

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE
ENGINEERING AND TECHNOLOGY

**R-LZMF:
ROBUST LOCAL ZERNIKE MOMENT BASED FEATURES**

M.Sc. THESIS

Gökhan ÖZBULAK

Department of Computer Engineering

Computer Engineering Programme

JANUARY 2015

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE
ENGINEERING AND TECHNOLOGY

**R-LZMF:
ROBUST LOCAL ZERNIKE MOMENT BASED FEATURES**

M.Sc. THESIS

**Gökhan ÖZBULAK
(504101506)**

Department of Computer Engineering

Computer Engineering Programme

Thesis Advisor: Prof. Dr. Muhittin GÖKMEN

JANUARY 2015

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

**G-YZMÖ:
GÜRBÜZ YEREL ZERNİKE MOMENT TABANLI ÖZELLİKLER**

YÜKSEK LİSANS TEZİ

**Gökhan ÖZBULAK
(504101506)**

Bilgisayar Mühendisliği Anabilim Dalı

Bilgisayar Mühendisliği Programı

Tez Danışmanı: Prof. Dr. Muhittin GÖKMEN

OCAK 2015

To my family and grandmother,

FOREWORD

I would like to express my gratitude to my thesis advisor Prof. Dr. Muhittin Gökmen for his interest and encouragement during the development of this thesis. His critical guidance on my research made this thesis possible.

I would also like to thank to my family and grandmother, who have been there and supported me for years with believing my journey on this thesis to be successfully concluded.

January 2015

Gökhan ÖZBULAK
Computer Engineer

TABLE OF CONTENTS

	<u>Page</u>
FOREWORD	ix
TABLE OF CONTENTS	xi
ABBREVIATIONS	xiii
LIST OF TABLES	xv
LIST OF FIGURES	xvii
SUMMARY	xix
ÖZET	xxi
1. INTRODUCTION	1
1.1 Literature Review	3
1.1.1 Interest point detection	3
1.1.2 Zernike moments.....	6
1.2 Organization of the Thesis	7
2. AN OVERVIEW OF IMAGE MOMENTS	9
2.1 Image Moments	9
2.1.1 Geometric moments	10
2.1.2 Complex moments	10
2.1.3 Orthogonal moments	11
2.2 Zernike Moments	12
2.2.1 Global Zernike moments.....	12
2.2.2 Local Zernike moments	13
3. ROBUST INTEREST POINT DETECTION	15
3.1 Principles of Robust Interest Point Detector Design.....	15
3.2 Proposed Interest Point Detection Algorithms: LZMF and R-LZMF.....	16
3.2.1 Normalization.....	16
3.2.2 Corner detection	18
3.2.3 Non maximum suppression.....	23
3.2.4 Scale-space	25
3.2.5 Parameter settings	27
4. EXPERIMENTAL RESULTS	31
4.1 Evaluation Criteria	31
4.2 Dataset	32
4.3 Results	33
5. CONCLUSIONS AND FUTURE WORK	41
5.1 Conclusions	41
5.2 Future Work	42
REFERENCES	45
CURRICULUM VITAE	47

ABBREVIATIONS

LoG	: Laplacian-of-Gaussian
DoG	: Difference-of-Gaussian
NMS	: Non Maximum Suppression
NCC	: Normalized Cross-Correlation
SIFT	: Scale Invariant Feature Transform
SURF	: Speeded-Up Robust Features
LZM	: Local Zernike Moment
LZMF	: Local Zernike Moment based Features
R-LZMF	: Robust Local Zernike Moment based Features
SSD	: Sum of Squared Distance
FAST	: Features from Accelerated Segment Test
ORB	: Oriented FAST and Rotated Brief
CenSurE	: Center Surround Extrema
BRISK	: Binary Robust Invariant Scalable Keypoints
FoV	: Field of View
LBP	: Local Binary Patterns
GPU	: Graphics Processing Unit
CUDA	: Compute Unified Device Architecture
HoG	: Histogram of Oriented Gradients

LIST OF TABLES

	<u>Page</u>
Table 3.1 : Parameter settings for LZMF and R-LZMF	29
Table 4.1 : Image sets in “Rotation” sequence.....	32
Table 4.2 : Image sets in “Zoom” sequence.....	33
Table 4.3 : Image sets in “Zoom&Rotation” sequence.....	33
Table 4.4 : Numbers of keypoints detected by LZMF and others in Marseil set.....	35
Table 4.5 : Numbers of keypoints detected by R-LZMF and others in Asterix set ..	39

LIST OF FIGURES

	<u>Page</u>
Figure 2.1 : Visual representations of Zernike polynomials of order 0 to 5 [23]	13
Figure 3.1 : Zernike moments used for corner detection [23]	18
Figure 3.2 : Gray-level corner models fitting onto the unit circle [15].....	19
Figure 3.3 : Detection results of LZMF with synthetic corner images	22
Figure 3.4 : Detection results of LZMF with checkerboard images	23
Figure 3.5 : Detection results of LZMF with real floor images.....	23
Figure 3.6 : Detection results of LZMF with real wall images.....	24
Figure 3.7 : Shifting window of Harris corner detector on the corners [22].....	25
Figure 3.8 : Detection results of R-LZMF with checkerboard images	27
Figure 3.9 : Parameter evaluation for scale-space based on average repeatability...	29
Figure 4.1 : Performance comparison of LZMF with other detectors	34
Figure 4.2 : Average repeatabilities of LZMF and other detectors	35
Figure 4.3 : Some detection results of LZMF on Monet image set	36
Figure 4.4 : Performance comparison of R-LZMF with other detectors	37
Figure 4.5 : Average repeatabilities of R-LZMF and other detectors.....	38
Figure 4.6 : Some detection results of R-LZMF on Crolles image set	39
Figure 4.7 : Some detection results of R-LZMF on Laptop image set	40

**R-LZMF:
ROBUST LOCAL ZERNIKE MOMENT BASED FEATURES**

SUMMARY

Feature extraction techniques are widely used in computer vision in order to transform image data represented as a set of pixels, which mostly consists of redundant information, into a set of features, which has meaningful information for specific problem, for further processing. In this sense, feature extraction can be considered as a dimensionality reduction approach.

Extracted features from the image should be meaningful and relevant to the domain where they are used. They should also have discriminative characteristic in order to specify special structures in the image such as corners, edges and blobs. These kind of structures are used in computer vision to detect the objects in interest, recognize the faces or match the corresponding regions between images.

In computer vision, image matching is one of the fundamental problems where feature extraction comes into play. This problem, which is also named as correspondence problem, requires matching corresponding regions in the images that are taken from same scene with different points of view. Correspondence matching is very crucial when the depth of the object is evaluated in a stereo system, 3D structure of the scene is constructed from the images taken from same scene or motion of the object needs to be tracked in temporal domain. One of the optimum solutions for these kind of problems is to extract interest points/keypoints as features from the images and use them for further processing.

An interest point is an image point that represents distinctive characteristic in that location and its local neighborhood. It should also be detected repeatedly in the images that are geometrically and photometrically transformed forms of each other. In other words, an interest point detected in one image taken from a scene should also exist in another image taken from same scene with a different point of view. In general, corner or blob detectors with some other mechanisms comprise the interest point detection algorithm so that an interest point detector can be characterized as corner or blob based. A robust interest point detector should be invariant to geometric transformations such as scale, rotation, translation and photometric transformations such as illumination. It shouldn't also be affected from background clutter and occlusion in the images.

Designing an interest point detection algorithm is an active topic in computer vision. There are so many studies on corner or blob based interest point detectors. SIFT and SURF are maybe the best-known and used interest point detectors in the literature. Both detectors provide blob based interest point detection mechanism and are very successful in terms of repeatability and distinctiveness. One other successful detector is Harris-Laplace detector and it's based on Harris corner detector. All these detectors are widely used in the literature for feature extraction and performance comparison.

In this thesis, a novel rotation and translation invariant local Zernike moment based interest point detection algorithm is presented and named as Local Zernike Moment based Features (LZMF). LZMF is then extended to have scale-invariant characteristic by constructing image pyramid in scale-space. Final detector is scale, rotation and translation invariant, and also robust to background clutter and occlusion. This final detector is named as Robust Local Zernike Moment based Features or R-LZMF shortly. R-LZMF is a corner based interest point detector and uses local Zernike moments as convolutional operators in order to detect corners in spatial-space. In this way, descriptive power of Zernike moments is utilized in local sense by applying them to the image pixels and thus structure of corners can be successfully exposed. Here, the critical decision is about which order of Zernike moment should be used for corner detection and it's also investigated in this study.

Performance of proposed interest point detection algorithms, LZMF and R-LZMF, are evaluated on the Inria Dataset by using repeatability score, which is the main criterion for detector accuracy, and the performance of proposed algorithms is compared to well known interest point detectors such as Harris, SIFT, SURF, CenSurE, BRISK for LZMF and SIFT, SURF, CenSurE, ORB, BRISK for R-LZMF. Evaluation results on "Rotation", "Zoom" and "Zoom&Rotation" sequences of the Inria Dataset show that LZMF and R-LZMF outperform almost all interest point detectors to be compared in terms of repeatability score. Distinctiveness performance of LZMF and R-LZMF are also presented by applying the detectors on to the synthetic and real images that contain corner points.

R-LZMF interest point detection algorithm is expected to be partly invariant to affine transformations because it's invariant to scale, rotation and translation and an affine transformation may be considered as a combination of these transformations. However this should be verified as a future work. LZMF and R-LZMF can also be extended to have a descriptor. This makes them a complete schema, which includes both detector and descriptor, as in SIFT and SURF.

G-YZMÖ: GÜRBÜZ YEREL ZERNİKE MOMENT TABANLI ÖZELLİKLER

ÖZET

Öznitelik çıkarma teknikleri bilgisayarla görmede yaygın bir şekilde kullanılmaktadır. Bu teknikler ile hedeflenen çoğunlukla artıklı bilgi içeren ve piksel kümesi şeklinde ifade edilen imge verisini üzerinde çalışılan problem için anlamlı hale getirecek öznitelik kümesine dönüştürmektir. Böylece, gerçekleştirilen dönüşüm neticesinde elde edilen bu öznitelik kümesi daha ileri işlemler için kullanılabilir, anlamlı ve daha düşük boyutta bir biçime dönüşmüş olur. Bu anlamda, öznitelik çıkarma işlemini bir boyut indirgeme yaklaşımı olarak görmek mümkündür.

