

Local Descriptor based on Texture of Projections

N V Kartheek Medathati^{*}
Center for Visual Information Technology
International Institute of Information Technology
Hyderabad, India
nvkartheek@research.iiit.ac.in

Jayanthi Sivaswamy
Center for Visual Information Technology
International Institute of Information Technology
Hyderabad, India
jsivaswamy@iiit.ac.in

ABSTRACT

The aim of a local descriptor or a feature descriptor is to efficiently represent the region detected by an interest point operator in a compact format for use in various applications related to matching. The common design principle behind most of the mainstream descriptors like SIFT, GLOH, Shape context etc is to capture the spatial distribution of features using histograms computed over a grid around interest points. Histograms provide compact representation but typically lose the spatial distribution information. In this paper, we propose to use projection-based representation to improve a descriptor's capacity to capture spatial distribution information while retaining the invariance required. Based on this proposal, two descriptors based on the CS-LBP are introduced. The descriptors have been evaluated against known descriptors on a standard dataset and found to outperform, in most cases, the existing descriptors. The obtained results demonstrate that proposed approach has the advantages of both the statistical robustness of histogram and the capability of the projection based representation to capture spatial information.

Keywords

Image Matching, Local Descriptor, Texture of Projections

1. INTRODUCTION

Local features play an important role in a wide variety of applications such as wide-baseline matching, image retrieval, robot localization, object recognition etc and due to their applicability a lot of feature descriptors have been developed in the recent times. The general aim of the descriptors is to capture the distribution of features such as gradient orientation or response to a particular kind of filter around the interest points. Traditionally this has been done with help of histograms of features computed over a defined grid.

^{*}Corresponding author

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ICVGIP '10, December 12-15, 2010, Chennai, India

Copyright 2010 ACM 978-1-4503-0060-5/10/12 ...\$10.00.

Several different kinds of descriptors can be generated based on the combination of features extracted and the grid over which the histograms are computed. For example Scale Invariant Feature Transform[5] is a 3D histogram of gradient orientations computed over a 4x4 square grid where as Gradient Location and Orientation Histogram[6] (GLOH) is computed over a log-polar grid. A recent study [10] has shown that the computation of most of the local descriptors that have been proposed till date can be divided into two main stages: The first stage is a feature extraction stage where features like gradients or Gabor filter responses are extracted over a given patch. The second stage is a pooling stage where the gradient like features within the patch are represented using histograms computed over pre-defined grids.

An important factor that influences the performance of these descriptors, apart from the choice of features, is the grid over which the histograms are computed. The grid plays a key role in determining the level of spatial distribution information that can be captured. Towards addressing this issue the problem of descriptor construction has been posed in [10] as an optimization problem to maximize the similarity scores of true matches while maximizing the dissimilarity of the negative matches from an independently computed ground truth data. Alternate ways to encode the spatial distribution information into the histograms have not been investigated.

In contrast to the steerable filters or gradient based features that were used earlier, a local variational feature has been investigated resulting in the local binary pattern (LBP) descriptor [7]. The LBP is popular for texture description. This has been extended for a matching scenario by making some improvements in terms of compactness and by constructing a descriptor (called Center-Surround LBP (CS-LBP)) on the same lines as SIFT using a 4x4 grid to compute histograms [2]. The results obtained by CS-LBP are shown to be better than the ones obtained using SIFT.

The discriminating power of a descriptor is good if it can capture the spatial distribution information and at the same time remain invariant to common photometric and geometric distortions. Spatial distribution information can be better captured by computing histograms over finer grids but this approach makes the descriptor more susceptible to geometric or photometric distortions. [2] reports that an empirical study on the optimal grid size revealed that a 4x4 grid, similar to SIFT, was the best compared to coarser ones such as 2x2 or much finer grids such as 8x8. Coarse grids lose the required granularity for information capture while fine

grids tend to be sensitive to distortions. The demanding task for a descriptor is to be as representative as possible without losing its invariance. Since finer grids have been shown to be susceptible to distortions, we wish to investigate an alternative route to incorporate the spatial distribution information.

