Rotationally and Illumination Invariant Descriptor based on Intensity Order

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

Master of Technology

in

Electronic Systems and Communication

By

Buddarthi Suresh

Roll no: 710EE1066

Department of Electrical Engineering

National Institute of Technology

Rourkela, Orissa, India

May 2015

Rotationally and illumination invariant descriptor based on intensity order

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

Master of Technology

in

Electronic Systems and Communication

By

Buddarthi Suresh

Roll no: 710ee1066

Under the supervision of

Prof. Dipti Patra

Department of Electrical Engineering

National Institute of Technology

Rourkela, Orissa, India

May 2015

DEPARTMENT OF ELECTRICAL ENGINEERING

NATIONAL INSTITUTE OF TECHNOLOGY, ROURKELA

ODISHA, INDIA-769008

CERTIFICATE

This is to certify that the thesis entitled "Rotationally and illumination invariant descriptor based on intensity order", submitted by **Buddarthi Suresh (Roll. No.710EE1066),** in partial fulfillment of the requirements for the award of **Master of Technology** in **Electrical Engineering** during session 2014-2015 at National Institute of Technology, Rourkela is a bonafide record of research work carried out by him under my supervision and guidance.

To the best of my knowledge, the matter embodied in the thesis has not been submitted to any other university/institute for the award of any Degree or Diploma

Prof. Dipti Patra

Place: Rourkela

Dept. of Electrical Engineering

National Institute of Technology

Rourkela-769008

Acknowledgements

I would like to thank my guide Prof.Dipti patra for her invaluable suggestions and support. Her constant supervision and encouragement made this work successful. I am very fortunate to have worked under supervision and I am very thankful to god for giving this great opportunity.

And I Cannot forget the support and motivation provided by fellow lab mates in “Image Processing and Computer Vision lab” I thank Smita Pradhan, Rajashree Nayak and Yogananda Patnaik, Umesh Kumar sahoo, Pranav GS,Rajkumar Maharaju and Saka harsha Vardhan .

I thank all my friends Thanda Shesha Salpa sai, Srinath Koilakonda, Suresh Gurjar, Sarbajeith jena, Ruben kanduna, Rakesh Kommineni, Vinay Balmuri and Dakkamadugula Ravi.

At last I thank my family for believing me more than, I believe myself and for having all the hopes on me .

710ee1066

Buddarthi suresh

Content

INDEX OF IMAGES

Abstract

In this thesis, a novel method for local feature description where local features are grouped in normalized support regions with the intensity orders is proposed. Local features extracted using this kind of method are not only gives advantage of invariant to rotation and illumination changes, but also converts the image information into the descriptor. These features are calculated with different ways, one is based on gradient and other one is based on the intensity order. Local features calculated by the method of the gradient performs well in most of the cases such as blur, rotation and large illuminations and it overcome the problem of orientation estimation which is the major error source for false negatives in SIFT.

In order to overcome mismatching problem, method of multiple support regions are introduced in the proposed method instead of using single support region which performs better than the single support region, even though single support region is better than SIFT.

The idea of intensity order pooling is inherently rotational invariant without estimating a reference orientation. Experimental results show that the idea of intensity order pooling is efficient than the other descriptors, which are based on estimated reference orientation for rotational invariance.

Index Terms—Local image descriptor, Normalization of support regions, rotation invariance, illumination, matching, intensity orders, SIFT.

Chapter1: Introduction

1. Image descriptors

Local image descriptors which are calculated from local interest regions have become widely popular in image processing applications. These descriptors are much famous and useful for a number of vision jobs, such as object-tracking, object recognition and classification and also image stitching.

Local image descriptor should possess very high discriminative ability so that the point described by the descriptor should be distinctive enough for image matching. In the same way any image under any kind of changes such as illumination, scale, blur, rotation and the viewpoint changes can be easily recognized from the large data base, where the descripted key points could be sensitively matched with the images, which were captured already from different environmental image conditions. Increasing distinctiveness along with the maintained robustness is important for a good descriptor designing. Here utmost importance given to the designing local image descriptors for interest regions.

As a result of the new communication skills and the immense use of Internet in our humanity, the amount of audio-visual material available in digital format is increasing substantially. Hence, it has been required to design some schemes that allow us to define the content of several types of audiovisual aid information in order to search and organize them.

The audio-visual descriptors are in charge of the contents description. These descriptors have a good knowledge of the objects and events found in a video, image or audio and they allow quick and efficient searches of the audio-visual content.

This system can be related to the search engines for textual substances. Although it is certain, that it is relatively easy to find text with a supercomputer but much more difficult to find concrete audio and video parts. For instance, imagine someone pointed a scene of a happy person. The happiness is a feeling and it is not evident its shape, color and texture explanation in images.

The description of the audio-visual content is not an artificial task and it is essential for the effective use of this type of archives. The tuning system that deals with audio-visual descriptors is the MPEG-7 (*Motion Picture Expert Group - 7*)

1.1 Types of Descriptor

In computer_vision, **visual descriptors** or **image descriptors** are descriptions of the image features of the contents in images, videos, or processes or solicitations that produce such descriptions. They describe simple features such as the shape, the color, the texture or the motion, among others.

In the last decade so many descriptor have been proposed based the type of application such scale invariant, affine invariant, rotation invariant Descriptors which are robust to many environmental changes like noise, blur. Descriptors are the basic character or basic step to find out the connection between pixels contained in a digital image and what humans remember after having witnessed an image or a cluster of images after some minutes.

Visual descriptors are typically divided in to two main groups:

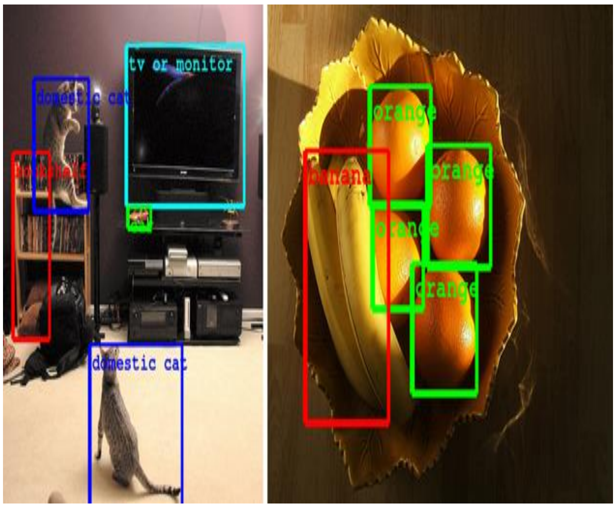
1. **General descriptors**: hold low level descriptors which give a description about color, shape, regions, textures and motion.
2. **Specific descriptors**: shows information regarding objects and events in the scene. A major examples are face recognition, object tracking, object recognition etc..

1.2 Descriptor applications

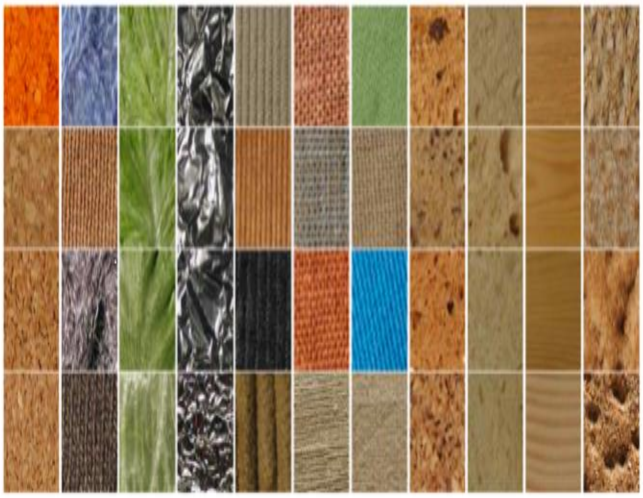
These are the most significant applications among all:

- Multimedia brochures search machines and identifiers object recognition, image matching.
- Visual descriptors allow a very detailed and existing search of any video or image by means of different search limitations. For instance, the search of films where a known actor appears, the search of videos containing the Everest Elevation, etc.
- Custom-made microelectronic news service.
- In possibility of an involuntary connection to a TV frequency distribution a soccer match, for example, whenever a player approaches the goal area.
- Regulation and sifting of strong and unwanted videos like pornographic matter which are dangerous threat to society. Also, approving some of the image contents in the internet.