İmgeden çıkarılan öznitelikler, kullanılacakları alan için anlamlı ve alakalı olmalıdır. Ayrıca, imgedeki özel yapıları niteleyebilen ayırt edici bir karakteristiğe de sahip olmalıdır. İmgedeki bu özel yapılara köşe, kenar ve nokta bulutu örnek olarak verilebilir. Bu tip özel yapılar bilgisayarla görmede ilgilenilen nesnelere saptarken, yüzleri tanıırken ya da resimler arasında karşılıklı bölgeleri eşleştirirken sıklıkla kullanılmaktadır.

Bilgisayarla görmede; imge eşleştirme, öznitelik çıkarmanın işin içine girdiği temel problemlerden biridir. Aynı zamanda, karşılık problemi olarak isimlendirilen bu problemde aynı sahneden değişen bakış açılarıyla alınmış olan imgelerin karşılık gelen bölgelerinin doğru ve tam bir şekilde eşleştirilmesi beklenir. Karşılık eşleştirme; bir stereo sistemde ilgilenilen nesnenin derinliği ölçülürken, aynı sahneden elde edilen imgeler ile sahnenin 3D yapısı çıkartılırken ya da bir nesnenin hareketi zamansal alanda takip edilirken oldukça kritiktir. Bu tip problemlere getirilen en uygun çözümlerden biri ele alınan imgelerden öznitelik olarak ilgi noktaları/anahtar noktalar çıkartmak ve bu noktaları ileri işlemler için kullanmaktır.

İlgi noktası, imgede belirli bir bölgede ayırt edici bir karakteristik gösteren bir imge noktasıdır. Bu ilgi noktasının, birbirinin geometrik ve fotometrik dönüşümü olan resimlerin hepsinde tekrar ederek saptanabilmesi gerekmektedir. Bir başka deyişle, belli bir sahneden elde edilen bir imgede saptanmış bir ilgi noktasının aynı sahneden değişik bir bakış açısıyla elde edilmiş başka bir imgede de saptanabilmesi beklenmektedir. Genellikle, köşe veya nokta bulutu saptayıcılar bazı diğer mekanizmalar ile birleştirilerek ilgi noktası saptayıcı şemasını oluşturmaktadır. Dolayısıyla, bir ilgi noktası saptayıcı köşe veya bulut noktası tabanlı olarak karakterize edilebilmektedir. Gürbüz bir ilgi noktası saptayıcıdan beklenen ise ölçek, döndürme, öteleme gibi geometrik dönüşümlere ve ışıklandırma gibi fotometrik dönüşümlere karşı değişimsiz olmasıdır. Ayrıca, gürbüz bir ilgi noktası saptayıcının arkaplan karışıklığına ve engellere karşı da dayanıklı olması beklenmektedir.

İlgi noktası saptayıcı algoritmaların tasarımı bilgisayarla görmede aktif bir konu olarak araştırılmaya devam etmektedir ve bu konuda çok sayıda köşe ve bulut noktası tabanlı ilgi noktası saptayıcı sunulmuştur. SIFT ve SURF belki de en iyi bilinen ve en çok kullanılan ilgi noktası saptayıcılar olarak gösterilebilir. Her iki ilgi

noktası saptayıcı da nokta bulutu tabanlı bir saptama gerçekleştirmekte, yinelenebilirlik ve ayırt edicilik bakımından oldukça başarılı sonuçlar vermektedir. Bir başka başarılı ilgi noktası saptayıcı ise Harris köşe saptayıcısından yararlanan Harris-Laplace ilgi noktası saptayıcıdır. Bütün bu ilgi noktası saptayıcılar, literatürde, öznelilik çıkarma ve performans karşılaştırması için sıklıkla kullanılmaktadır.

Bilgisayarla görmede; resim momentleri, sıklıkla kullanılan bir başka yaklaşımdır. Bir resim momenti, imgeyi bir fonksiyon gibi ele alan ve imgedeki pikselleri sayıl bir nicele dönüştüren bir izdüşümdür. Bu izdüşüm, bir imge fonksiyonunun polinomsal tabana dönüşümünü sağlayarak imgedeki yapılardan anlamlı öznelilikler çıkartılabilmesine imkan verir. İkili bir imgede alan hesaplaması veya imgenin merkez koordinatlarının bulunması resim momentlerine verilebilecek bazı örneklerdir. Geometrik, karmaşık ve dikgen olarak sınıflandırılan imge momentleri; karakter tanıma, imge geri çatılma ve yüz tanıma gibi problemlere başarılı bir şekilde uygulanabilmektedir.

Son yıllarda ivme kazanan imge momentlerinden biri de Zernike momentleridir. Zernike polinomlarını polinomsal taban olarak kullanan bu dikgen momentler global ve yerel olarak imgeye uygulanabilmekte ve uygulandığı imge bölgesinin bir birim çember üzerinde tanımlanmış Zernike polinomları üzerine izdüşümlerini çıkartabilmektedir. Zernike momentleri, tüm imgeye uygulandığında imgenin kendisini, yerel olarak bir piksel etrafında uygulandığında ise o piksel ve çembersel komşuluğundaki bölgeyi sayıl bir niceliğe dönüştürür. Bu nicelik, Zernike polinomlarının doğası gereği, döndürmeye karşı değişmezdir.

Bu tez çalışmasında, özgün bir döndürme ve ötelemeye değişimsiz Yerel Zernike Moment (YZM) tabanlı ilgi noktası saptama algoritması sunulmaktadır. Sunulan bu algoritma Yerel Zernike Moment tabanlı Özellikler (YZMÖ) olarak isimlendirilmiştir. Aynı çalışmada, YZMÖ, ölçek değişimsiz karakteristiğe sahip olması için ölçek-uzayında imge piramidi oluşturmak suretiyle genişletilmiştir. Sonuçta elde edilen ilgi noktası saptayıcı ölçek, döndürme ve ötelemeye değişimsiz olmakla birlikte arkaplan karışıklığına ve engellere karşı sağlam bir karakteristik göstermektedir. Elde edilen bu ilgi noktası saptayıcı Gürbüz Yerel Zernike Moment tabanlı Özellikler (G-YZMÖ) olarak isimlendirilmiştir. G-YZMÖ, köşe tabanlı bir ilgi noktası saptayıcıdır ve uzamsal-uzayda köşeleri saptamak için YZM'leri evrişimsel işlemler şeklinde kullanır. Bu sayede, köşe noktalarının yapılarını başarılı bir şekilde ortaya çıkarmak için Zernike momentleri imge piksellerine yerel olarak uygulanarak bu momentlerin betimlemesal gücünden faydalanılmış olunur. Burada, kritik olan köşe saptama için kaçınıcı dereceden Zernike momentleriyle çalışılacağıdır ve bu tez çalışmasında uygun moment derecesi de araştırılmıştır. Önerilen algoritmalarla, uzamsal-uzayda ilgi noktası saptama işlemi, temel olarak şu adımlarla gerçekleştirilir: Giriş imge, gri tonlu imgeye dönüştürüldükten sonra ışık değişimlerinin etkisini minimuma indirmek için önce tüm imgede sonra da YZM'lerin evrişimsel işlemler olarak uygulanacağı çember şeklindeki yerel imge bölgesinde normalizasyona tabi tutulur. Daha sonra, uygun evrişimsel işlemler vasıtasıyla her bir piksel için Zernike tepki haritaları çıkartılır ve bu harita önceden belirlenmiş eşik değerleri ile eşiklendirilerek aday ilgi noktaları belirlenir. Belirlenen her bir aday ilgi noktası için bu ilgi noktasının merkezde olacağı kare bir pencere açılır. Bu ilgi noktasının Zernike tepki değeri pencere içine düşen diğer aday ilgi noktalarına ait Zernike tepki değerleri ile karşılaştırılır. Eğer, merkezdeki aday ilgi noktası en yüksek Zernike tepki değerine sahipse gerçek bir ilgi noktası olarak

işaretlenir, değilse dikkate alınmaz. Bu saptama şeması, giriş resminden elde edilen imge piramidindeki tüm imgelere uygulanarak saptanan ilgi noktaları birleştirilir ve çıktı olarak verilir.

Önerilen ilgi noktası saptama algoritmaları olan YZMÖ ve G-YZMÖ için performans ölçümü Inria veri kümesi üzerinde yinelenebilirlik skoru kullanılarak gerçekleştirilmiştir. Yinelenebilirlik skoru, ilgi noktası saptayıcısının başarımını ölçmek için kullanılan temel kriterdir. Sunulan ilgi noktası saptama algoritmalarının Inria veri kümesinde elde edilen yinelenebilirlik skorları iyi bilinen ilgi noktası saptayıcılar ile karşılaştırılmıştır. Bu kapsamda; YZMÖ için Harris, SIFT, SURF, CenSurE, BRISK ve G-YZMÖ için SIFT, SURF, CenSurE, ORB, BRISK ilgi noktası saptayıcıları ile performans karşılaştırması amacıyla çalışılmıştır. Inria veri kümesindeki "Rotation", "Zoom" ve "Zoom&Rotation" imge dizileri ile elde edilen sonuçlar, YZMÖ ve G-YZMÖ ilgi noktası saptayıcılarının karşılaştırıldıkları neredeyse bütün ilgi noktası saptayıcılara yinelenebilirlik skoru bakımından üstünlük sağladığını göstermektedir. Sunulan algoritmaların ayırt edicilikleri bakımından performansları ise üretilen sentetik ve gerçek test resimlerindeki köşe noktalarının saptanmasındaki başarımları ile gösterilmektedir.

G-YZMÖ ilgi noktası saptama algoritmasının kısmi olarak ilgin dönüşümlere değişimsiz olduğu düşünülmektedir. Çünkü; bir ilgin dönüşümü aslında ölçek, döndürme, öteleme dönüşümlerinin bir kombinasyonu şeklinde düşünülebilir ve de G-YZMÖ bu dönüşümlerin hepsine değişimsiz olduğu için dolaylı olarak ilgin dönüşümü karşı da değişimsiz olmalıdır. Bu durumun ileride doğrulanması planlanmaktadır. Ayrıca, YZMÖ ve G-YZMÖ ilgi noktası saptayıcıları, bir betimleyiciyi de içerecek şekilde genişletilecek ve böylece SIFT ve SURF'te olduğu gibi ilgi noktası saptayıcı ve betimleyiciden oluşan tam bir şema sunulmuş olacaktır.