Radon transformation or projection based representation is known to be effective in capturing the spatial distribution information of the image pixels. It has been successfully used for development of shape descriptors like R-Transform[9], Histogram of Radon transform[8], Radon Representation based feature descriptor [4] etc. Even though these methods have been shown to be useful for robust shape description, Radon representation based features are not used in the context of building local descriptors for image matching. It has been mainly due to two reasons: the descriptors are mainly designed for handling binary shape images and their potentially high dimensionality. In this paper, we investigate how projection based information can be incorporated into a local feature descriptor. We do this by studying binary patterns computed in the Radon transform domain. We propose to construct a descriptor by concatenating the CS-LBP pattern of a patch along with CS-LBP pattern of its projections. The rest of the paper is organized as follows. In section 2 we briefly review the CS-LBP operator, section 3 gives a brief description of our idea using binary images as an example, section 4 describes the construction of our descriptor and section 5 provides details of experimental methodology.

2. CENTER-SURROUND LOCAL BINARY PATTERN

The CS-LBP descriptor is a modified and extended version of LBP descriptor proposed in [7] for the purpose of image matching. The CS-LBP operator is computed by comparing the center symmetric pairs of pixels instead of comparing the center pixel with all the surrounding pixels which is done in generating the LBP. The main advantage of computing the centre symmetric rather than centre to surround difference, is that the length of the descriptor is reduced significantly (by half). The CS-LBP is computed for a pixel using the neighbors at radius R , as follows

$$CS-LBP_{R,N,T}(x,y) = \sum_{i=0}^{\frac{N}{2}} s(n_i - n_{i+\frac{N}{2}}) 2^i \quad (1)$$

where $s(x) = 1, \forall x > 0$ and $s(x) = 0$ otherwise. CS-LBP as an operator is shown to be good at capturing the local variational patterns amongst the pixels. It has several desirable properties such as ease of computation and illumination invariance.

3. PROPOSED DESCRIPTOR: TEXTURE OF PROJECTIONS

Consider the segmented binary shapes as shown in Fig.1. The corresponding sinograms or the Radon transform of these binary shapes are also presented. It can be observed from this figure that different shapes give rise to different kind of textures in the sinograms. In the shape description literature, most of the techniques which use Radon transform have only attempted to bring in invariance to translation and rotation by computing shift invariant representations like Histogram of Radon transform [8] or Fourier

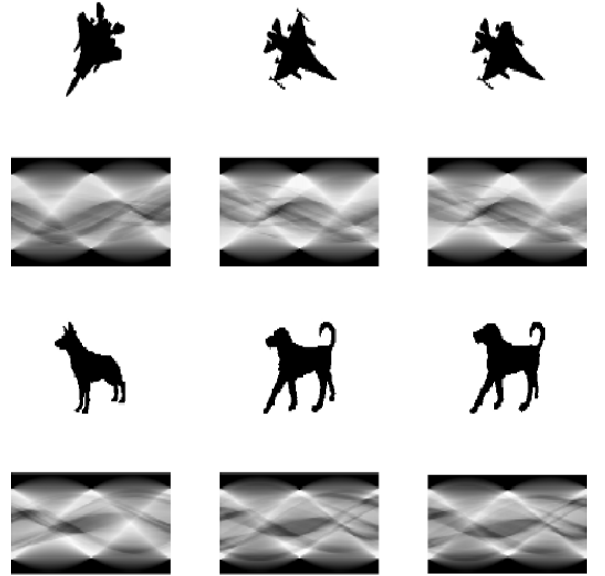


Figure 1: Binary Shapes and Their Projections

transform of the Radon transform [11]. We argue that at an abstract level, shape can be characterised by measuring the variability of its projections at different angles and radial distances. A variational descriptor like CS-LBP, captures just this kind of information and also provides a compact representation and should serve an ideal starting point. Hence, it should be possible to develop a good shape representation by computing the CS-LBP on the projection space. However, in the case of greyscale images, this approach would be inadequate. This can be explained as follows. When a greyscale image is projected (at a particular angle) an obtained ray sum can be due to any combination of pixel values along the ray. This implies that the variational pattern of the projections alone is insufficient to represent a region in an image. If in addition a histogram of the region is also provided, it can serve as an additional constraint and help improve the discriminability of the representation. This can further be illustrated via an example based on the popular Su Do Ku puzzle. Let us consider a 3×3 grid which has to be filled with numbers k using two simple rules: The sum of pixels along a column or row has to be a constant value $3k$ and any specific pixel value can occur along a row or column only once. The puzzle is guaranteed to have a unique solution only when both rules are applied since, in the absence of the second rule, there are multiple solutions to the problem.