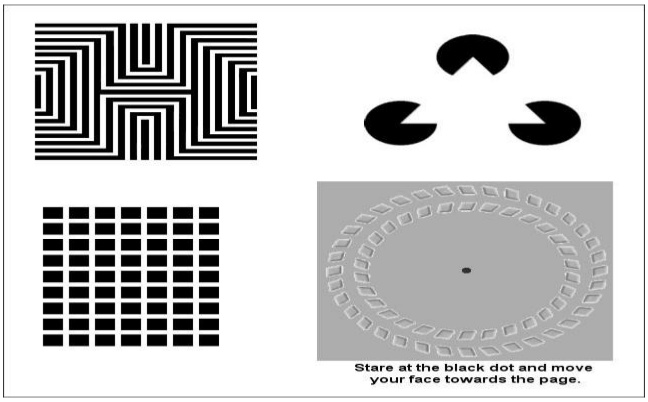
Object recognition:



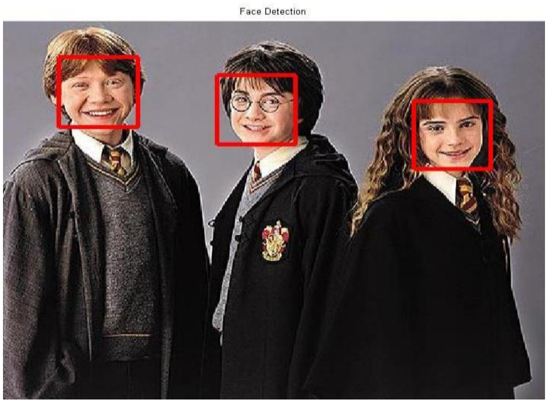
Texture classification:



Pattern recognition:



Face recognition:



Above are the some of the examples of useful application of descriptors. Such as face recognition, pattern recognition, texture recognition etc.

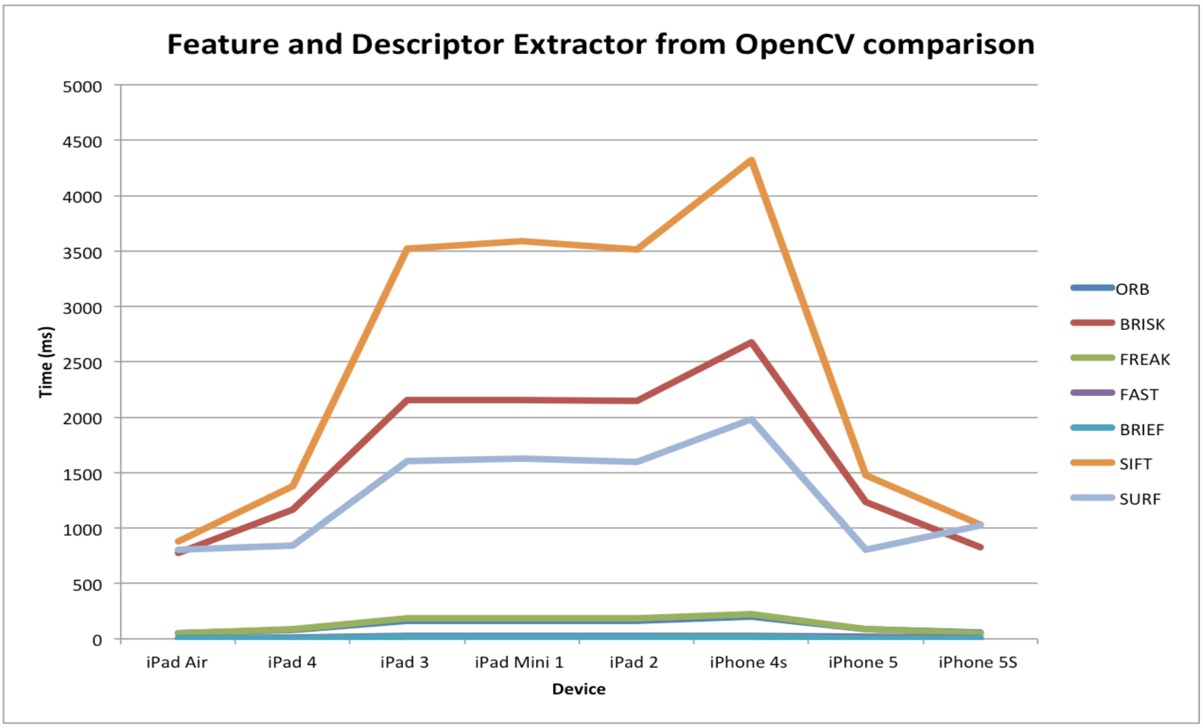


Figure: 1.2.1 Feature extractor and descriptor performance on iOS iPad and iPhones

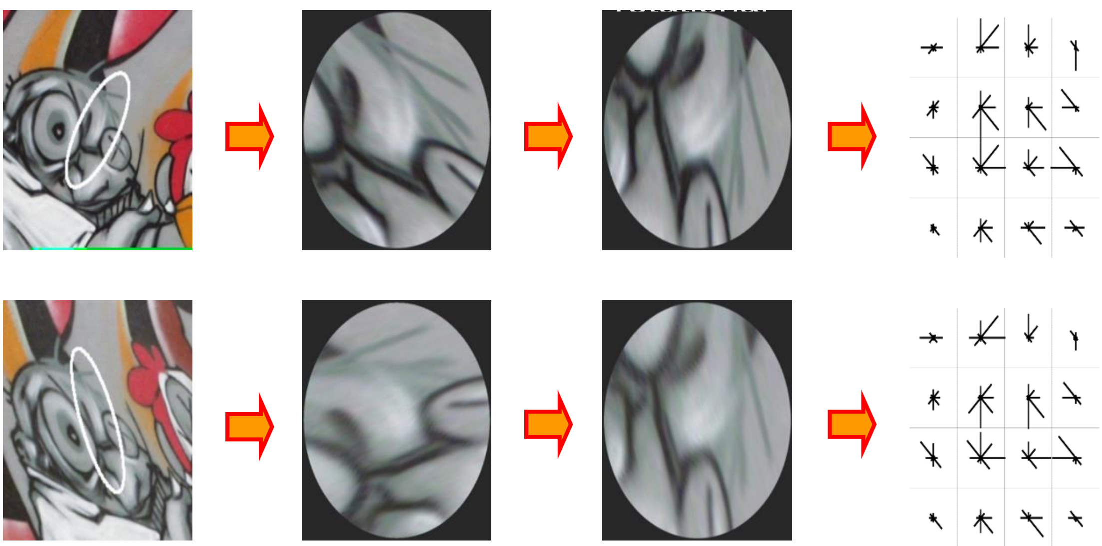


Figure: 1.2.2: Local image Descriptor

1.3 Motivation

In computer vision most of the applications are based on one of the most famous and highly rated descriptor i.e. Scale Invariant Feature Transform (SIFT), it has given all the efficient scale invariant features but there is a scope for the development in terms of rotationally invariance ,the features of SIFT can give rotationally invariant local features but there is some degradation in the amount of spatial information in the encoded data.

Therefore SIFT and other descriptors based on SIFT (i.e. SURF, RIFT) are showing some losses in the information encoded it. The main reason which blocks to achieve rotation invariance is the gradient orientation histogram cells are accumulated in rings around the interest point, some amount of spatial information is lost in this process, leading to degradation of its distinctiveness of the descriptor.

So this drawback of conventional descriptor construction methods leaves the space for improvement in the rotational invariance of the region descriptor leads to the motivation of my work. I choose intensity order local features because it is invariant to rotation and illumination changes .Local features computed from different support regions are linked with their intensity order is experimented and analyzed from the emerging or developing descriptors like intensity based descriptors such as MROGH and MRRID .

1.4 OBJECTIVES:

- Descriptors are mostly used in many image recognition applications and so many methods have been developed during the decade.
- In feature matching process mainly there are three steps: detection, description and matching.
- Even though each and every step in feature matching process, my aim is to design feature descriptor. The descriptions of the local features should be distinctive enough to remove false matches mean while remaining robust to the changes in images like rotation, translation, and noise.
- The method of interest region description is based on intensity order of the local features whereas the local features are calculated using gradients and intensity order.

Chapter 2 Background Theory

2.0 SIFT

SIFT stands for Scale Invariant Feature Transform descriptor this is authored by Prof. David Lowe from University of British Colombia. This is one the most influential work in the field of the computer vision which is having 1500 citations, published in 2004 and heavily used in lots of applications it is key point descriptor, interest point detector [3].

Harris detector is also interest point detector but is not scale invariant [1]. Leading thing in sift is it is scale invariant that means images of different sizes and images of different viewpoints and different depths. When you want to take two images of the same scene taken from a different depths in the scale of the objects , this image is different from other image then we will have problem in finding these interest points so the advantage of this is invariant to scale . this is patented one ,if anybody want to use this can pay to university of British Colombia it also have motivation of what neurons of humans or living beings is doing is the same thing what sift is doing. It change the image into scale invariant coordinates.

2.1 Features of SIFT:

We want to extract distinct invariant features, these features can match from image belongs to the same group. They are invariant to rotation, even though it is not directly taking care of the rotation.