1. INTRODUCTION

In computer vision, image matching, or correspondence problem, is a fundamental research area in order to tackle some important problems such as object recognition, stereo vision, image registration and motion tracking.

Correspondence problem is about matching same regions from different images, which are taken from same scene, under varying viewing conditions. These kinds of images are geometrically or photometrically transformed versions of each other and searching for corresponding regions between such images is a hard problem. Scale change because of zoom in/out by camera lens, rotation and translation in the camera cause geometric transformations whereas change in lighting condition is an example of photometric transformations. Under these transformations, corresponding regions still need to be detected and matched with high repeatability.

General approach to correspondence problem is to consider it in three steps. First, interesting points, which represent distinctive regions in the image, are detected. Corners and blobs are good candidates for such regions. These points also need to be repeatedly detected across the images in interest. Next, the region around each detected interesting point is described as a feature vector. Finally, feature vectors are compared to each other by some distance criteria such as Euclidean or Mahalanobis in order to match the most similar regions whose feature vectors have the least distance/most similar.

Image matching is widely considered in computer vision. For instance, in a stereo system where two images belonging to same scene exist, same regions in both images need to be detected and matched in order to evaluate the depth of the object. For image stitching problem, corresponding regions, which are detected and matched in the low-resolution images of same scene, are used for concatenation in order to build a high-resolution panorama image. In object recognition framework, detected distinctive regions are described as feature vectors and then stored in a database in order to match them with feature vectors of objects in given query image. Thus, the matched objects in query image can be recognized.

The first phase in corresponding problem is to detect distinctive regions in the images. This is very critical phase and designates the performance of subsequent phases such as description and matching. As mentioned before, a distinctive region is represented well by an interesting point in the image. An interest(ing) point, or keypoint, is a local image feature that is extracted from regions where high information or distinctiveness exists. Corners and blobs are such structures that show interesting characteristics whereas flat regions such as walls and surfaces don't contain any information and they don't exhibit interesting characteristics. Hence, in general, interest point detection algorithms are designed with corner or blob detection mechanisms. Locality of interest points makes interest point detection algorithms robust to background clutter and occlusion. Locality also provides translation-invariant characteristic to interest point detection algorithms because local regions move together in case of image translation and thus information in local regions is preserved.

Interest point detection is a feature extraction approach and it's important in order to represent the image in more compact way. By representing the image with a set of feature vectors extracted from detected interest points, all redundant information may be discarded and meaningful information is preserved only. This helps search space to be reduced dramatically for subsequent phases of the problem in interest.

A robust interest point detection algorithm is expected to be invariant to geometric and photometric transformations as well as robust to background clutter and occlusion. Translation invariance problem and robustness to background clutter/occlusion are solved by locality characteristic of interest points. Rotation invariance requires special characteristics inherent to corner detection algorithm. For scale invariance problem, a general approach is to build an image pyramid in scale-space in order to sample the input image in various scale levels and apply interest point detection algorithm for each image in the pyramid. This approach is named as multi-scale or scale-space representation. Invariance to photometric transformations such as illumination changes can be overcome by applying some normalization procedures on the image.

Image moments are widely used transformations in computer vision and applied to many problems such as character recognition, image reconstruction and face recognition. An image moment basically projects the image function on to some

polynomial basis in order to transform the image to the scalar quantities and extract meaningful features from the image. For instance, area of the surfaces in a binary image or central coordinates of a given image can be evaluated by image moments. Image moments can be classified as geometric, complex or orthogonal. One of the well-known orthogonal image moments is Zernike moment that uses Zernike polynomials as polynomial basis and projects the image on to these polynomials defined in a unit circle. Zernike moments can be applied on to the whole image globally or image pixels locally. Scalar quantities evaluated by applying Zernike moments on to the images aren't affected from image rotations and this is a reason of Zernike moments to be rotation-invariant.

In this thesis; two novel interest point detection algorithms are proposed. First one is named as Local Zernike Moment based Features (LZMF). LZMF is invariant to rotation, translation and illumination changes as well as robust to background clutter and occlusion but suffers from scale changes. To tackle scale invariance problem, a second algorithm is introduced and named as Robust Local Zernike Moment Based Features (R-LZMF). R-LZMF builds an image pyramid in scale-space in order to be invariant against scale changes. In both algorithms, local Zernike moments are used as convolutional operators for corner detection with some other mechanisms. Final detectors, LZMF and R-LZMF, are tested on the Inria Dataset with well-known interest point detectors such as Scale Invariant Feature Transform (SIFT) and Speeded-Up Robust Features (SURF) and they outperform in almost all cases in terms of repeatability score that is the main evaluation criterion for interest point detection performance.

1.1 Literature Review

In this section, well-known interest point detection algorithms are examined with their advantages and drawbacks. Studies about application of Zernike moments to some computer vision and pattern recognition problems in global and local manner are also covered.

1.1.1 Interest point detection

One of the earliest interest point detectors was developed by Harris *et al.* and named as Harris corner detector [1]. This detector searches for large intensity changes in a

shifted window by using Sum of Squared Distance (SSD) and considers such locations as corner. Harris detector is rotation-invariant but not scale-invariant.

Andrew Witkin introduced scale-space concept in his seminal work [2] by exposing signals in different scale levels in order to show how signal behavior changes from fine scales to coarse scales. He also showed that smoothing the image with Gaussian filters of increasing standard deviation (σ) has ability to suppress fine details and expose coarse structures. Koenderink, in [3], showed that Gaussian filter is the unique filter for building scale space. Lindeberg verified this uniqueness and proposed an automatic scale selection mechanism in order to find the characteristic scales of interest points in the image [4]. Lindeberg showed that scale-normalized Laplacian-of-Gaussian (LoG) operator, $\sigma^2 \nabla^2 G$, should be used to provide true scale invariance and he used this operator to detect blob-like structures in the image. One drawback of using LoG operator is that it's not a fast operator to apply on the image although it's very good at exposing blobs.

Lowe, in his study [5], proposed Scale Invariant Feature Transform (SIFT) that consists of an interest point detector and descriptor. For interest point detection, he used scale-normalized LoG as in the study of Lindeberg to detect blobs in the image. However, by considering slowness of LoG operator, he suggested using Difference-of-Gaussian (DoG) to approximate LoG operator. DoG is the difference of two images convolved with Gaussian filters of consecutive scales and it's a fast alternative of LoG operator. In SIFT, an image pyramid is built in scale-space by convolving the input image with Gaussian filters of scales differing by a constant factor and difference of two Gaussian images with consecutive scales in the pyramid is taken to apply LoG operator. Thus, scale-space representation and application of LoG operator are both realized in a very efficient way. This is one of the key factors to make SIFT very fast during interest point detection. One drawback of using LoG/DoG is that they also yield high responses in the neighborhood of contours or straight edges, which don't exhibit interest point characteristic. For accurate keypoint localization, SIFT applies some other mechanisms such as interpolation with quadratic Taylor expansion.

Mikolajczyk *et al.* combined Harris corner detector with scale-normalized LoG for interest point detection in [6]. This schema was named as Harris-Laplace detector

and it extracts complementary features in the image by using Harris detector, which responds to corners and textured regions, and LoG, which responds to blobs. Harris-Laplace detector detects Harris corners in spatial-space because Harris detector is the most reliable detector when rotation, illumination change and perspective deformation occur in the image [7]. However, as mentioned before, Harris detector is not scale-invariant and can't detect corners in images with different resolutions. It also fails to determine characteristic scale because it can't reach to maximum frequently in scale-space. Hence, when building image pyramid in scale-space, LoG operator is applied to the locations where Harris corners are detected and local maxima are searched in order to detect interest points. Harris-Laplace detector has a performance up to a scale factor of 4 for scaled images according to [6].

Harris-Laplace detector was extended by Mikolajczyk *et al.* by determining the shape of the elliptical region with the second moment matrix in [25,26] and it was named as Harris-Affine detector. A second affine region detector, which was named as Hessian-Affine, was also proposed in [25,26]. Hessian-Affine detector utilizes Hessian matrix for interest point detection in spatial space and uses Laplacian for scale-space. Harris-Affine and Hessian-Affine detectors have significant invariance to affine transformations when compared to Harris-Laplace detector.

Bay *et al.*, in [8], introduced Speeded-Up Robust Features (SURF) as an interest point detection and description schema. In SURF, Hessian-based detector is used instead of Harris-based counterpart during interest point localization because Hessian function is more stable and repeatable. As in SIFT and Harris-Laplace, SURF uses LoG operator in scale-space in order to determine characteristic scales of detected interest points and this operator is approximated by using determinant of Hessian matrix. Scale-space representation of SURF, however, is different than SIFT and Harris-Laplace. Here, the pyramid is built for LoG operator instead of input image. In other words, the input image is not down-sampled, but rather, LoG filter is up-sampled in scale-space. Up-scaling the filter instead of down-scaling the image prevents aliasing problems occurred in the image when the image pyramid is built. SURF is also faster than SIFT because up-scaled filters are used with efficient integral image method.

Rosten *et al.* developed a fast interest point detector in [9] and named it as Features from Accelerated Segment Test (FAST). FAST tests each image pixel for cornerness by looking its 16 pixel-circular neighborhood and if some contiguous pixels in this neighborhood are brighter/darker than the pixel in test then this pixel is detected as corner. This method also learns from image pixels by applying decision tree to increase its accuracy. Interest points detected by FAST are not multi-scale features, in other words, FAST is not scale-invariant. Oriented FAST and Rotated Brief (ORB), proposed in [10] by Rublee *et al.*, is a combination of FAST keypoint detector and BRIEF descriptor [11]. ORB modifies FAST to work with image pyramid for scale invariance and it also modifies BRIEF descriptor to make it rotation invariant. Center Surround Extrema (CenSurE) is another scale and rotation invariant interest point detector proposed by Agrawal *et al.* in [12]. In CenSurE, a center-surround filter is applied to the image at all locations and scales, and Harris function is used for eliminating weak corner points. Leutenegger *et al.* proposed a rotation and scale-invariant key point detector named as Binary Robust Invariant Scalable Keypoints (BRISK) in [13]. BRISK uses a novel scale-space FAST-based detector for scale-invariant interest point detection and considers a saliency criterion by using quadratic function fitting in continuous domain.