Thus, we propose to combine two kinds of information for achieving higher discriminability without losing the essential invariance of a local descriptor for greyscale images: i) a variational pattern computed in projection space and ii) the histogram information computed in the spatial domain. For capturing the first type of information we propose using CS-LBP of projections and for the latter, we propose using CS-LBP again as it contains the histogram information albeit of the variational pattern in the raw image. This choice to compute both parts of the proposed descriptor with the same base descriptor CS-LBP, should also help gain insight into the value addition that projection based information

can provide to a descriptor's performance in different tasks such as matching.

3.1 Descriptor Computation

In constructing the proposed descriptor, we considered two variants of the same idea.

- Type 1

In the first variant, given a patch, the CS-LBP is computed in the spatial domain and in the projection domain over the *entire* patch. The projection domain representation for the patch is obtained by computing the Radon transform of the patch. This descriptor is henceforth referred to as "PLBP".

- Type 2

In the second variant, the patch is subdivided first into $n \times n$ blocks and CS-LBP of the projections of each block is computed. A histogram of length n^2 is computed for each CS-LBP pattern. The length of the total CS-LBP histograms thus formed is $16 * n^2$, are combined with the original CS-LBP histogram of the patch to construct the final descriptor. This method is henceforth referred to as 'PLB1'.

4. IMPLEMENTATION DETAILS

We used a Hessian-Affine region detector for detecting interest points. The detected regions were first affine normalized to a size of 41x41 before computing the descriptors. This size is as per the standards of detector literature.

The CS-LBP used for computation is our own implementation of the algorithm. Since the main idea behind this paper is to check the validity of the idea behind combination of the complementary information, basic parameter settings were used for the CS-LBP operator. The CS-LBP implementation was with the following parameters Radius = 1, Number of nearest neighbours = 8 and Threshold = 0. These parameters were also reported to be performing well in [2]. As given in the CS-LBP implementation, we use a 4x4 grid to finally compute the histograms for spatial domain representation of the patch along with the projection domain representation of the patch. The number of projections for Radon transform computation was empirically set to 60.

In the computation of PLBP and PLB1 normalization of the descriptors is done in a similar fashion to that of the SIFT as follows: Initially the descriptor is normalized to unit length. Then all the bins having a maximum value of 0.2, are clipped to 0.2 and the descriptor is re-normalized.

5. EVALUATION

The proposed descriptors were evaluated on the standard dataset using the standard matching protocol provided by [6]. The underlying performance measure for this protocol is the recall versus false positive ratio. The performance of the designed method was compared with some state of the art descriptors like SIFT, GLOH, Shape Context and native CS-LBP. Computation of all these descriptors except CS-LBP, have been with the binaries provided by the Robotics group at Oxford [3] have been used.

The standard dataset contains different image sets with different geometric and photometric transformations. It covers six different types of changes for has both pairs of structured and textured scenes. The transformations provided

are: viewpoint change, scale change, image rotation, image blur, illumination change, and JPEG compression. For each category there are a set of six images with established ground truth homographies. For a given detector and descriptor pair, the performance of the descriptor is measured using the following steps,

1. Accurate number of correspondences are measured by projecting the regions detected on one image on to other and if the overlap error is below a threshold, then the patches are said to be corresponding. The overlap threshold is set to 0.5 in our case.

2. The ground truth number of correspondences also depend on the matching strategy used. Here, we test the descriptor using two matching methods.

- Nearest Neighbour Method: Two points are said to be corresponding if the distance between their descriptors is the minimum and is below a threshold. This implies there is one to one matching.

- Threshold based or Similarity Based Matching: Two points are said to be corresponding if the distance between their descriptors is below a threshold. Here, a point can have many correspondences. Even though it might look counter intuitive to consider this kind of matching Schmid *et al* have reasoned that when matching is performed on a large database of descriptors, it is very useful.

3. Finally for performance evaluation, the threshold parameter is varied to obtain a plot of recall versus (1-precision).

$$recall = \frac{\text{No. of correct matches}}{\text{No. of correspondences}} \quad (2)$$

$$1 - precision = \frac{\text{No. of false matches}}{\text{Total no. of matches}} \quad (3)$$

We have evaluated the descriptors based on both nearest neighbour based and threshold based matching. The following sections provide the results obtained and their analysis.