Actually Harris detector is rotationally invariant, that means if we take an image there is an objective to find Harris points. We have to rotate the image with respect to z – axis after that we have to apply Harris detector to get the Harris points exactly same as in the rotated version and also in the original image.

SIFT is not directly a rotationally invariant descriptor but all the steps are pretty robust to rotation.

Affine Transformation: more than rotation it is called affine transformation. In simple transformation there is scale, rotation and then there is a shear and so on.

Change in 3D- view point: As exemplified earlier the biggest problem in image processing or computer vision is the picture of the same scene from different view point is not the same as the original image and it is very different, it is very difficult to match this kind of images and also noise and all those things.

2.2 Advantages:

SIFT feature is a local feature and it is not that sensitive to clutter and occlusion, hence we can say it is robust to clutter and occlusion because some parts are cladded and other parts are not.

It has distinctive features, lot of studies are made on different images. In real time implementation there are many implementations which are very popular applications in the modern world.

1. Locality
2. Distinctiveness
3. Quantity
4. Efficiency

2.2.1 Global feature:

When you take an image , considerer image is large image which is having lots of pixels are there , so you can find the global feature like histogram of gradients (HOG) which will in turn gives the distribution of intensity and distribution of color that is called Global feature

2.2.2 Local feature:

When we look at small neighborhood as we and looking at Harris- detector we are looking at 5*5 and 10*10 neighborhood based on that we can say there is a “local feature “

similarly SIFT is local feature. So there are other things when we keep emphasizing the sift algorithm we can get lots of interest points.

2.3 computation process to generate the set of image features

1. Scale and space extraction.
2. Key point localization.
3. Orientation assignment.
4. Key point description.

Mostly affine invariant regions are described as our interest region because they are invariant to scale and rotation changes as the name affine indicates that it as scale and rotation , hence affine invariant regions are more preferable than the presenting detectors like MSER and other region detectors .

Affine Invariant region detectors such as Hesian-Afiine and Harris-Affine are the most advanced detectors for detecting the interest regions. As Hessian detector detects the region into elliptical shape interest regions and whereas Harris detector detects the region into canonical shape interest regions. Then the detected regions are normalized to circular shape. This Normalization take place by the process of the obtained canonical or elliptical regions on the unit circle.

2.4 Overview on SIFT:

1. Overview of Algorithm
2. Scale Space and Difference of Gaussian
3. Key point Localization
4. Orientation Assignment
5. Descriptor Building

The basic aspect of many problems related to image processing is image matching and also object or scene recognition, solving for 3D structure from multiple images, stereo correspondence, and motion tracking. SIFT is the most efficient algorithm to describes image features that have many properties, that make them suitable for matching differing images of an object or scene.

These features are invariant to changes in image such as scaling and rotation, and partly 3D camera viewpoint and invariant to change in illumination. These are restricted in both the spatial and sequential domains, decreasing the chance of disruption by occlusion, clutter, or noise. Maximum features can be extracted from typical images with efficient methods.

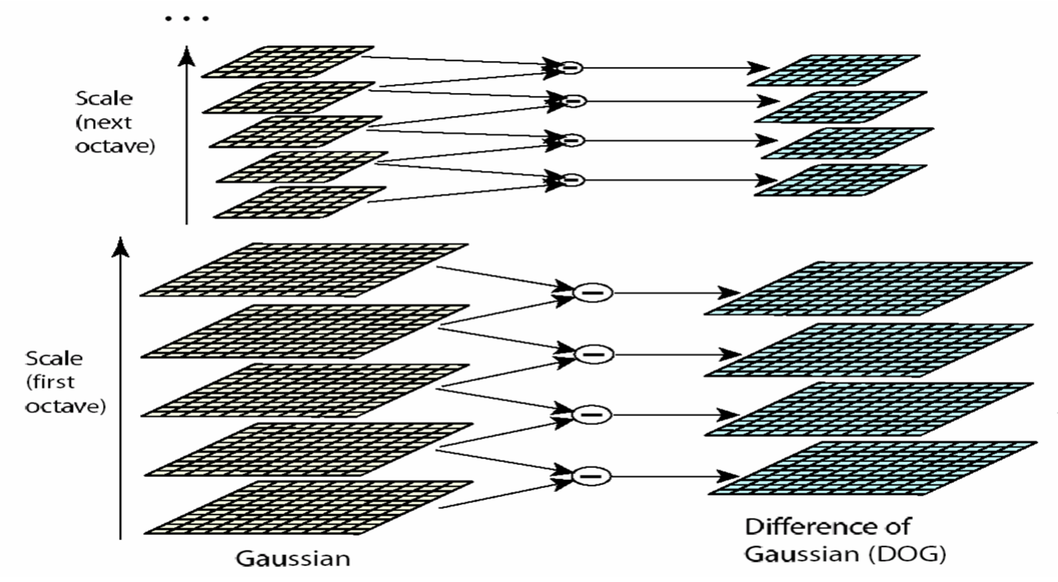
Along with the features which are highly distinctive, which allows a single feature to be correctly matched with high probability against a large database of features, providing a basis for object and scene recognition.

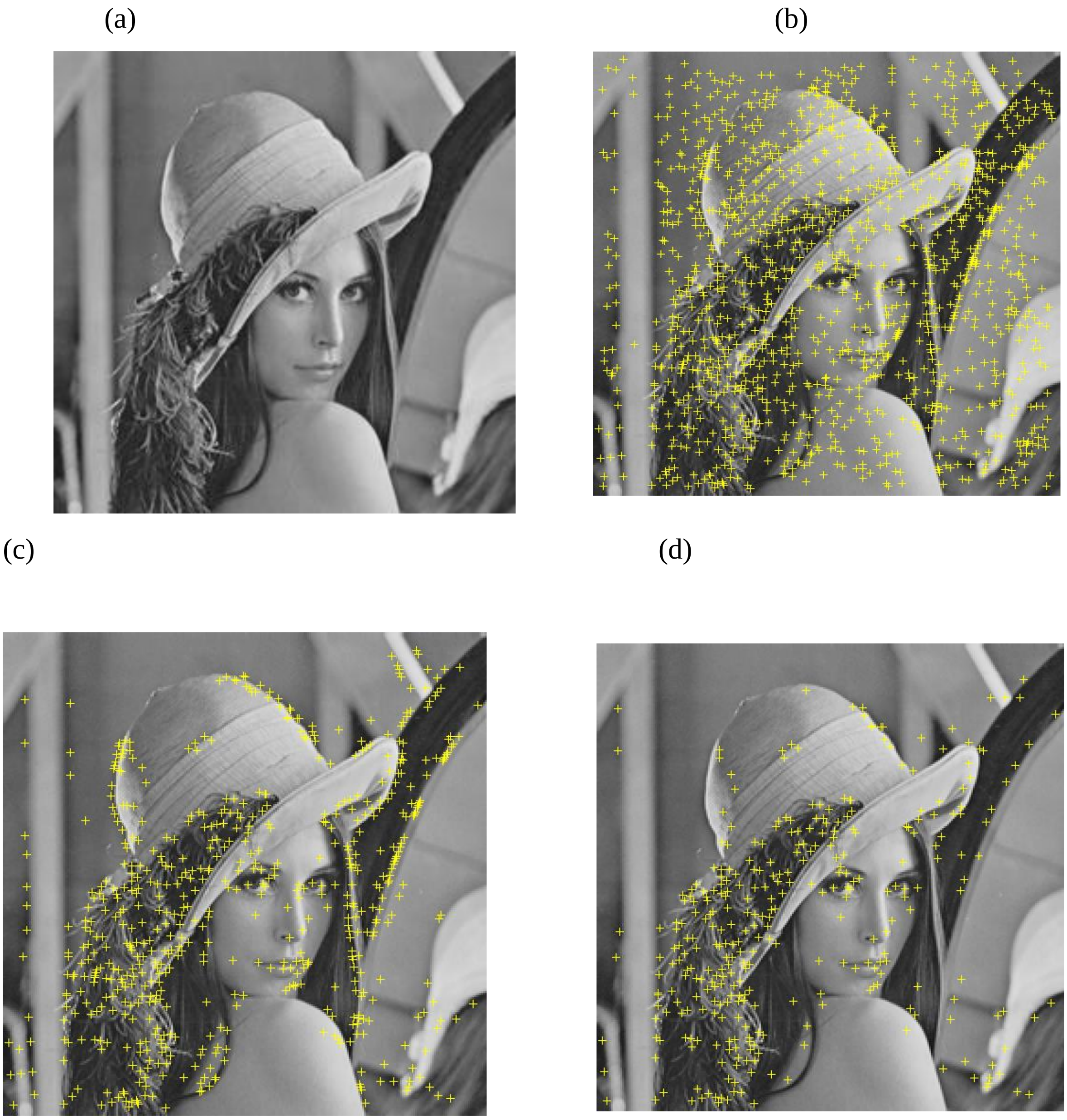
The expense for extracting the features is reduced by incorporating a linking cleaning method, where the costly method are induced at locations that pass an basic test.