1.1.2 Zernike moments

Khotanzad *et al.* proposed one of the earliest studies about image recognition by global Zernike moments in [14]. The method solves rotation invariance problem by using magnitude of Zernike moments; for scale and translation invariance, regular moments are used. Tests on a 26-class character set shows the superiority of Zernike moments. One other approach using Zernike moments is the study of Ghosal *et al.* in [15]. In this study, Ghosal proposed a unified framework for low-level image feature detection by using local Zernike moments. These low-level features are step/roof edges, gray-level corners and topological features. For corner detection, A_{22} and A_{20} Zernike filters are convolved with the image and the results are thresholded in order to find strong corner responses. In [16], Sariyanidi *et al.* applied local Zernike moments (LZM) for face recognition problem and proposed a novel LZM representation that outperforms Gabor and Local Binary Patterns (LBP) methods on FERET database.

1.2 Organization of the Thesis

This thesis is organized as follows: Chapter 2 provides an overview of image moments including Zernike moments in detail. Chapter 3 contains a detailed explanation of proposed interest point detection algorithms, LZMF and R-LZMF. In this chapter, decisions on parameter settings of proposed algorithms are also justified. Chapter 4 discusses the experimental results with evaluation criterion and dataset used throughout the experiments. The performance of both algorithms, LZMF and R-LZMF, are demonstrated separately in this chapter. Chapter 5 concludes the thesis with some future directions.

2. AN OVERVIEW OF IMAGE MOMENTS

In this chapter, a brief introduction of image moments is given. Image moments are examined based on projected polynomial basis as geometric, complex and orthogonal with providing some examples for their applications. Zernike moment, which is a special orthogonal moment based on Zernike polynomials, are explained and it's shown that how Zernike moments are applied on the image in global and local sense. Rotational invariance of Zernike moments is also proved in the end of section.

2.1 Image Moments

A moment is a scalar quantity that characterizes a function to capture its significant features [18]. In mathematical domain, a moment is defined as a projection of a function onto a polynomial basis. For instance, Fourier transform is a projection onto a basis of harmonic functions.

Moments are used for digital images in order to tackle some computer vision problems such as character recognition, image reconstruction. By considering an image as a piece-wise continuous real function, a general moment M_{pq} for an image function, $f(x, y)$, of order $r = p + q$, where p and q are non-negative integers, is defined as

$$M_{pq} = \iint_D p_{pq}(x, y) f(x, y) dx dy \quad (2.1)$$

Where $p_{pq}(x, y)$ is a polynomial basis function.

Moments can be classified based on polynomial basis used for projection as geometric, complex and orthogonal and they are examined in following subsections in a nutshell.

2.1.1 Geometric moments

A geometric moment is defined on a standard power basis $p_{pq}(x, y) = x^p y^q$ as

$$m_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q f(x, y) dx dy \quad (2.2)$$

Geometric moment of order $0 = 0 + 0$, m_{00} , gives mass of the image. The mass is area of the object in binary images. A geometric moment of order 0 is defined as

$$m_{00} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy \quad (2.3)$$

Geometric moment of order $1 = 1 + 0 = 0 + 1$, m_{10} and m_{01} , represents the center of mass of the image and is defined as

$$m_{10} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f(x, y) dx dy \quad (2.4)$$

$$m_{01} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f(x, y) dx dy \quad (2.5)$$

The coordinates of the center of mass, \bar{x} and \bar{y} , are

$$\bar{x} = \frac{m_{10}}{m_{00}}, \quad \bar{y} = \frac{m_{01}}{m_{00}} \quad (2.6)$$

In practice, the center of mass is used to represent the position of an image in Field of View (FoV) [19].

2.1.2 Complex moments

A complex moment, which is projected on the polynomial basis $p_{pq}(x, y) = (x + iy)^p (x - iy)^q$ where i is the imaginary unit, is defined as

$$c_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x + iy)^p (x - iy)^q f(x, y) dx dy \quad (2.7)$$

Where p and q are non-negative integers and $i = \sqrt{-1}$.

A complex moment of order r is a linear combination of geometric moments of the same order and it's expressed as

$$c_{pq} = \sum_{k=0}^p \sum_{j=0}^q \binom{p}{k} \binom{q}{j} (-1)^{q-j} i^{p+q-k-j} m_{k+j, p+q-k-j} \quad (2.8)$$

Complex moments show rotation-invariant characteristic and are preferred for the images where rotational transformations occur. However, they suffer from information loss, suppression and redundancy, which shouldn't exist in a good moment invariant [19]. Therefore, complex moment invariants are not used as image features.

2.1.3 Orthogonal moments

An orthogonal moment is defined on the polynomial basis, $p_{pq}(x, y)$, whose elements satisfy the orthogonality condition as

$$\iint_{\Omega} p_{pq}(x, y) p_{mn}(x, y) dx dy = 0 \quad (2.9)$$

For any indexes $p \neq m$ or $q \neq n$.

Orthogonal moments yield good image features that are non-redundant and require low computing precision. Image reconstruction, which can't be performed by geometric moments in spatial domain directly, can be achieved by

$$f(x, y) = \sum_{k,j} M_{kj} p_{kj}(x, y) \quad (2.10)$$

This is an optimal reconstruction because the mean-square error is minimized when using only a finite set of moments.

Orthogonality for orthogonal moments can be provided on rectangle or unit disk. Legendre, Chebyshev and Zernike moments are some examples of orthogonal moments.

2.2 Zernike Moments

In this section, Zernike moments are examined based on how they're applied on the images. First, application of Zernike moments to whole image is shown and then projection of local intensity profiles onto Zernike polynomials is considered. Finally, rotational invariance of Zernike moments is explained with providing proper method to apply on the images.

2.2.1 Global Zernike moments

Zernike, in [20], introduced a complete and orthogonal set of complex polynomials, which are named as Zernike polynomials, on the unit disk where $x^2 + y^2 \leq 1$. Zernike polynomials are defined as

$$V_{nm}(x, y) = V_{nm}(p, \theta) = R_{nm}(p)e^{jm\theta} \quad (2.11)$$

Where $R_{nm}(p)$ is radial polynomial, n is order of polynomial, m is number of iteration, p is length of vector from origin to (x, y) and θ is angle between p and x -axis in counter-clockwise direction. There are some constraints on n and m parameters such as $n \geq 0$, $n - |m| = \text{even}$ and $|m| \leq n$. $R_{nm}(p)$ is defined as

$$R_{nm}(p) = \sum_{s=0}^{\frac{n-|m|}{2}} \frac{(-1)^s p^{n-2s} (n-s)!}{s! \left(\frac{n+|m|}{2} - s\right)! \left(\frac{n-|m|}{2} - s\right)!} \quad (2.12)$$

Visual representations of Zernike polynomials can be examined in Figure 2.1. Teague used Zernike polynomials as orthogonal image moments in [21] for a two-dimensional pattern recognition problem. Given an image function of $f(x, y)$, a Zernike moment of order n and repetition m is defined as

$$A_{nm} = \frac{n+1}{\pi} \iint_{x^2+y^2 \leq 1} f(x, y) V_{nm}^*(p, \theta) dx dy \quad (2.13)$$

Where $*$ in $V_{nm}^*(p, \theta)$ denotes the complex conjugate. The equation in (2.13) is discretized in order to work with digital images of size $M \times N$ as

$$A_{nm} = \frac{n+1}{\pi} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} f(i,j) V^*(p_{ij}, \theta_{ij}) \Delta x_i \Delta y_j \quad (2.14)$$

Where $x_i, y_j \in [-1,1]$, $p_{ij} = \sqrt{x_i^2 + y_j^2}$, $\theta_{ij} = \tan^{-1} \frac{y_j}{x_i}$ and $\Delta x_i = \Delta y_j = 2/N\sqrt{2}$.

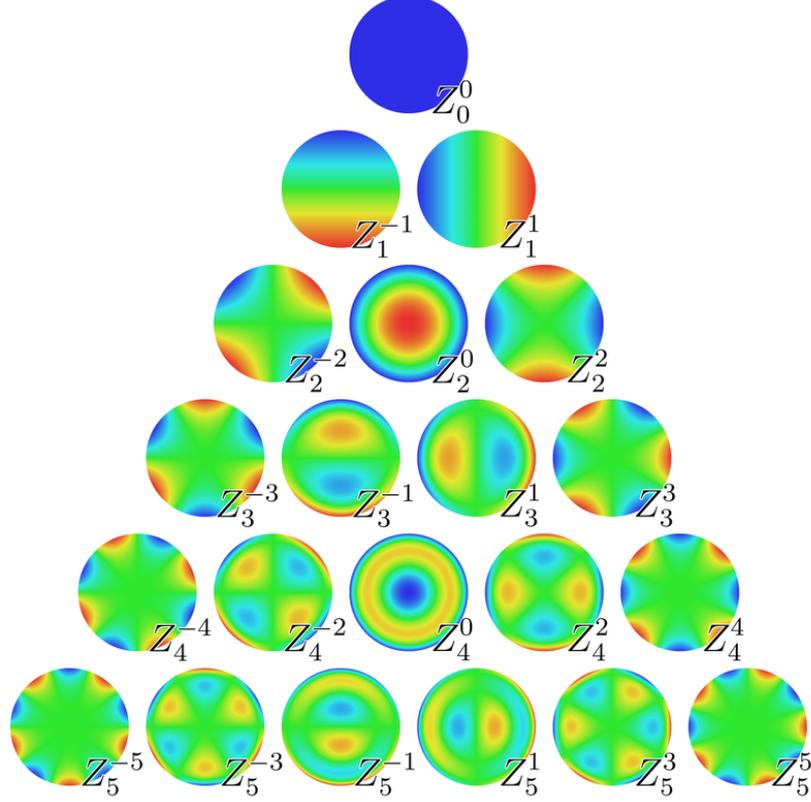


Figure 2.1 : Visual representations of Zernike polynomials of order 0 to 5 [23].

2.2.2 Local Zernike moments

As seen from (2.14), a Zernike moment, A_{nm} , is a measurement about intensity profile of the whole image. It's also possible to project the local intensity profiles onto Zernike polynomials by fitting the unit circle on pixels of the image. The image moments using Zernike polynomials in this way are named as Local Zernike Moments or LZM shortly. As proposed in [15], LZMs have potential to expose low-level image features such as gray-level corner points or step edge points.