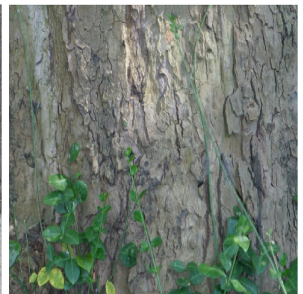
5.1 Results and Analysis

The performance results for matching the image pairs shown in Fig.2 are shown in the Fig.3 and 4. The performance is compared with some standard descriptors like SIFT [5], GLOH [6] and Shape Context [1]. The axis of all the graphs is scaled between 0 and 1. The X-axis of the graph plot 1- precision and the Y-axis of the graph plots recall. The legend used for the plots is as follows, SIFT - Scale Invariant Feature Transform, GLOH - Gradient Location and Orientation Histogram, SCON - Shape Context, CLBP - Center - Surround Local Binary Pattern (Our implementation), PLBP, PBP1.

5.2 Performance for different transformations

- View point change

Based on the graphs for the Graffiti and wall images, we can observe that the performance of the PLBP is superior in handling view point change over all descriptors including PBP1. This implies a) that texture

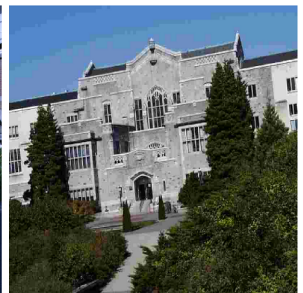
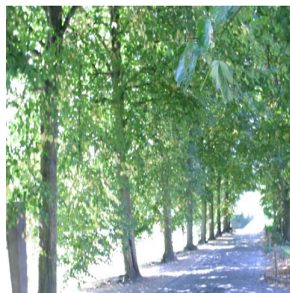
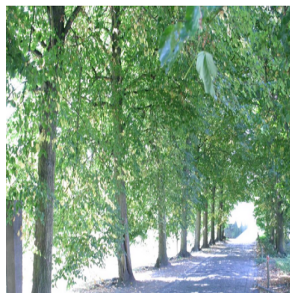


(e) Graf

(f) Wall

(g) Boat

(h) Bark



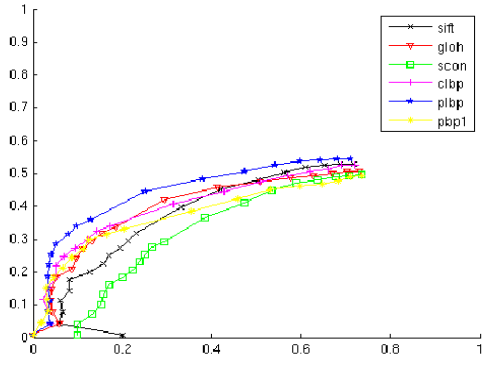
(m) Bikes

(n) Trees

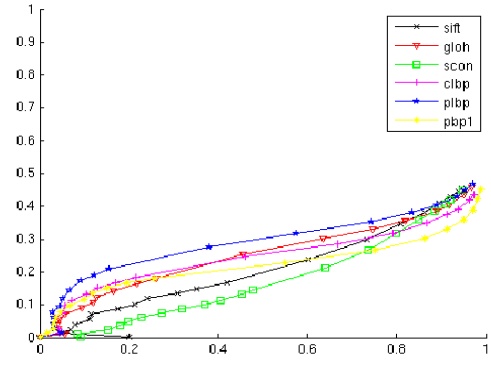
(o) Leuven

(p) UBC

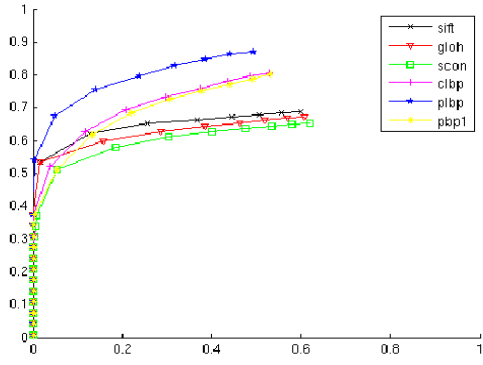
Figure 2: Image Pairs used for evaluation