2.4.1 Scale-space extreme detection

In the first stage of computation of SIFT[3], it searches over all scales and image locations. Difference-of-Gaussian function is used to determine the potential points for affine invariance.

Figure2.4.2 Scale-space extreme detection





(a) Taken Image (b) DoG extrema points (1284 points) (c) When filtering low contrast extrema with the given threshold 0.05 (505 points) (d) when filtering edge points with then given curvature ratio $r = 20{:}0$ (348 points)

2.4.3 Key point localization

At each sample location, a brief model is used to determine the location and scale. Key points are selected depend on measures of their weight.

2.4.4 Orientation assignment

We can use one or more than one orientations to assign each key point location depending on the directions of the local image gradient. In future every operation will be performed on image to provide invariance to affine transformations for those images which are changed relative to the assigned orientation, scale, and location for each feature.

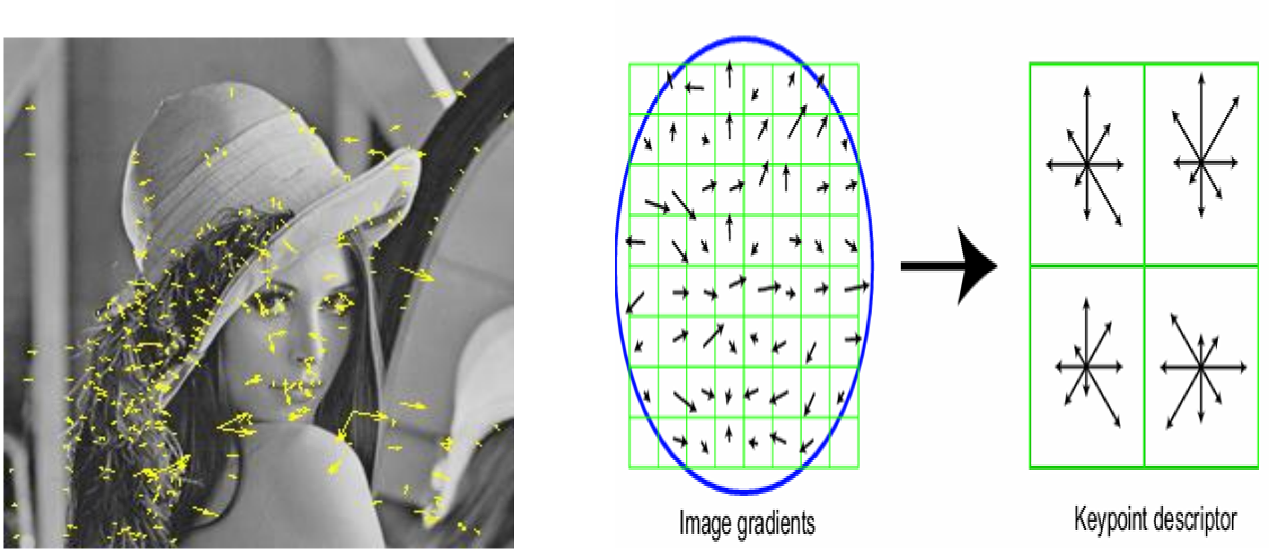
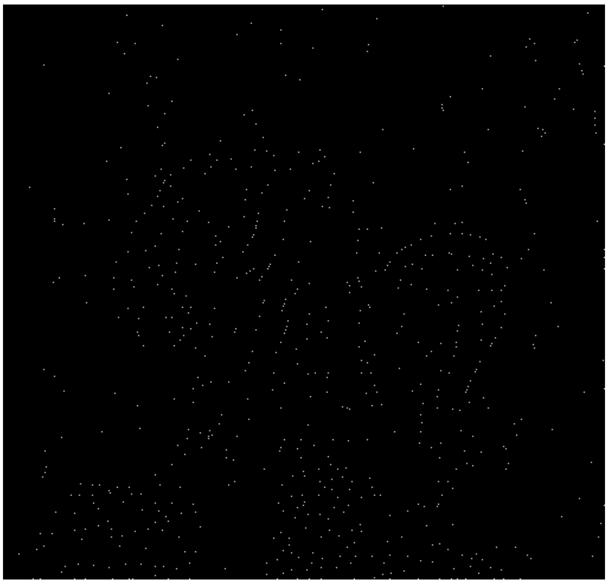


Figure 2.4.4: orientation assignment

2.4.5 Key point descriptor

The local image gradients are calculated with respect to the scale in the region around each key point. These are finally transformed to represent the significant levels of local distortion in shape and illumination changes.

Input image: grey scale image:



Obtained keypoints:

2.4.6 Results of the input image rose using SIFT Descriptor:

y = 256 256

z = 128 128

Time for Gaussian scale space construction: 5.422 s

Time for Differential scale space construction: 0.014 s

Time for finding key points: 0.086 s

Total number of key points extracted are: 675

Time for calculating descriptor: 0.591 s

Chapter 3: Descriptors based on Binary Patterns

3.1 Local Binary Patterns: Local Binary Pattern (LBP) is a modest and very effective texture operator where the labels of the image pixels are threshold by the neighbourhood of each pixel resulting a binary number. LBP give high distinctiveness and robustness to most of the changes in the image hence it became popular in its kind. It can be seen as a unifying approach to the traditionally divergent statistical and structural models of texture analysis. Perhaps the most important property of the LBP operator in real world applications is its robustness to monotonic grey-scale changes caused by illumination variations. Another important property is its computational simplicity, which makes it is possible to analyse images in challenging real-time settings.

The value of LBP code of pixel (x_c, y_c) is given by:

$$LBP = P = \sum_{p=0}^{P-1} s(g_p - g_c)2^p$$

(3.1)

Where,

$$S(x) = \begin{cases} 1, \Rightarrow x \geq 0 \\ 0, \Rightarrow otherwise \end{cases} \tag{3.2}$$

Figure 3.1sample

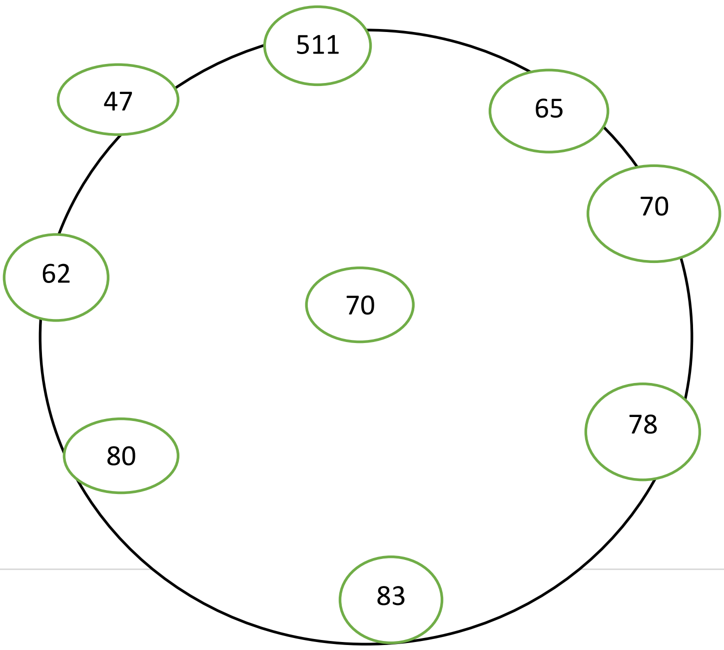
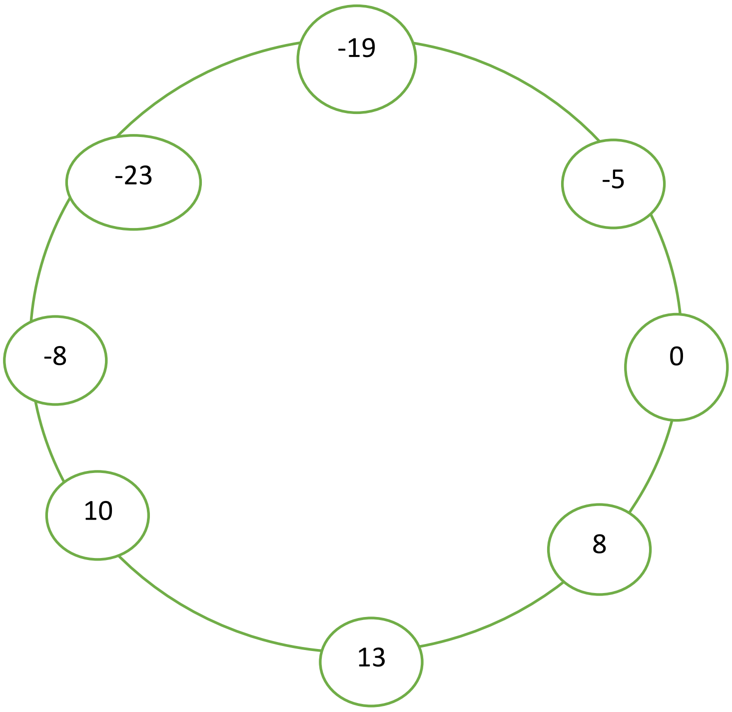


Figure 3.2 difference of pixels



Nc is the center pixel and N1,N2…..N7 are the neighboring pixels , here we compare the center pixel with the each and every pixel and using the threshold we compute binary patterns.