A LZM representation is obtained by convolving the image with moment-based operator, V_{nm}^k , that is a 2D convolution filter of size $k \times k$ and defined as

$$V_{nm}^k(i,j) = V_{nm}(p_{ij}, \theta_{ij}) \quad (2.15)$$

Convolution of the image with $V_{nm}^k(i, j)$ is formulated as

$$A_{nm}^k(i, j) = \sum_{p, q = -\frac{k-1}{2}}^{\frac{k-1}{2}} f(i-p, j-q) V_{nm}^k(p, q) \quad (2.16)$$

Zernike moments are based on complex Zernike polynomials. Therefore, they are constructed as real and imaginary convolutional filters. Imaginary filter is not considered when there is no repetition ($m = 0$). In this case, real Zernike filter, $Re(V_{nm}^k)$, is convolved with the image only and real Zernike moment representation, $Re(A_{nm})$, is taken into account.

Zernike moments show rotation-invariant characteristic. If an image, $f(x, y)$, is rotated by an angle of α w.r.t. x -axis as $f'(x, y)$ then relationship between Zernike moments of the original and rotated images, A_{nm} and A'_{nm} , is as follows

$$A'_{nm} = \iint f'(x, y) V_{nm}^*(p, \theta) dx dy = A_{nm} e^{-jm\theta} \quad (2.17)$$

As seen from (2.17), although there is a phase shift because of image rotation, the magnitude of Zernike moment, $|A_{nm}|$, remains same. Thus, the magnitude of Zernike moments can be used to be invariant against image rotations. The magnitude of Zernike moment is defined as

$$|A_{nm}| = \sqrt{[Re(A_{nm})]^2 + [Im(A_{nm})]^2} \quad (2.18)$$

3. ROBUST INTEREST POINT DETECTION

In this chapter, details of proposed interest point algorithms, LZMF and R-LZMF, are explained. First, principles about designing a robust interest point detection algorithm are given in a nutshell. Then, LZMF and R-LZMF algorithms are explained with their major steps. Some detection results on synthetic and real images are also provided in order to prove the capabilities of proposed algorithms.

3.1 Principles of Robust Interest Point Detector Design

A robust interest point detector should be invariant to geometric and photometric transformations occurred in the images that are taken from same scene under different conditions. Scale, rotation and translation changes in the images cause geometric transformations whereas change in lighting conditions is an example of photometric transformations. An interest point detector should also be robust against background clutter and occlusion.

Invariance against scale changes is usually handled by building image pyramid in scale-space. All possible scale levels are sampled by convolving the image with a scale-space operator such as Gaussian filter. Thus, all structures represented in different scales can be exposed. Gaussian filter is known to be the best scale-space operator [3,4].

Rotational invariance is a characteristic of interest point operator in general. Zernike moments, for instance, have a property of rotational symmetry that makes them invariant against rotations. Harris corner detector is an operator that shows rotational invariance.

Locality is a key characteristic that gives robustness to translational changes, background clutter and occlusion. Local features in the image move together in case of translation. Hence, information around a local feature never changes and this makes local feature based interest point detector translation-invariant. Local features

also eliminate the effect of background clutter and occlusion because local regions have small sizes in the image [6].

The change in lighting is encountered frequently in the images. A general approach is to normalize the image in order to decrease the effect of lighting changes to the minimum. In this way, the range of pixel values is fitted to some intervals so that large variations because of changes in lighting may be eliminated. The normalization can be local in the image patch or global for the whole image.

3.2 Proposed Interest Point Detection Algorithms: LZMF and R-LZMF

In this section, proposed interest point detection algorithms, LZMF and R-LZMF, are explained in details. LZMF algorithm steps are briefly as follows. The input image is first converted to the gray-scale image if not. Then, the gray-scaled image is globally normalized by L_2 normalization. Before corner detection, the local image patch, where corner detector is applied on, is normalized by fitting to standard normal distribution. $|A_{42}|$ and $\frac{|A_{42}|}{|A_{40}|}$ response maps are obtained by convolving the image with $Re(V_{42}^k)$ and $Im(V_{42}^k)$ and $Re(V_{40}^k)$. Both $|A_{42}|$ and $\frac{|A_{42}|}{|A_{40}|}$ response maps are thresholded by different global threshold values and the image pixels, which pass thresholding tests, are considered as candidate corner points. Non Maximum Suppression (NMS) is applied on candidate corner points in spatial-space as 2D for thinning interest point detection density. Resulting interest points are the output of LZMF.

The procedure for R-LZMF algorithm is briefly as follows. An image pyramid is built in scale-space by using Gaussian filter. LZMF is applied on each scale layer of the image pyramid in order to detect candidate interest points. Each candidate interest point is analyzed based on cornerness in scale-space for 3D Non Maximum Suppression (3D-NMS). Resulting interest points are the output of R-LZMF.

3.2.1 Normalization

The input image must be gray-scale in order to apply normalization procedure, therefore, multi-channel images must be converted to single-channel. Working with gray-scale image is efficient because single-channel has enough information for

interest point detection and search-space for interest points is greatly reduced in this way.

There are two normalization steps for gray-scale input image. One is global and applied to the whole image, second is local and applied in the unit circle where Zernike filter is operated for convolution. The reason of applying such normalization procedures is to make proposed interest point detectors more robust against effect of outliers in the image.

L_2 normalization is applied to the whole image as first normalization step in order to reduce effect of changing lighting conditions in the image. In this normalization, each pixel is divided by sum of squared root of pixel intensity values to fit its value in range of 0 and 1. This normalization is formulated for image $f(x, y)$ with size of $M \times N$ as below

$$\|f(x, y)\|_{L_2} = \frac{f(x, y)}{\sqrt{\sum_{i=0}^{M-1} \sum_{j=0}^{N-1} f(i, j)^2}} \quad (3.1)$$

In Figure 3.5 (a) and (b), the result of applying such normalization is shown. As seen from Figure 3.5 (b), applying L_2 normalization sweeps noisy detections out. The number of interest points also reduces from 45 to 26 as seen from Figure 3.5 (a) to (b) without missing any true corner points. Similarly, in Figure 3.6 (a) and (b), applying L_2 normalization reduces the number of keypoints from 77 to 43 with discarding noisy interest points.

Before convolving the image with a Zernike filter, a second normalization step is locally applied to the regions where unit circle of the filter is fitted. Pixel intensity values falling in unit circle of Zernike filter are subtracted from mean and divided by standard deviation. Thus, local intensity profile in unit circle is fitted to standard normal distribution with $\mu = 0$ and $\sigma = 1$. Here, mean and standard deviation are the statistics of local intensity profile falling in unit circle. This approach is similar to normalization in Normalized Cross-Correlation (NCC) and makes proposed interest point detectors more robust to local intensity variations.

3.2.2 Corner detection

Corners and blobs are good structures to be interesting points in the image [1,4,5,6,8,9,10,13]. By following this fact, proposed interest point detection algorithms were designed to be corner based and local Zernike moments were used for corner detection phase.

Ghosal *et al.*, in [15], used $|A_{22}|$ to measure cornerness. However, $|A_{22}|$ is a Zernike moment that responds to the edge points closer to the true corner points as well. Therefore, these nearby edge points should be suppressed in order to respond to true corners only. In [15], this problem was tackled by evaluating $\left|\frac{A_{22}}{A_{20}}\right|$. According to Ghosal, $\left|\frac{A_{22}}{A_{20}}\right|$ approaches one at nearby edge points whereas it reaches high values for true corner points. Briefly, a point in interest is considered as a true corner point if both $|A_{22}|$ and $\left|\frac{A_{22}}{A_{20}}\right|$ have high responses for that point. The corner detection model of Ghosal was first applied for LZMF. However, this model didn't yield expected results with test images in terms of repeatability criterion. For test images, it had repeatability scores of around 50% at most. Therefore, different orders of Zernike moments were investigated for cornerness and suppression of nearby edge points. This was concluded with using similar but more complex Zernike moments, $|A_{42}|$ and $|A_{40}|$. These Zernike moments had superior repeatability performance on test images as will be discussed in Chapter 4. Visual representations of Zernike moments of A_{20} (top-left), A_{22} (top-right), A_{40} (bottom-left) and A_{42} (bottom-right) can be examined in Figure 3.1. As seen from Figure 3.1, Zernike moment of A_{42} looks like a corner, which is considered as an intersection of two edges, and using A_{42} for convolution, which searches for maximum similarity of local region with applied filter, exposes regions that are similar to corner. Symmetrical and rotational invariance properties of A_{42} make it possible to detect every case of edge intersections in image plane.

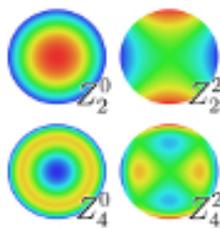


Figure 3.1 : Zernike moments used for corner detection [23].

Theoretical explanation of why A_{42} works for corner detection is as follows. Zernike polynomial of V_{42} in Cartesian coordinate system is defined as

$$V_{42}(x, y) = \sqrt{10}(4x^4 - 4y^4 - 3x^2 + 3y^2) \quad (3.2)$$

The equation in (3.2) can be rewritten with ignoring constant factor $\sqrt{10}$ as

$$\begin{aligned} V_{42}(x, y) &= 4x^4 - 4y^4 - 3x^2 + 3y^2 \\ &= 4(x^4 - y^4) - 3(x^2 - y^2) \\ &= 4(x^2 - y^2)(x^2 + y^2) - 3(x^2 - y^2) \\ &= (x^2 - y^2)(4x^2 + 4y^2 - 3) \end{aligned} \quad (3.3)$$

By converting Cartesian coordinates, x and y , to polar coordinates with $x = \rho \cos \theta$ and $y = \rho \sin \theta$, (3.3) is rewritten as

$$\begin{aligned} V_{42}(\rho, \theta) &= (\rho^2 \cos^2 \theta - \rho^2 \sin^2 \theta)(4\rho^2 \cos^2 \theta + 4\rho^2 \sin^2 \theta - 3) \\ &= \rho^2(\cos^2 \theta - \sin^2 \theta)[4\rho^2(\cos^2 \theta + \sin^2 \theta) - 3] \\ &= \rho^2 \cos 2\theta (4\rho^2 - 3) \\ &= (4\rho^4 - 3\rho^2) \cos 2\theta \end{aligned} \quad (3.4)$$

As in [15], intensity profile of a corner projected onto a unit circle is defined by a Zernike moment after the corner profile is rotated by a degree of $-\phi$ in order to align the bisector of the corner with the x-axis. This operation can be examined in Figure 3.2 (a) and (b).

