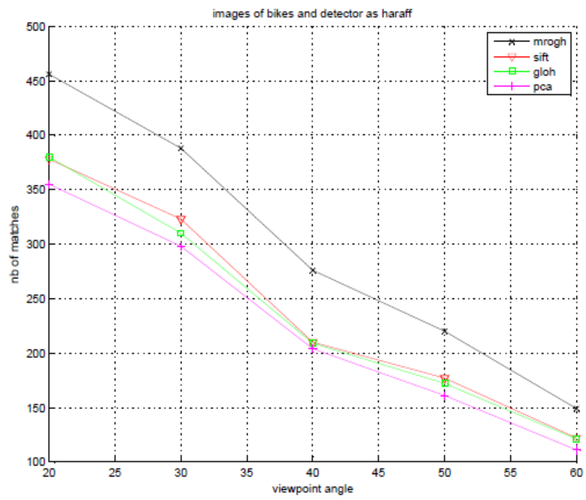

Figure 6 Matching score images of bike and detector as haraff

### *C. Performance Evaluation Based on 3D Objects*

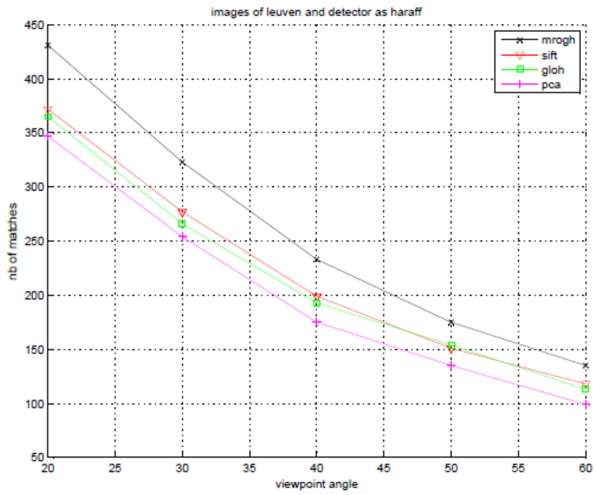
Moreels and Perona [15] evaluated different combinations of feature detectors and descriptors based on 3D objects. We evaluated the proposed descriptors following their work.The Hessian-Affine detector was used for feature extraction since it was reported with the best results combined with SIFT in [15]. As in [15], interest point matching was conducted on a database containing both the target features and a large amount of features from unrelated images so as to mimic the process of image retrieval/object recognition.




(a)




(b)




(c)

Figure 6 ROC curves of four different local descriptors on the data set

It can be seen from Figures 6 (a), 6 (b) and 6(c) that the matching performances of SIFT, GLOH, PCA and MROGH.

## V. Conclusion

In this paper, we have presented an experimental evaluation of interest region descriptors in the presence of real geometric and photometric transformations. The goal was to compare descriptors computed on regions extracted with recently proposed scale and affine-invariant detection techniques. Note that the evaluation was designed for matching and recognition of the same object or scene..

## References


[1] N. Snavely, S. M. Seitz, and R. Szeliski. Photo tourism: Exploring photo collections in 3d. In ACM Transactions on Graphics (SIGGRAPH Proceedings), volume 25, pages 835– 846, 2006.

[2] K. Mikolajczyk and C. Schmid. Indexing based on scale invariant interest points. In Proceedings IEEE International.

[3] D. Nist´er and H. Stew´enius. Scalable recognition with a vocabulary tree. In Proceedings IEEE Conference on Computer Vision and Pattern Recognition, pages 2161–2168, 2006.

[4] D.G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," Int'l J. Computer Vision, vol. 60, no. 2, pp. 91-110, 2004.

[5] M. Heikkila, M. Pietikainen, and C. Schmid, "Description of Interest Regions with Local Binary Patterns," Pattern Recognition, vol. 42, pp. 425-436, 2009.

[6] J. Chen, S. Shan, G. Zhao, X. Chen, W. Gao, and M. Pietikainen, "A Robust Descriptor Based on Weber's Law," Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2008.

[7] H. Bay, A. Ess, T. Tuytelaars, and L.V. Gool, "SURF: Speeded Up Robust Features," Computer Vision and Image Understanding, vol. 110, no. 3, pp. 346-359, 2008.

[8] F. Tang, S.H. Lim, N.L. Change, and H. Tao, "A Novel Feature Descriptor Invariant to Complex Brightness Changes," Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 2631-2638, 2009.

[9] R. Gupta, H. Patil, and A. Mittal, "Robust Order-Based Methods for Feature Description," Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 334-341, 2010.

[10] K. Mikolajczyk and C. Schmid, "A Performance Evaluation of Local Descriptors," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 27, no. 10, pp. 1615-1630, Oct. 2005.

[11] E. Mortensen, H. Deng, and L. Shapiro, "A SIFT Descriptor with Global Context," Proc. IEEE Conf. Computer Vision and Pattern Recognition, vol. 1, pp. 184-190, 2005.

[12] H. Cheng, Z. Liu, N. Zheng, and J. Yang, "A Deformable Local Image Descriptor," Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2008.

[13] S. Lazebnik, C. Schmid, and J. Ponce, "A Sparse TextureRepresentation Using Local Affine Regions," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 27, no. 8, pp. 1265-1278, Aug.2005.

[14] http://www.robots.ox.ac.uk/~vgg/research/affine/, 2012.

[15] P. Moreels and P. Perona, "Evaluation of Features Detectors andDescriptors Based on 3D Objects," Int'l J. Computer Vision, vol. 73,no. 3, pp. 263-284, 2007.


**Ashutosh Singh Kushwaha** obtained his B.Tech from HBTI, Kanpur, India. He is now a student of M.Tech (Electronics and Communication Engineering) in Kanpur Institute of Technology, Kanpur, India.

**Mr. Vaibhav Purwar** is a working as Professor and Head of Department of Electronics and Communication Engineering, Kanpur Institute of Technology, Kanpur, India..