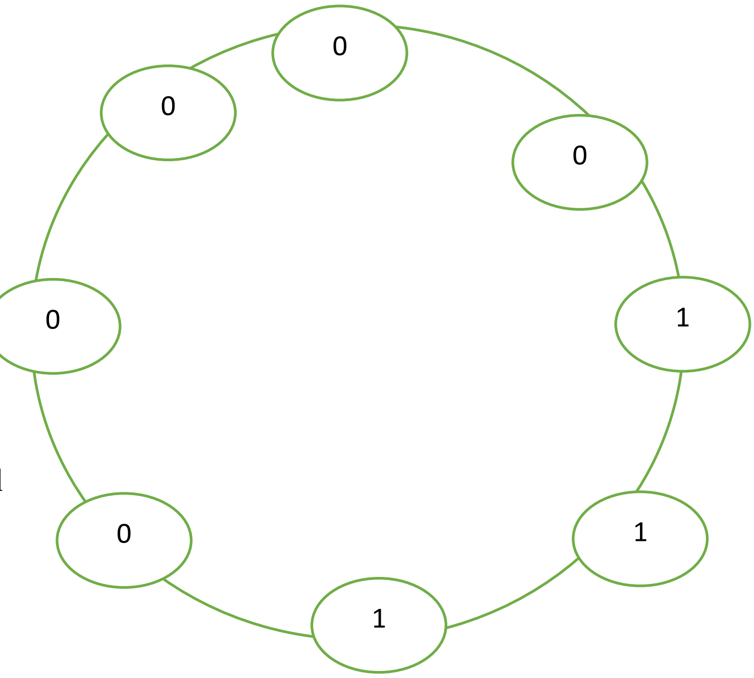


Figure 3.3 threshold

1*1+1*2+1*4+1*8+0*32+0*64+0*128=15

Figure 3.4 binary conversion

78	99	50
54	54	49
57	12	13

threshold

1	1	0
1		0
1	0	0

Binary code: 11000011

3.2 Center Symmetric -Local Binary patterns:

- The LBP[11] operator produces labels for the image rather long histograms and is therefore difficult to use in the context of a region descriptor. To address the problem we modified the scheme of how to compare the pixels in the neighbourhood. Instead of comparing each pixel with the centre pixel, we compare centre-symmetric pairs of pixels .This halves the number of comparisons for the same number of neighbours. We can see that for 8 neighbours, LBP produces 256 (2^8) different binary patterns, whereas for CS-LBP[11]produces only 16 (2^4) different patterns. Furthermore, robustness on flat image regions is obtained by threshold the grey level differences with a small value. Nc is the center pixel and N1,N2…..N7 are the neighboring pixels , here we compare the center pixel with the each and every center symmetric-neighbor pixels and using the threshold we compute binary patterns

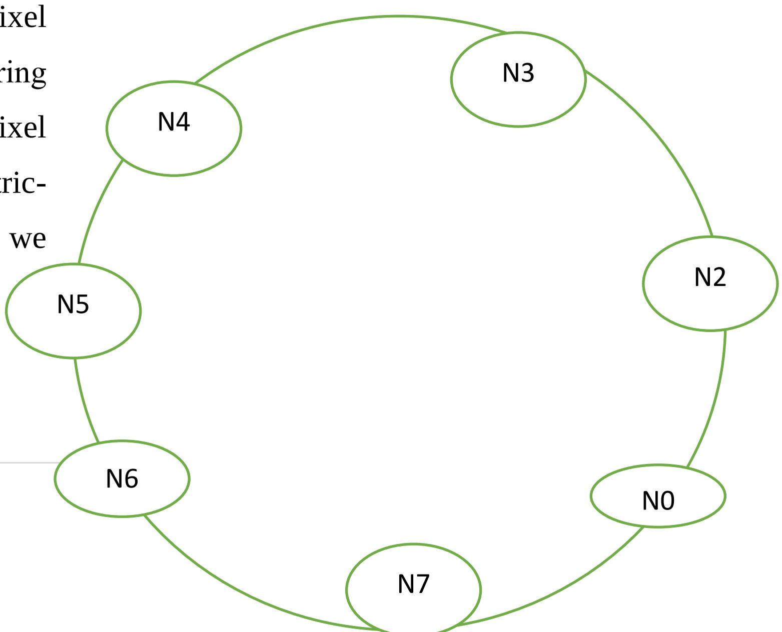


Figure 3.2 Neighborhood:

Nc is the center pixel and N1,N2…..N7 are the neighboring pixels , here we compare the center pixel with the each and every center symmetric- neighbor pixels and using the threshold we compute binary patterns.

LBP=

S(N0-NC)*2^0 +

S(N1-NC)*2^1 +

S(N2-NC)*2^2 +

S(N3-NC)*2^3 +

S(N4-NC)*2^4 +

S(N5-NC)*2^5 +

S(N6-NC)*2^6 +

S(N7-NC)*2^7 +

CS-LBP=

S(N0-NC)*2^0 +

S(N1-NC)*2^1 +

S(N2-NC)*2^2 +

S(N3-NC)*2^3 +

Figure 3.2.2 LBP and CS-LBP Calculation

78	99	50
54	54	49
57	12	13

threshold

1	1	0
1		0
1	0	0

Binary code: 11000

Figure 3.2.3 binary conversion

Chapter 4: Descriptor based on Intensity Order

4.1 THE PROPOSED MROGH:

MROGH stands for Multi support Region Order-Based Gradient Histogram (MROGH). The main idea of this method is to group the calculated rotationally invariant local based on their intensity orders. Here, we don't assign a reference orientation for its rotationally invariant property, instead we use the idea of rotationally invariant-coordinate system in which the obtained local features are inherently invariant to rotation.

Sample points are divided into number of groups using intensity order, after that the obtained local features which are invariant to rotation are grouped altogether in the respective groups using intensity order. Then the local features in different groups of the sample points are group with their intensity order so as to construct the descriptor.

The idea of intensity order is rotationally invariant so the obtained local features based on the intensity order (which is inherently rotation invariant) is also rotationally invariant. Hence there is requirement of reference orientation for its rotational invariant property.

Method for local feature description where local features are grouped with their intensity orders in normalized support regions. Local features arranged according to the intensity orders is invariant to rotation and illumination changes and encodes the information into a descriptor. Local features are calculated in two different ways, one based on gradient using MROGH and other based on intensity using MRRID.

Descriptor based on local features ,calculated from the gradient based method performs well in most of the cases but, in the case of blur and large illuminations changes descriptors based on local features, calculated from the intensity based method performs better. Over all gradient based method gives the better performance than the intensity based method. Experimental results show that the idea of intensity order pooling rules over the descriptors depended on estimated reference orientation for rotational invariance property, as it is inherently rotational invariant without estimating a reference orientation , which results in major error source for the descriptors like SIFT (Scale Invariant Feature Transform) and other famous descriptors like SURF and DAISY

4.2 Affine Normalized Regions:

Mostly affine invariant regions are described as our interest region because they are invariant to scale and rotation changes as the name affine indicates that it as scale and rotation , hence affine invariant regions are more preferable than the presenting detectors like MSER and other region detectors .

Affine Invariant region detectors such as Hesian-Afiine[2] and Harris-Affine[1] these two are the most advanced detectors for detecting the interest regions. As hessian detector detects the region into elliptical shape interest regions and whereas Harris detector detects the region into canonical shape interest regions. And this detected regions are Normalized to Circular shape. This Normalization take place by the process of the obtained canonical or elliptical regions on the unit circle.

In reference to the other local descriptors, in this method also we formulate the descriptor using the normalized region, this normalized region should be in the shape of circular region having radius 20.5 pixels and the normalized region with minimal patch is in size of 41*41 pixels.

If the detected region larger than the normalized region then the detected region is smoothed by the kernel which is of Gaussian (the size ratio of the detected region and the normalized region is set as the standard deviation for the smoothing Gaussian kernel.