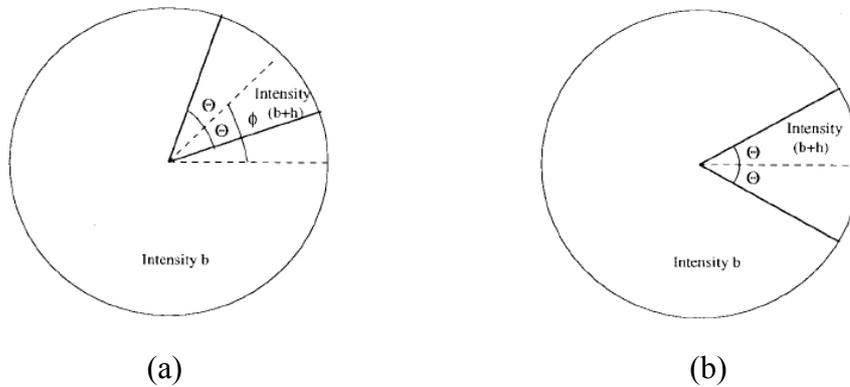


Figure 3.2 : Gray-level corner models fitting onto the unit circle [15].

Relationship between original and rotated corner functions, $f(x, y)$ and $f'(x, y)$, over their projections onto Zernike moments with order of 4 and repetition of 2, A_{42} and A'_{42} , is defined as

$$\begin{aligned} A'_{42} &= A_{42}e^{-2j\theta} \\ &= \iint_{x^2+y^2 \leq 1} (4x^4 - 4y^4 - 3x^2 + 3y^2)f'(x, y) dy dx \end{aligned} \quad (3.5)$$

The equation (3.5) defined in Cartesian coordinate system is rewritten in polar coordinate system by using (3.4) as

$$\begin{aligned} A'_{42} &= b \int_{\theta=0}^{2\pi} \int_{\rho=0}^1 (4\rho^4 - 3\rho^2) \cos 2\theta \rho d\rho d\theta + h \int_{\theta=-\Theta}^{\Theta} \int_{\rho=0}^1 (4\rho^4 - 3\rho^2) \cos 2\theta \rho d\rho d\theta \\ &= b \int_{\theta=0}^{2\pi} -\frac{1}{12} \cos 2\theta d\theta + h \int_{\theta=-\Theta}^{\Theta} -\frac{1}{12} \cos 2\theta d\theta \\ &= 0 + h \left(\frac{-\sin 2\Theta + \sin -2\Theta}{24} \right) \\ &= -h \frac{\sin 2\Theta}{12} \end{aligned} \quad (3.6)$$

Where b and h are the intensities of the outside and inside of the corner respectively, see Figure 3.2.

Based on (3.6), the magnitude of A'_{42} , $|A'_{42}|$, has a low response where flat regions occur (h is small) and edge points exist ($\Theta = 90^\circ$) in the image as similar to corner model of Ghosal *et al.* [15]. Therefore, $|A'_{42}| = |A_{42}|$ is suitable for representing cornerness.

Zernike moment with order 4 and repetition 2, A_{42} , is based on Zernike polynomial of $V_{42}(x, y)$ defined as

$$V_{42}(x, y) = V_{42}(p, \theta) = R_{42}(p)e^{2j\theta} \quad (3.7)$$

Where $R_{42}(\rho)$ is derived as

$$\begin{aligned}
R_{42}(p) &= \sum_{s=0}^1 \frac{(-1)^s p^{4-2s} (4-s)!}{s! (3-s)! (1-s)!} \\
&= \frac{(-1)^0 p^4 4!}{0! 3! 1!} + \frac{(-1)^1 p^2 3!}{1! 2! 0!} \\
&= 4p^4 - 3p^2
\end{aligned} \tag{3.8}$$

By substituting (3.8) in (3.7), Zernike moment filter, $V_{42}^k(i, j)$, can be defined as

$$\begin{aligned}
V_{42}^k(i, j) &= V_{42}(p_{ij}, \theta_{ij}) \\
&= (4p^4 - 3p^2) e^{2j\theta} \\
&= 4p^4 e^{2j\theta} - 3p^2 e^{2j\theta}
\end{aligned} \tag{3.9}$$

Zernike moment A_{42} is constructed as real and imaginary filters, $Re(V_{42}^k)$ and $Im(V_{42}^k)$, because repetition is not zero ($m \neq 0$).

Zernike polynomial for Zernike moment with order 4 and repetition 0, $V_{40}(x, y)$, is defined as

$$V_{40}(x, y) = V_{40}(p, \theta) = R_{40}(p) \tag{3.10}$$

Where $R_{40}(\rho)$ is formulated as

$$\begin{aligned}
R_{40}(p) &= \sum_{s=0}^2 \frac{(-1)^s p^{4-2s} (4-s)!}{s! (2-s)! (2-s)!} \\
&= \frac{(-1)^0 p^4 4!}{0! 2! 2!} + \frac{(-1)^1 p^2 3!}{1! 1! 1!} + \frac{(-1)^2 p^0 2!}{2! 0! 0!} \\
&= 6p^4 - 6p^2 + 1
\end{aligned} \tag{3.11}$$

Zernike moment filter, $V_{40}^k(i, j)$ is then defined by substituting (3.11) in (3.10) as

$$\begin{aligned}
V_{40}^k(i, j) &= V_{40}(p_{ij}, \theta_{ij}) \\
&= (6p^4 - 6p^2 + 1) e^{0j\theta} \\
&= 6p^4 - 6p^2 + 1
\end{aligned} \tag{3.12}$$

Repetition of Zernike moment A_{40} is zero ($m = 0$) and hence real filter, $Re(V_{40}^k)$, is the only convolution filter to be considered.

LZMF algorithm uses $|A_{42}|$ to measure cornerness of each pixel in the image. The real and imaginary Zernike filters, $Re(V_{42}^k)$ and $Im(V_{42}^k)$, are convolved with the

image in order to get $Re(A_{42})$ and $Im(A_{42})$ representations. These representations are used to get the magnitude of A_{42} , $|A_{42}|$, as follows

$$|A_{42}| = \sqrt{[Re(A_{42})]^2 + [Im(A_{42})]^2} \quad (3.13)$$

A global threshold, which is named as "corner threshold", t_c , is applied to $|A_{42}|$ response map in order to detect candidate corner points. The pixels whose $|A_{42}|$ value is higher than the "corner threshold" are considered as candidate corner points and the rest is discarded.

$|A_{42}|$ response map may have high responses where edge points, which are closer to true corner points, exist. These nearby edge points are suppressed by dividing the magnitude of $|A_{42}|$ to $|A_{40}|$ as $\frac{|A_{42}|}{|A_{40}|}$. The real Zernike filter, $Re(V_{40}^k)$, is convolved with the image in order to get $Re(A_{40})$ representation. As a note, there is no imaginary representation for A_{40} as $Im(A_{40})$ because repetition is zero ($m = 0$). The magnitude of A_{40} , $|A_{40}|$, is obtained by

$$|A_{40}| = \sqrt{[Re(A_{40})]^2} \quad (3.14)$$

$\frac{|A_{42}|}{|A_{40}|}$ response map is thresholded by a second global threshold, which is named as "nearby edge threshold", t_e , in order to suppress nearby edge points. The candidate corner points whose $\frac{|A_{42}|}{|A_{40}|}$ is higher than the "nearby edge threshold" are retained and the rest is eliminated.

LZMF algorithm was tested with synthetic and real images for corner detection. In Figure 3.3 (a) and (b), it's seen that the corners of synthetic corner image and its rotated version by a degree of 30 were successfully detected. Detected corner points are indicated as red circles.

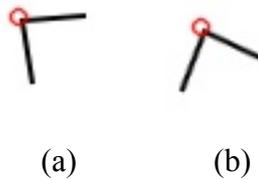


Figure 3.3 : Detection results of LZMF with synthetic corner images.

In Figure 3.4 (a), (b) and (c), it's shown that corners on the checkerboard and its two rotated versions by degrees of 40 and 70 were detected respectively. As seen from Figure 3.4, all corner points on the checkerboards are detected by LZMF although it's rotated by degrees of 40 and 70. Detections are indicated as red circles on all three checkerboards.

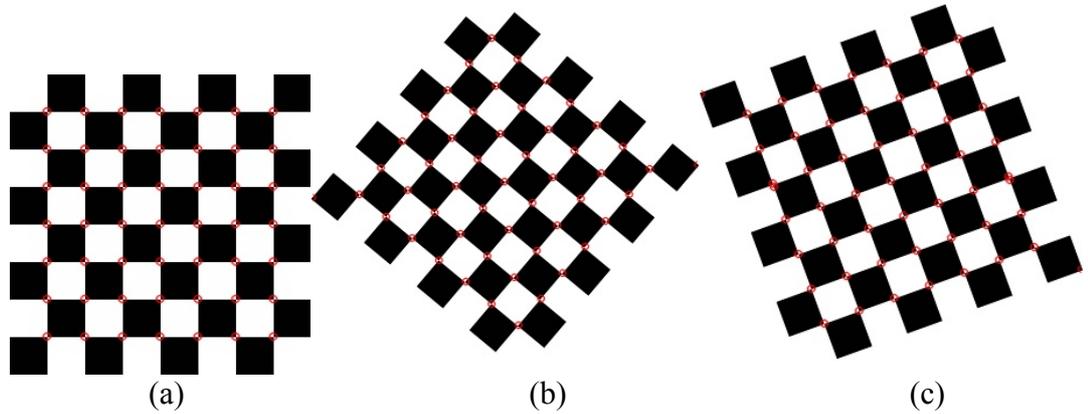


Figure 3.4 : Detection results of LZMF with checkerboard images.

Some detection results for LZMF on real images can be examined in Figure 3.5 and Figure 3.6. Detection rate for floor image in Figure 3.5 is 12/12=1 (100%) whereas wall image in Figure 3.6 has a detection rate of 15/18=0.83 (83%).