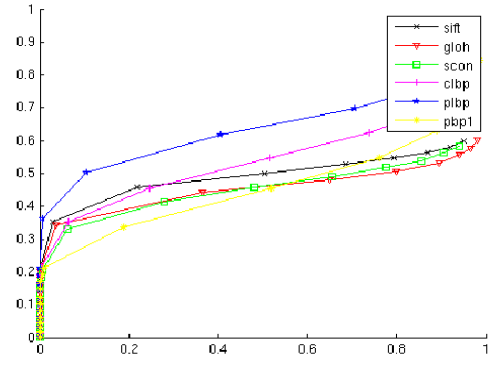
(a) Graffiti: Nearest Neighbour



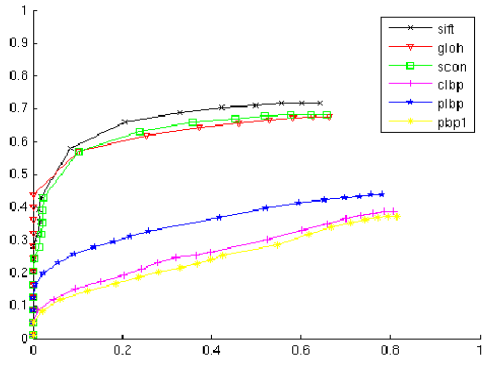
(b) Graffiti: Similarity Matching



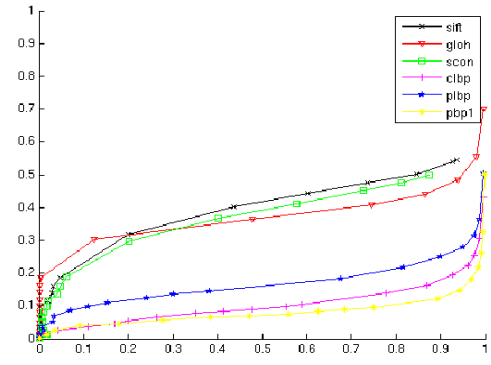
(c) Wall: Nearest Neighbour



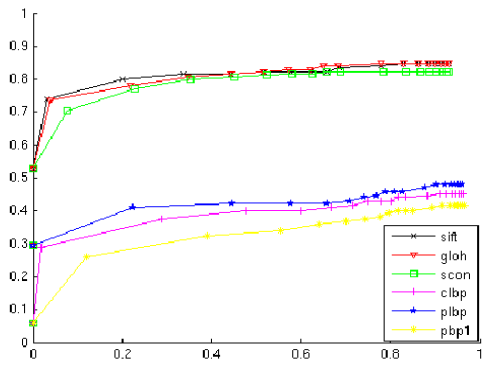
(d) Wall: Similarity Matching



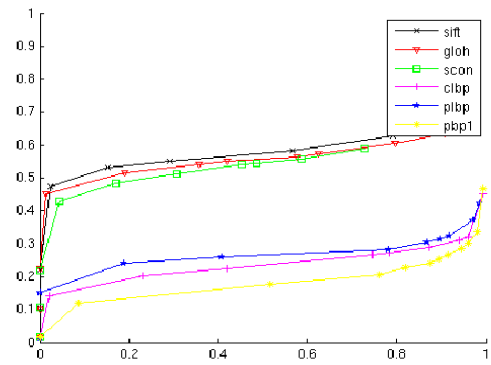
(e) Boat: Nearest Neighbour



(f) Boat: Similarity Neighbour

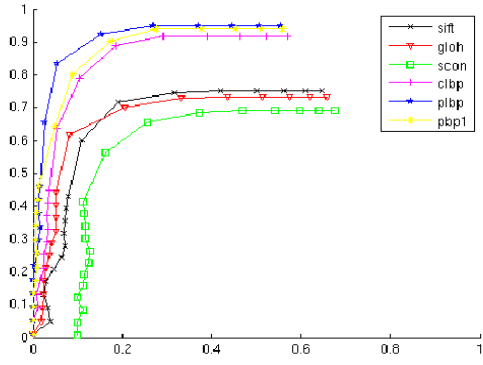


(g) Bark: Nearest Neighbour

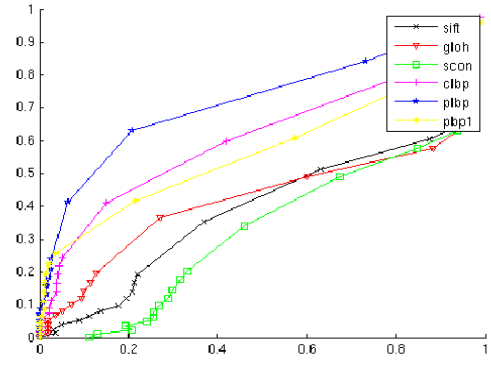


(h) Bark: Similarity Matching

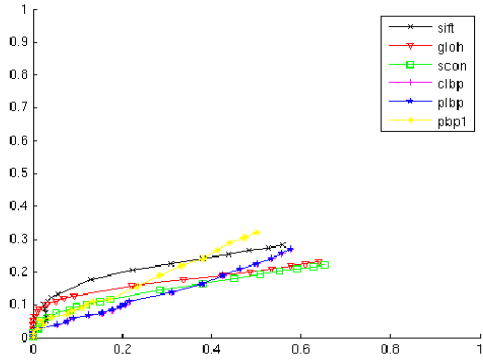
Figure 3: Performance of Various Descriptors over Hessian Affine Regions



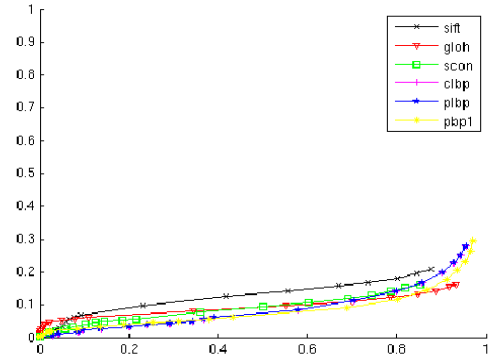