Detected region can be represented by a symmetrical matrix $A \in R^{2\times 2}$, for any point X in the region,

It satisfies $$X^T AX \leq 1 \qquad (4.2.1)$$

$A = \frac{1}{c^2} E$, here "E" should be an identity matrix, then only we obtained region which is a circular region and with radius "c"; In case if "E" is not an identity matrix then the detected region may be an elliptical region. Normalization process tries to warp the given detected region into a canonical circular region as shown in Fig. 4. The sample point Xo belonging to the normalized region satisfies

$$X'^T X' \leq r^2$$

(4.2.2)

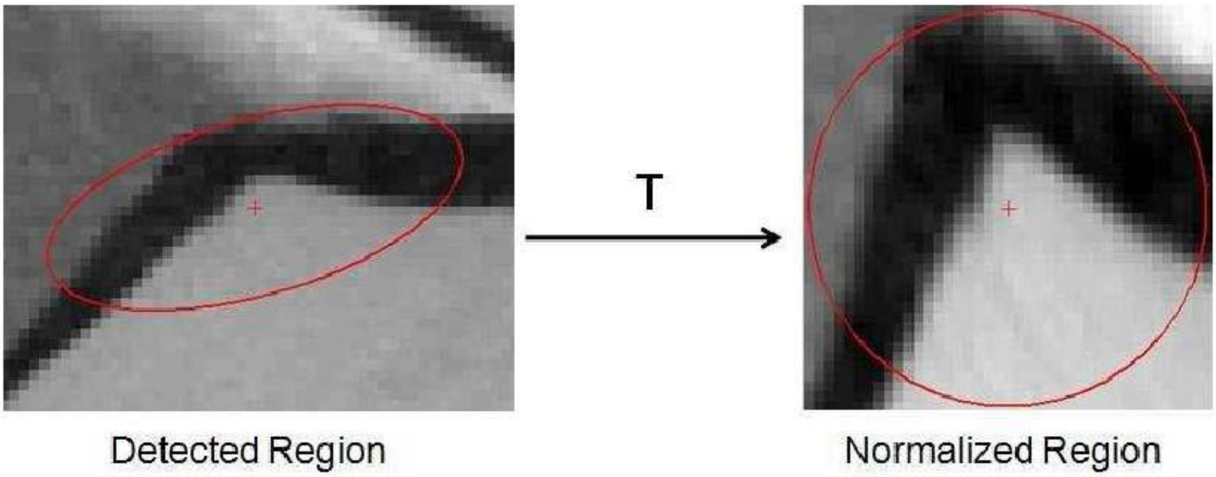


Fig. 4. The affine normalization of a detected region to the canonical
Circular region (normalized region).

Where r is the radius of the normalized region, which is set to 20.5 pixels in this experiment.

Combining (2) and (3),
We have

$$X = \frac{1}{r} A^{-1/2} X' = T^{-1} X' \qquad (4.2.3)$$

Hence, points X in the detected region can be calculated by mapping it to the each sample point X' in the normalized region and considers the intensity of X as the intensity of X' in the normalized region, i.e., I (x') =I(x).
Usually, X is not exactly located at a grid point, so I(x) is obtained by bilinear interpolation. Fig. 4 gives an example of the normalized region.

4.3 Support Region Partition Based on Intensity Orders

In the RIFT [8] we divide the given support region into number of rings and group the calculated local features of the sample points with the already divided rings. In RIFT [8] we do it with the rotation invariant property of the region descriptor. But in this case there is lot of losses in the information encoded which in turn leads to the reduction in the distinctiveness of the obtained descriptor.

We can say, grouping local features circularly gives the rotation invariance the cost of descriptor distinctiveness. Hence, most of the famous methods are dividing the support region into numerable regions so as to achieve the rotation invariance, ex. SIFT[3], DAISY, CS-LBP[11], OSID, and so on.

Unluckily, to be rotationally invariant, there is a need of reference orientation for these sub divided support regions.
Instead of dividing the support region geometrically we opt a different based on intensity order pooling where we divide the sample points into different groups depending upon their intensity order. There is no need of being neighbors for these sample points as they are depending on intensity. Hence there is no need of depending on the estimated orientation because our approach is based on the intensity order which is inherently rotation invariant

R is a support region with n sample points, and I (X_i^j) the intensity of sample pointX_i^j. Our aim is to partition *R* into *k* groups according to the intensity orders of sample points.

$$R = \{X_1, X_2 \ldots\ldots X_n\}$$

(4.3.1)

- finally, set of sorted sample points according to the intensity order in non-descending order is obtained as:

$$\{X_{f(1)}, X_{f(2)}, X_{f(3)} \ldots\ldots X_{f(n)} = I\left(X_{f(1)}\right) \leq I\left(X_{f(2)}\right) \leq \ \ldots\ldots I(X_{f(n)})\ \}$$

(4.3.2)

4.3.1 Gradient based local feature:

Using intensity order we combine gradient supplies in multiple support regions:

- In a support region gradient are computed in locally rotation invariant way.
- In order to encode spatial information The rotation invariant gradients are adaptively combined spatially based on their intensity orders.
- To improve discriminative ability we use multiple support in designing descriptor.

1. Step1: we calculate gradient which is locally rotation invariant for each sample point in the detected region according to Eq. (4.3.1.1) and Eq. (4.3.1.2)
2. Step2: sample points are arranged based on the intensities and they are divided into k segments based on their orders.
3. Step3: gradient calculated are combined organized in each of the k segments.

$$D_y = I(X_i^3) - I(X_i^7) \tag{4.3.1.1}$$

$$D_x = I(X_i^1) - I(X_i^5) \tag{4.3.1.2}$$

$$m(X_i) = \sqrt{D_x(X_i)^2 + D_y(X_i)^2} \tag{4.3.1.3}$$

$$\theta(X_i) = \tan^{-1}\{D_y(X_i)/D_x(X_i)\} \tag{4.3.1.4}$$

$\theta(X_i)$ Is linearly assigned to the two adjacent bins depending on the distances to them weighted by $m(X_i)$.

- When $\alpha(\theta(X_i), dir_j < 2\pi/d$

$$f_j^G = \left\{ m(X_i) \frac{\left(2\pi/d - \alpha(\theta(X_i), dir_j)\right)}{2\pi/d} \right\} \tag{4.3.1.5}$$

- Otherwise $f_j^G = \{0\}$ (4.3.1.6)

$$F_G(X_i) = \{f_1^G, f_2^G \ldots\ldots f_d^G\}\ ,\ here.\ (0{,}2\pi)\ \ is\ divided\ into\ d\ equal\ bins$$

$$as\ \ dir_j = (2\,\pi/d) \times (i-1), \mathrm{i} = 1{,}2{,}3 \ldots\ldots \mathrm{d}$$