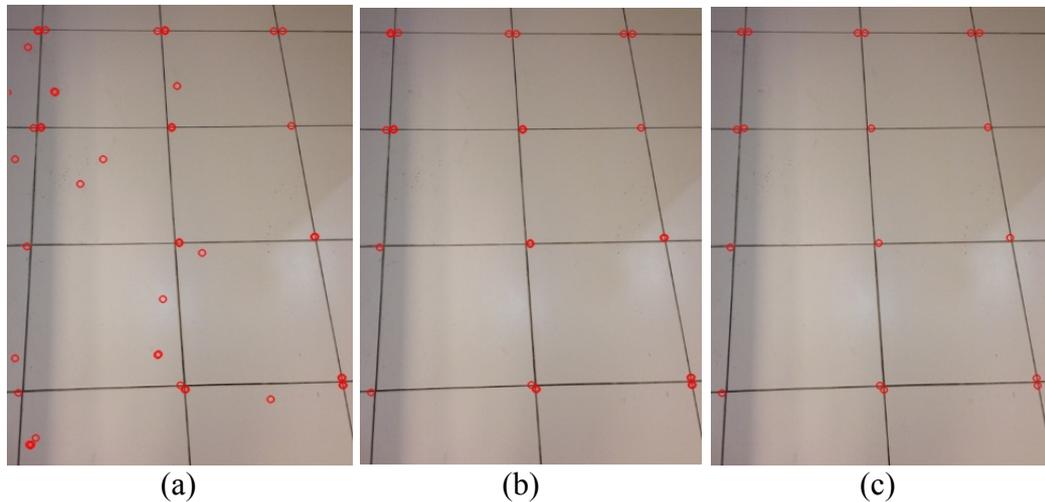


Figure 3.5 : Detection results of LZMF with real floor images.

3.2.3 Non maximum suppression

Candidate corner points detected by Zernike filters, V_{42}^k and V_{40}^k , usually appear at true corner points and their neighborhood in the image. The reason of such clustering

is that V_{42}^k and V_{40}^k have high response around corner points. In Figure 3.5 and Figure 3.6, this effect can be seen by looking at density of detections around true corners.

Density of detected interest points to be high around true corner points is not an expected situation. For a robust interest point detection algorithm, it's expected that interest points representing true corner point should be retained and the rest should be discarded. In other words, those detected points having maximum cornerness in a local neighborhood should be taken into account only. For proposed interest point detection algorithms, the cornerness criterion is the response of V_{42}^k . This thinning procedure is named as Non Maximum Suppression (NMS) and, in this way, more accurate interest points are retained and weak ones are discarded.

Nearby interest points detected in spatial-space by V_{42}^k and V_{40}^k are swept out with NMS as follows for each candidate interest point: i) a 5x5 window is centralized on each candidate interest point, ii) the candidate interest point is compared with other candidate interest points detected in 5x5 local neighborhood based on cornerness value, V_{42}^k , iii) the candidate interest point is retained if its V_{42}^k response is the maximum or discarded otherwise.

The result of applying 5x5 window for NMS is shown in Figure 3.5 (c) and Figure 3.6 (c). As seen from Figure 3.5 (b) and (c), the number of interest points reduces from 26 to 18. Similarly in Figure 3.6 (b) and (c), this number reduces from 43 to 24. The keypoint number can also be further reduced to exact number of true corner points, 12, by using larger windows such as 7x7 or 9x9. However, this would increase the computation time because of more comparison in the window.

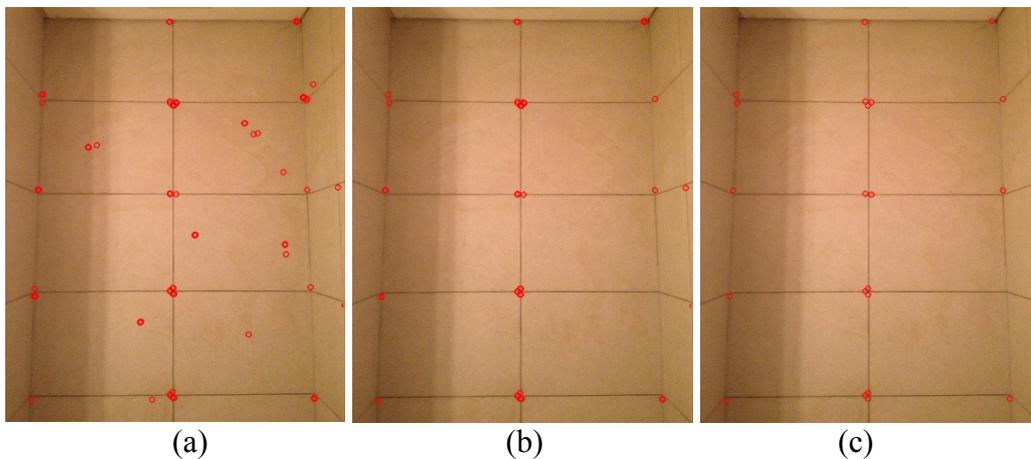


Figure 3.6 : Detection results of LZMF with real wall images.

3.2.4 Scale-space

Scale-space representation exposes a signal in different scale levels in order to show how signal behavior changes from fine scales to coarse scales. In image domain, this is important when the structures in different scales exist in the image and they still need to be detected without considering in which scale they are. For instance, a table can be in different sizes in different images taken from same scene with varying scales. The corners of this table would not be detected in most of the images by a corner detector like Harris because of its lack of invariance against scale. As seen from Figure 3.7, the reason of missing corner points for Harris detector is that shifting window can't determine high change in both x and y axes (λ_1, λ_2) for a corner point much bigger than it. The solution to this is to sample a given image in different scales so that the corner operator have chance to respond the structure well. This kind of sampling images in different resolution levels is named as scale-space representation.

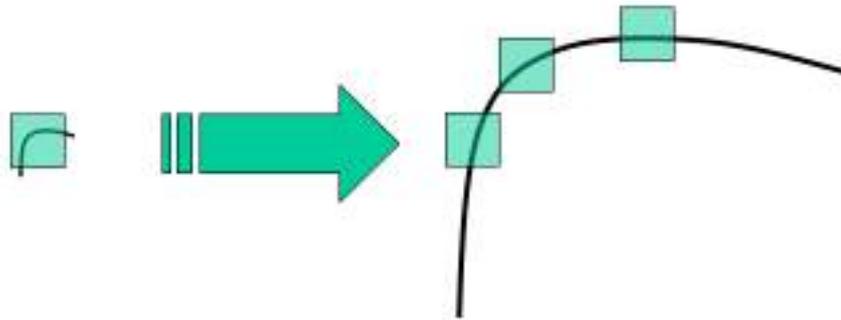


Figure 3.7 : Shifting window of Harris corner detector on the corners [22].

In scale-space, the input image is repeatedly convolved with Gaussian filters of scales differing by a constant factor for blurring and each blurred image constitutes a scale level, $L(x, y)$. As mentioned earlier, Gaussian filter is the only filter that can be used as scale-space operator [3,4]. Here, the blurring effect is to suppress fine details and expose coarse details. For instance, in a given image containing a car and a key, blurring the image continuously with Gaussian filter would suppress the structures related with the key object, which was exposed earlier, and expose the structures of the car object. A scale level $L(x, y)$ for an image $f(x, y)$ is defined as

$$L(x, y) = G(x, y, \sigma) * f(x, y) \quad (3.15)$$

$G(x, y, \sigma)$ is 2D Gaussian function with standard deviation of σ , which refers to scale size, and defined as

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \quad (3.16)$$

Scale-space representation for R-LZMF is divided into octaves for efficient computation. An octave is a stack of scale levels/layers, $L(x, y)$, with same resolution and σ for each scale layer is a constant factor of previous scale layer's σ in that octave. An image in one octave is a sub-sampled version of an image in previous octave. When σ in current octave is doubled, the image convolved with Gaussian filter of this σ is halved (sub-sampled) and used as first scale layer of next octave. Here, sub-sampling is the key factor for computational gain.

In R-LZMF, candidate interest points detected in the images (scale layers) of each octave are analyzed to eliminate weak ones, which have low cornerness, and to figure out characteristic scales of strong ones, which have maximum cornerness. Characteristic scale is the scale level where an interest point attains to local extremum (minimum or maximum according to the type of interest point operator) in scale-space. This is the moment where interest point exhibits the most interesting characteristic (cornerness, blobness etc.) in scale-space. A description extracted from the interest point with characteristic scale size would be independent of interest points detected at same relative location in different scaled images. In other words, the information, which the description of one interest point contains, would always be same for that interest point detected in images with different resolution.

Interest points detected in each octave is analyzed in scale-space as follows: Detected interest point in one scale layer of the octave is compared with other points in its 2D and 3D local neighborhood in terms of distinctiveness such as corner or blob response. Here, 2D local neighborhood refers to points around the point in interest in current scale layer and 3D local neighborhood refers to points around the point in interest relatively in lower and upper scale layers. If its distinctiveness is the extremum in both 2D and 3D local neighborhood then it's considered as an interesting point, otherwise it's discarded. This is a NMS approach in 3D actually and named as 3D Non Maximum Suppression (3D-NMS).

3D-NMS is applied to each octave in R-LZMF in a slightly different way. First, LZMF is applied on all image layers in the octave in order to detect candidate interest points in spatial-space (2D). As a note, extremum check in 2D is already realized by LZMF and there is no need to consider it again. For outermost scale layers, candidate interest points are directly retained without any scale-space analysis. For inner scale layers of the octave, a candidate interest point is compared with interest points in lower and upper scale layers based on V_{42}^k . This comparison again falls in 5x5 neighborhood of the point in interest for the adjacent scale layers. If V_{42}^k response of the interest point is the maximum among all interest points detected in local neighborhood of lower and upper scale layers then the candidate interest point is retained as real interest point and scale of layer in which the interest point resides on is considered as characteristic scale, otherwise the point in interest is discarded.

In Figure 3.8 (a), (b) and (c), the results of applying R-LZMF on checkerboard images are shown. As seen from Figure 3.8, all corner points in the original and scaled versions of checkerboard images with scale factors of 2 and 4 are successfully detected by R-LZMF and indicated as red circles. For the sake of visibility for detected corner points, scaled images are enlarged.