(a) Bikes: Nearest Neighbour



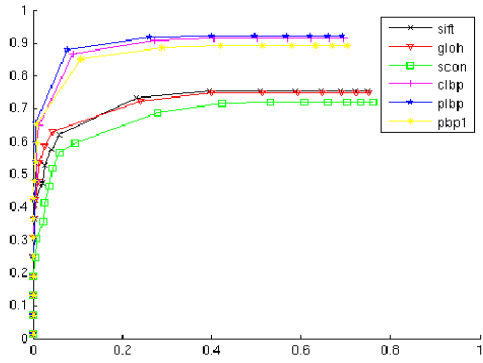
(b) Bikes: Similarity Matching



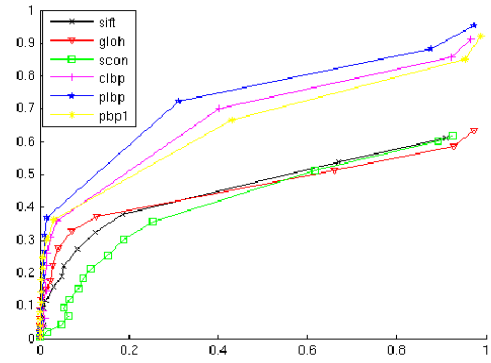
(c) Trees: Nearest Neighbour



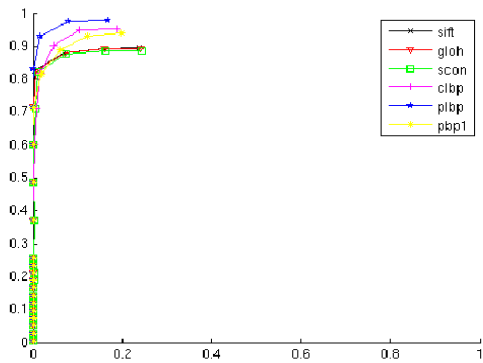
(d) Trees: Similarity Matching



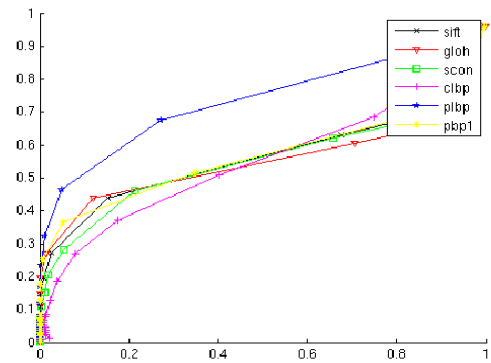
(e) Leuven: Nearest Neighbour



(f) Leuven: Similarity Matching



(g) UBC: Nearest Neighbour



(h) UBC: Similarity Matching

Figure 4: Performance of Various Descriptors over Hessian Affine Regions

of projections along with LBP is providing more invariant and robust information for matching; and b) texture of projections over smaller regions makes it more sensitive to changes which is to be expected.