4.3.2 Assembling of obtained local features using their intensity order

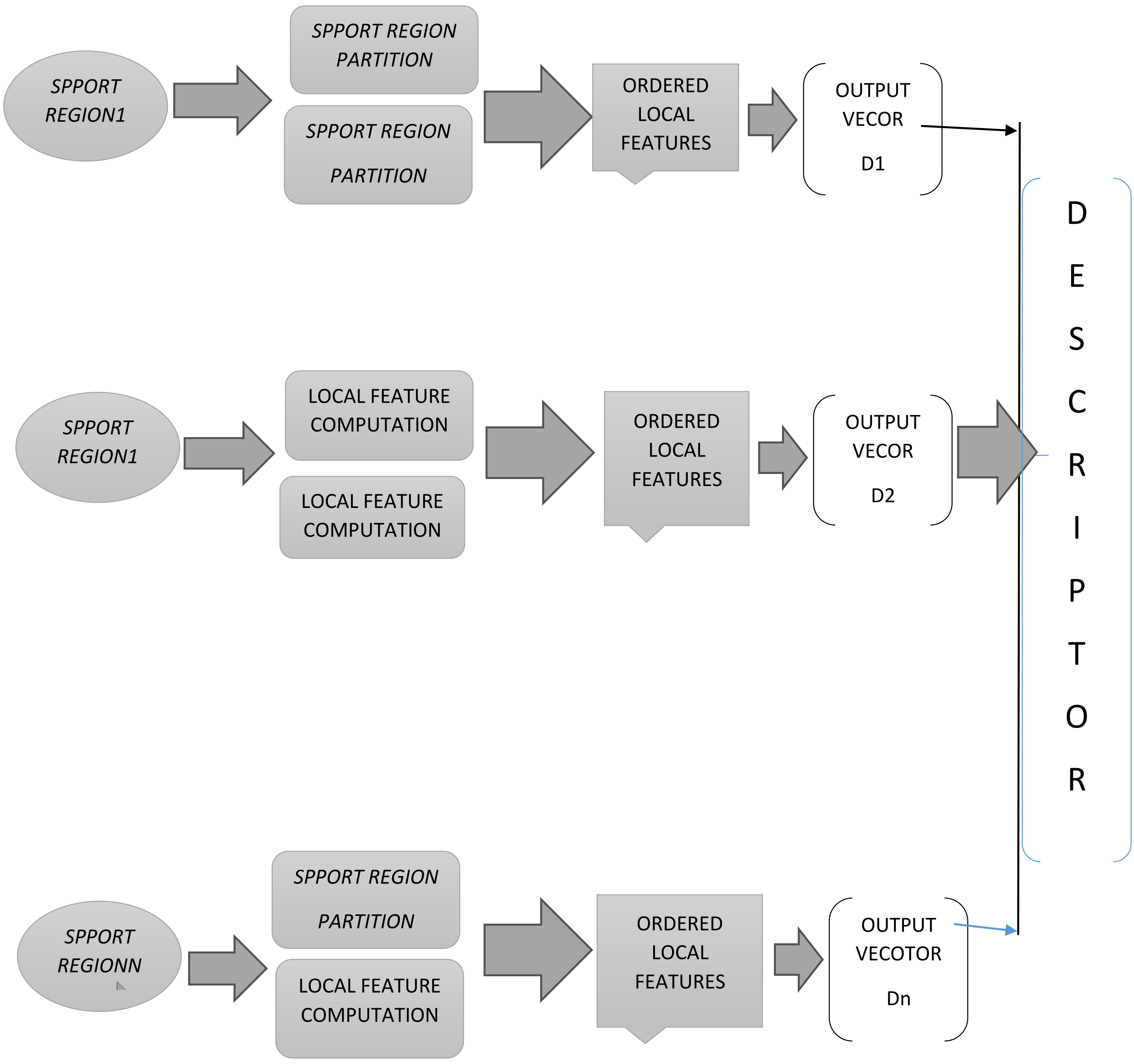
- The local features designed from the sample points are then combined with their intensity orders.as shown in the figure (4.4.1).
- At first they are grouped together to form a vector in each partition which is obtained based on the intensity orders of sample points.
- The gathered vectors of diverse partitions are linked together to represent this support region. We indicate it as

  D(R) = (F(R_1), F(R_2), ・ ・ ・ , F(R_k)) and

- F(R_i) is the collected vector of partition R_i, i.e.,

$$F(R_i) = \sum\nolimits_{X \varepsilon R_i} F_g\ (X) \tag{4.3.2.1}$$

Design of local descriptor:



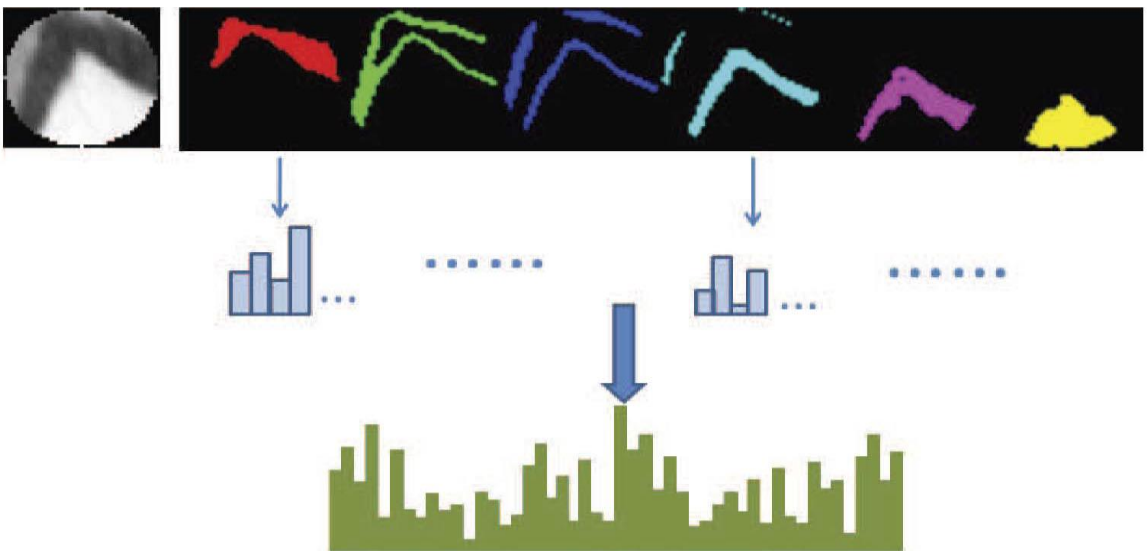


Figure 4.4.1 Process of assembling local features depending on their intensity orders for a support region

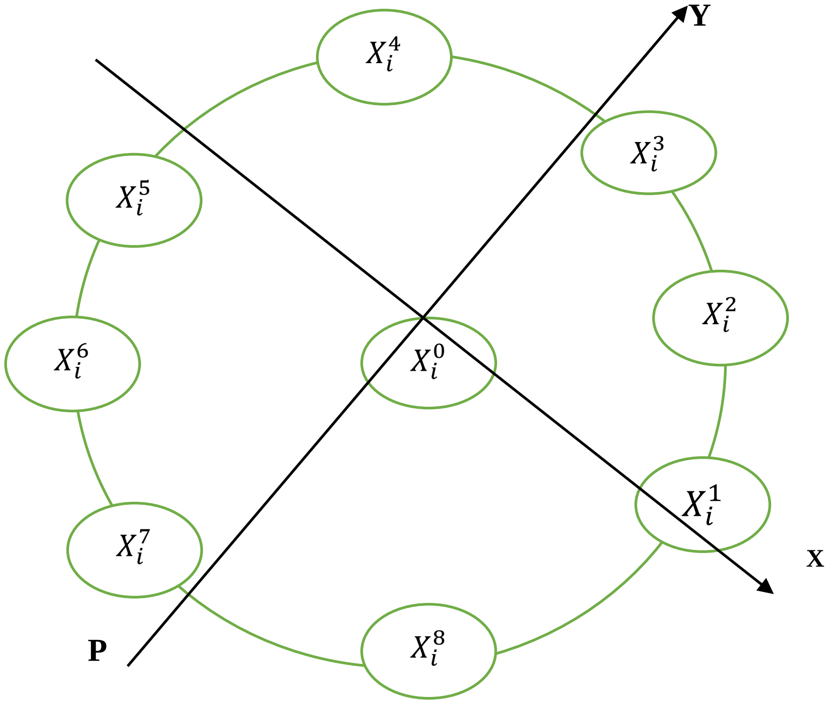
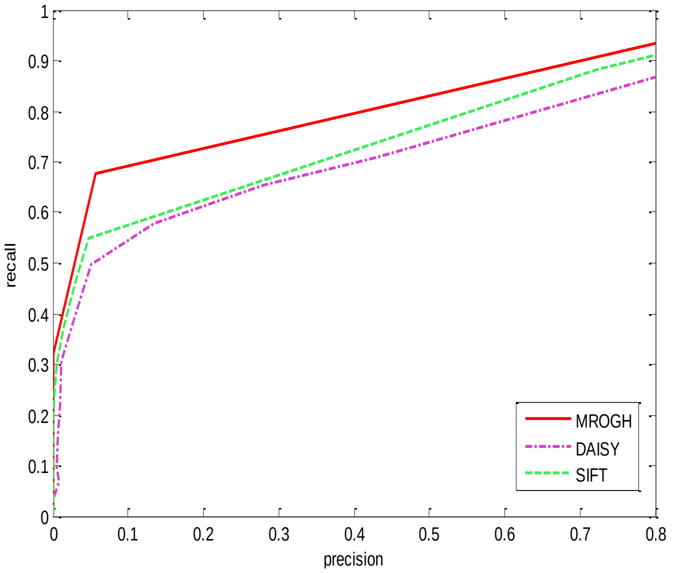


Figure. 4.4.2
The coordinate system which is used to compute rotation invariant local features locally of sample point X_i **.** P is taken as interest point which is the centre of the normalized region.

Chapter 5

Results and analysis:

Figure 5.1 Rotational changes:



This above comparison shows that use of our proposed method shows better results than the other descriptors like DAISY and SIFT [3].

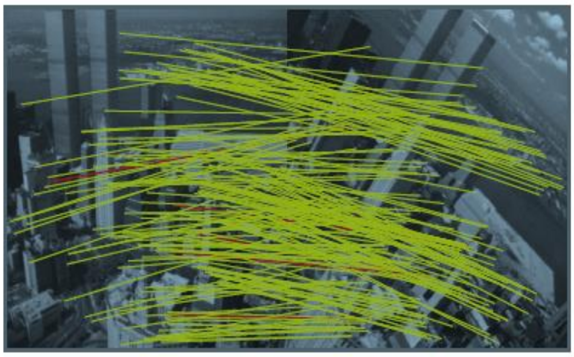
MROGH-Multi Region Order based Gradient Histogram

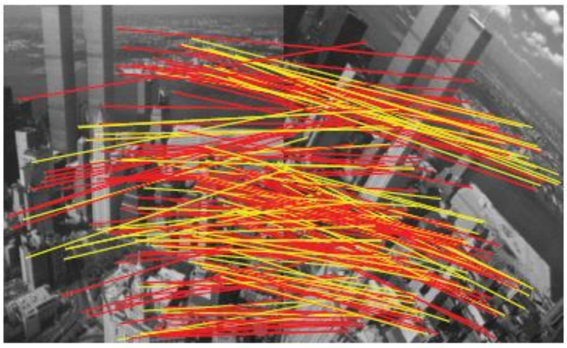
SIFT-Scale Invariant Feature Transform

RIFT-Rotationally Invariant Feature Transform

5.2 Showing matching points for SIFT and MROGH descriptor:

SIFT MROGH



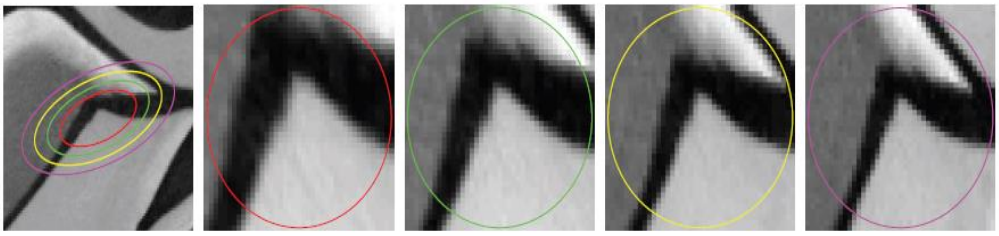


Results obtained from matching corresponding points in the image which are large orientation estimation errors i.e. ($\geq 20^0$) (a) by SIFT and (b) by our proposed descriptor. In the image red lines are the corresponding points that are matched by their descriptors are marked with, while yellow lines indicate those corresponding points that are unmatchable by their descriptors. Maximum of these corresponding points can be correctly matched by our proposed descriptor (MROGH)[5], but limited of them can be correctly matched by SIFT[3].

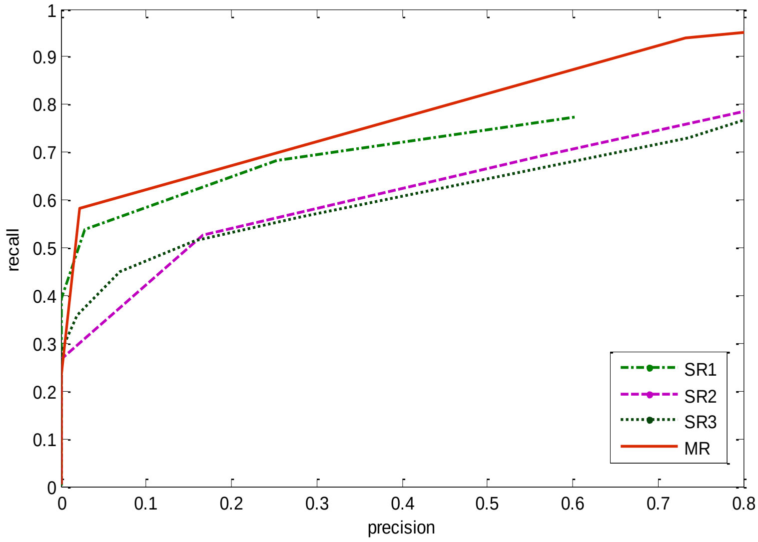
Parameter settings for the descriptors:

Type	Denotation	Parameters setting	depiction
MROGH	K	4,6,8	No. of Partitions
	D	4,8	No. of Orientation bins
	N	1,2,3,4	No. of Support regions

Comparison between single support region and multiple support region



(a) (b) (c) (d) (e)



(a)MR-Multi Support Region

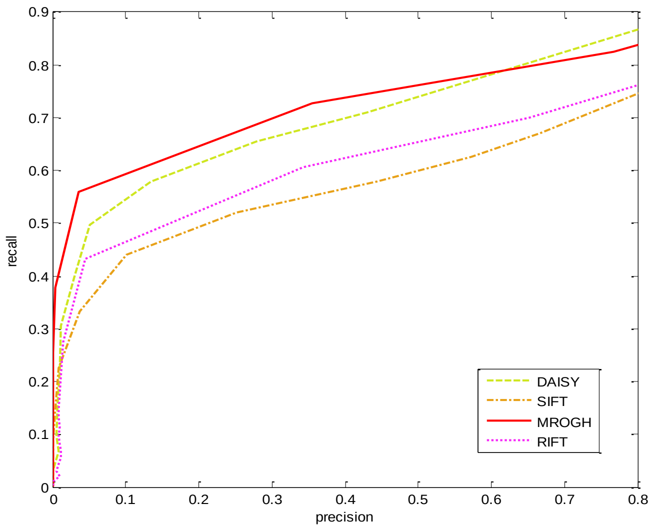
(b)SR1-Single Support Region 1

(c)SR2-Single Support Region2

(d)SR3-Single Support Region3

This above comparison shows that use of multiple support regions are giving better results than the single support region ,it also depicts Single support region is showing much better performance than the SIFT[3]. Multiple support regions to overcome mismatching proble

Performance comparison on illumination changes:



Performance on illumination changes:

MROGH-Multi Region Order based Gradient Histogram,

SIFT-Scale Invariant Feature Transform

RIFT-Rotationally Invariant Feature Transform

This above comparison shows that use of our proposed method shows better results than the other descriptors like DAISY ,RIFT and SIFT [3].

Datasets used for measuring accuracy of the descriptors:

Measure of accuracy on various datasets:

Dataset	SIFT	DAISY	MROGH
53 objects	53.02%	62.4%	**72.8%**
ZudBuD	73.07%	84.1%	**88.8%**
Kentucky	48.9%	58.3%	**75.1%**

The above table show the performance comparison of the MROGH over other available descriptors like SIFT and DAISY. In all the datasets the taken datasets ZudBuD dataset shows the highest percentage of accuracy as 88.8% and remaining datasets like 53 objects and Kentucky shows 72.8% and 75.1% respectively.

Conclusion and future scope:

In this work, local features are described and studies using intensity order method where local features are grouped with their intensity order of the given samples points. By using the idea of the intensity order pooling, we can achieve rotational and illumination invariance in the features. Local features calculated from the gradients can with stand all the challenges like rotation, scale, viewpoint changes. In the case of blur and large illuminations local features calculated from the intensity do better. Over all gradient based method gives the better performance than the intensity based method.

Here we have used multiple support regions to overcome mismatching problem, even though single support region is better than SIFT [3]. As the descriptors developed on estimated reference orientation are the major error source for the descriptors like SIFT, by using the idea of the of intensity order we can overcome this problem.

Mainly errors in many descriptors are due to depending on reference orientation for rotational invariance property. In comparison with those descriptors the proposed method is very stable as it doesn't depend on the reference orientation, which is the cause of major error for false negatives.

Descriptors are used in almost all basic computer vision applications like image matching, object recognition, object tracking etc. As our method shows it is more efficient than SIFT descriptor, hence it is more convenient to use this descriptor to get much more fruitful results.

- Local Image descriptors are computed by the process of rotation invariant method in spite of depending on the reference orientation.
- Errors in assigning reference orientation are the main cause of the false negatives.
- We use intensity order based local features for support region partition , the divided sample point partitions are invariant to illumination changes
- In comparison with other descriptors, our method is more stable.

References:

[1] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L.V. Gool, "A Comparison of Affine Region Detectors," Int'l J. Computer Vision, vol. 65, nos. 1/2, pp. 43- 72, 2005.

[2] 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.

[3] D. G. Lowe. Distinctive image features from scale-invariant key points. *IJCV*, 60:91–110, 2004.

[4] B. Fan, F. C. Wu, and Z. Y. Hu. Aggregating gradient distributions into intensity orders: A novel local image descriptor. In *Proc. CVPR*, pages 2377–2384, 2011..

[5] B. Fan, F. Wu, and Z. Hu, "Rotationally invariant descriptors using intensity order pooling," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 10, pp. 2031–2045, Oct. 2012.

[6] http://www.robots.ox.ac.uk/ vgg/research/affine/.

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

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

[9] E. Tola, V. Lepetit, and P. Fua, "A Fast Local Descriptor for Dense Matching," Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2008.

[10] 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.

[11] M. Heikkila, M. Pietikainen, and C. Schmid, "Description of interest regions with local binary patterns," *Pattern Recognition*, vol. 42, pp. 425–436, 2009.