- **Rotation and Zoom**

The graphs for the Boat and Bark images exhibit a compromised performance for all descriptors based on CS-LBP. This was discovered to be due to the following. The rotation correction routine implemented in our work does not account for scale as a consequence of which the normalisation is improper. With a correct scale-space implementation such as in [5], there is scope to address this problem and improve the performance. This is attested to by the better performance reported in [2] on these images.

- **Blur**

The difference in the graphs for the Bikes and trees illustrates that blur in a structured (bikes) compared to a textured (trees) scene indicates that the behaviour of the descriptor under blur depends on the content of the scene. This is due to the fact that a textured scene is more affected by blur, which is faithfully captured by the descriptor. We also observe that PBP1 performs better in blurred textured scene due to the fine grain texture information it is able to capture.

- **Illumination and JPEG compression**

LBP (and its variants including ours) by design handles illumination changes well. This can be seen from the graphs for Leuven. While all descriptors are robust to JPEG compression the textured projection information appears to give a slight edge to the performance as seen in graph for UBC image.

5.2.1 Dimensionality

One of the limitation of the proposed descriptors (PLBP and PBP1) is that the dimensionality of the descriptor is twice that of the CS-LBP operator. We believe that this is an aspect that can be addressed in the future using some dimensionality reduction techniques such as PCA or changing the histogram binning parameters. It is noteworthy however, that past attempts to increasing the number of patches from 4×4 to 8×8 have resulted in a poorer performance. This was evident in the performance test results given by Table 2, page 430, [2].

6. CONCLUSIONS AND FUTURE WORK

In this paper, we began by observing that spatial distribution information was lost in most of the existing approaches which use histograms. This was sought to be rectified by adding the information from projection space. Towards establishing the utility of this idea, we have proposed a method to incorporate spatial distribution information using variational patterns in projection domain. This is markedly different from the traditional way of addressing this issue by computing histograms on finer grids.

We have proposed two ways of executing the proposal for inclusion of projection-based information. These resulted in two descriptors based on the CS-LBP. Evaluation of these descriptors, using a standard evaluation protocol, have shown that the projection space has sufficient information to be

captured as the designed descriptor (PLBP) outperforms the traditional methods in most of the cases.

An interesting aspect of the proposed approach is that it shows that a variational pattern in projection domain can be used to capture useful information in grayscale images as well. In the future, we aim to test the proposed approach on tasks like object recognition.

7. REFERENCES

- [1] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(4):509–522, 2002.
- [2] M. Heikkilä, M. Pietikäinen, and C. Schmid. Description of interest regions with local binary patterns. *Pattern Recogn.*, 42(3):425–436, 2009.
- [3] <http://www.robots.ox.ac.uk/vgg/research/affine/>. Oxford, 2004.
- [4] G. Liu, Z. Lin, and Y. Yu. Radon representation-based feature descriptor for texture classification. *Image Processing, IEEE Transactions on*, 18(5):921–928, may 2009.
- [5] D. G. Lowe. Distinctive image features from scale-invariant keypoints, 2003.
- [6] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 27(10):1615–1630, 2005.
- [7] T. Ojala, M. Pietikainen, and T. Maenpää. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(7):971–987, jul 2002.
- [8] S. Tabbone, O. Ramos Terrades, and S. Barrat. Histogram of Radon Transform. A useful descriptor for shape retrieval. In *19th International Conference on Pattern Recognition - ICPR 2008*, Tampa United States, 2008.
- [9] S. Tabbone, L. Wendling, and J.-P. Salmon. A new shape descriptor defined on the radon transform. *Comput. Vis. Image Underst.*, 102(1):42–51, 2006.
- [10] S. Winder and M. Brown. Learning local image descriptors. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR07)*, Minneapolis, June 2007.
- [11] S. Xiao and Y. Wu. Rotation-invariant texture analysis using radon and fourier transforms. *Chin. Opt. Lett.*, 5(9):513–515, 2007.