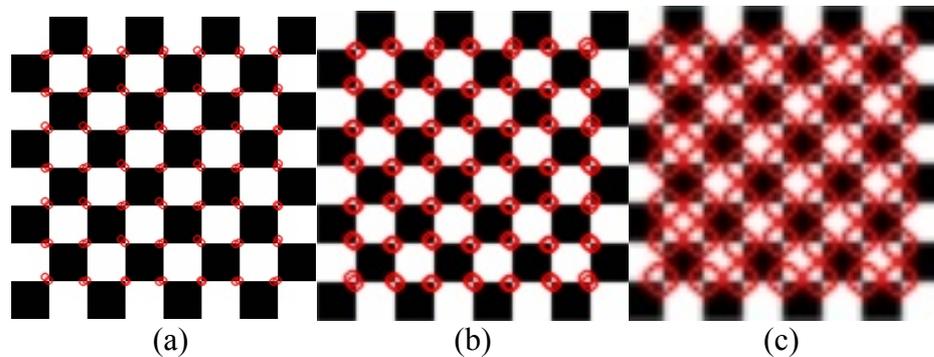


Figure 3.8 : Detection results of R-LZMF with checkerboard images.

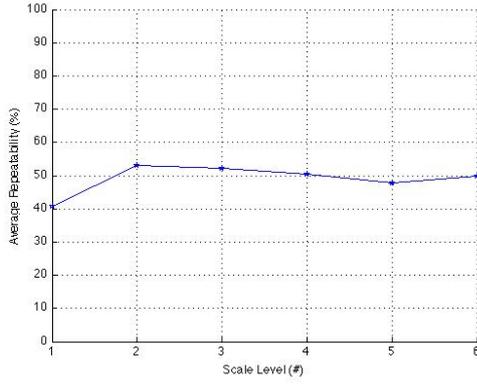
3.2.5 Parameter settings

R-LZMF can be considered as a combination of LZMF and scale-space representation. Therefore, parameter settings of proposed interest point detectors can be examined as configurations of LZMF and scale-space separately and these settings are used throughout the experiments discussed in Chapter 4.

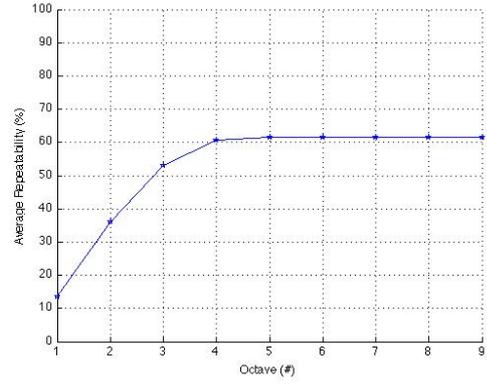
LZMF uses LZM representations, A_{42} and A_{40} , as convolutional filters, V_{42}^k and V_{40}^k , respectively. In the experiments, the optimal results were obtained by using convolution filters of size 9 (k) so that Zernike filters of size 9x9, V_{42}^9 and V_{40}^9 , are used during convolution operations. Threshold values for corner detection were evaluated empirically by trying a wide range of values and determined as 0.51 for "corner threshold" (t_c) and 5 for "nearby edge threshold" (t_e). Window size for NMS in spatial-space (w_{2D}) was determined as 5x5 by considering computation time. Here, there is a trade-off between interest point localization accuracy and computation efficiency, the window size should be chosen in a balance between these factors. A windows size of 5x5 was optimal throughout the experiments.

Scale-space is built based on four parameters and these are number of scale levels in one octave (l), number of octaves in scale-space (o), initial standard deviation for first scale level of the first octave (s_0) and window size for 3D-NMS (w_{3D}). These parameters should be fine-tuned in order to have a full coverage in scale-space. Belledonnes image from the Inria Dataset [17] was used to figure out the optimal parameters for R-LZMF and repeatability criterion was considered for fine-tuning. First, by assuming $o = 3$ and $s_0 = 1.7$, number of scale levels was determined and the best result was obtained by using two scale levels as seen in Figure 3.9 (a). In this case, however, 3D NMS is not applied because outermost scale layers are the only scale layers to be used. Second, by using optimal l ($l = 2$) and assuming $s_0 = 1.7$, optimum number of octaves was searched and, as seen in Figure 3.9 (b), working with four octaves had the best performance. As noted in Figure 3.9 (b), using more than 4 octaves doesn't increase the performance. Finally, by using optimal l ($l = 2$) and o ($o = 4$), s_0 was determined. Although the best performance was obtained with s_0 of 1.6 as seen in Figure 3.9 (c), s_0 of 1.8 had better performance throughout the experiments.

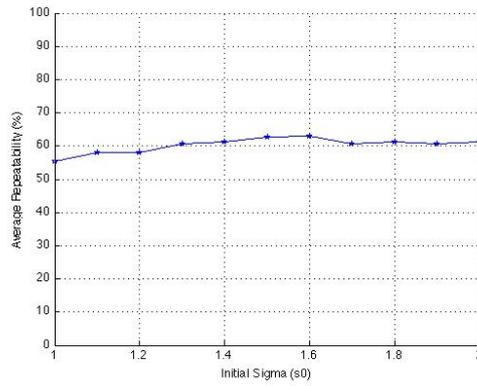
Parameter configuration, which is used throughout the experiments, for both LZMF and scale-space is summarized in Table 3.1.



(a)



(b)



(c)

Figure 3.9 : Parameter evaluation for scale-space based on average repeatability.

Table 3.1 : Parameter settings for LZMF and R-LZMF.

Parameter	Value	Definition
k	9	Zernike filter size
t_c	0.51	Corner threshold
t_e	5	Nearby edge threshold
w_{2D}	5	Window size of 2D-NMS
w_{3D}	5	Window size of 3D-NMS
o	4	Number of octave
l	2	Number of scale level
s_0	1.8	Initial standard deviation

4. EXPERIMENTAL RESULTS

In this chapter, performance evaluation of proposed interest point detection algorithms, LZMF and R-LZMF, is discussed. First, repeatability score, which is the main criterion to evaluate the performance of interest point detectors, is explained. Then, the Inria Dataset, which is used throughout the experiments, is presented. Finally, experimental results in terms of repeatability are explained by comparing LZMF and R-LZMF with well-known interest point detectors such as Harris, SIFT and SURF and the superiority of the proposed algorithms is highlighted.

4.1 Evaluation Criteria

The detector performance is evaluated using the repeatability score proposed in [7]. The repeatability is a measurement of the point correspondence between two images that are transformed (scaled, rotated or translated) forms of each other. The repeatability shows the stability of detector and it's key factor when comparing the performance of detectors. A robust interest point detector is expected to detect the most of the same structures in images even they are scaled, rotated or translated versions of each other. In other words, a robust interest point detector should have high repeatability scores in challenging datasets.

The repeatability score is evaluated by dividing the number of corresponding interest points detected in two images, which are taken from same scene under different conditions, to the minimum number of detected interest points. This score is formulated as

$$r_{12} = \frac{C(f_1, f_2)}{\min(m_1, m_2)} \quad (4.1)$$

Where m_1 and m_2 are numbers of interest points detected in images f_1 and f_2 respectively, and $C(f_1, f_2)$ is number of corresponding matches between images f_1 and f_2 .

One other criterion for evaluating interest point detector performance is the number of detected interest points. Lowe, in [24], emphasizes the importance of number of interest points detected in scenes where the objects are occluded, background clutter exists or the object in interest is too small. There should be many but distinctive interest points detected by interest point detector in order to tackle such problems. If an interest point detector generates many distinctive interest points then the chance of detecting objects in interest increases.

4.2 Dataset

The experiments for evaluating the performance of proposed interest point detection algorithms, LZMF and R-LZMF, were conducted on the Inria Dataset [17]. "Rotation" sequence of the Inria Dataset was used for testing rotation-invariant LZMF whereas "Zoom" and "Zoom&Rotation" sequences of the Inria Dataset were used in order to evaluate repeatability of scale, rotation and translation-invariant R-LZMF. Each sequence contains some image sets and each image set comes with an image and its transformed versions. In each image set, there are also homography matrices representing the transformations between the original image and its transformed forms.

"Rotation" sequence has four image sets named as Marseil, Monet, NewYork and VanGogh. Marseil image set has 18 images with size of 842x842. Monet image set has similarly 18 images with size of 842x842. NewYork image set has 35 images and size of each one is 512x512. VanGogh image set has 17 images with size of 512x348. Rotation information in image sets of "Rotation" sequence, however, is not provided so that the images are indexed by integer numbers on the x-axis of repeatability plots instead of rotation angles. "Rotation" sequence is summarized in Table 4.1.

Table 4.1 : Image sets in "Rotation" sequence.

Image Set	Number of Images	Resolution
Marseil	18	842x842
Monet	18	842x842
NewYork	35	512x512
VanGogh	17	512x348

"Zoom" sequence has six image sets named as Belledonnes, Asterix, Crolles, Bip, VanGogh and Laptop. Asterix, Crolles and VanGogh were the only image sets used from this sequence throughout the experiments. Asterix image set has 17 images with size of 512x348. Crolles image set comes with 10 images whose sizes are 760x555. VanGogh image set has 17 images and each image has a size of 512x348. This sequence is to test scale-invariance but doesn't contain scale information about scaled transformations. Table 4.2 can be examined to have information about "Zoom" sequence.

Table 4.2 : Image sets in "Zoom" sequence.

Image Set	Number of Images	Resolution
Asterix	17	512x348
Crolles	10	760x555
VanGogh	17	512x348

"Zoom&Rotation" sequence has nine image sets named as Boat, East_park, East_south, Inria, Inria_model, Resid, UBC, Laptop and Ensimag. East_park, Laptop and Resid were considered for performance evaluation. East_park image set has 11 images with size of 850x680. Laptop image set has 13 images with size of 760x574. Resid image set contains 11 images whose sizes are 850x680. This sequence is used to evaluate detector performance against scale and rotation invariance but doesn't contain scale or rotation information about images. Table 4.3 contains information about "Zoom&Rotation" sequence.

Table 4.3 : Image sets in "Zoom&Rotation" sequence.

Image Set	Number of Images	Resolution
East_park	11	850x680
Laptop	13	760x574
Resid	11	850x680

4.3 Results

LZMF detector was tested on Marseil, Monet, NewYork and VanGogh image sets of "Rotation" sequence by comparing its repeatability scores with well-known interest point detectors such as Harris, SIFT, SURF, CenSurE (STAR) and BRISK. OpenCV v2.4.8 was used to work with these interest point detectors and default parameter settings were applied. Detection results for Marseil, Monet, NewYork and VanGogh image sets are compared in Figure 4.1 (a), (b), (c) and (d) respectively.

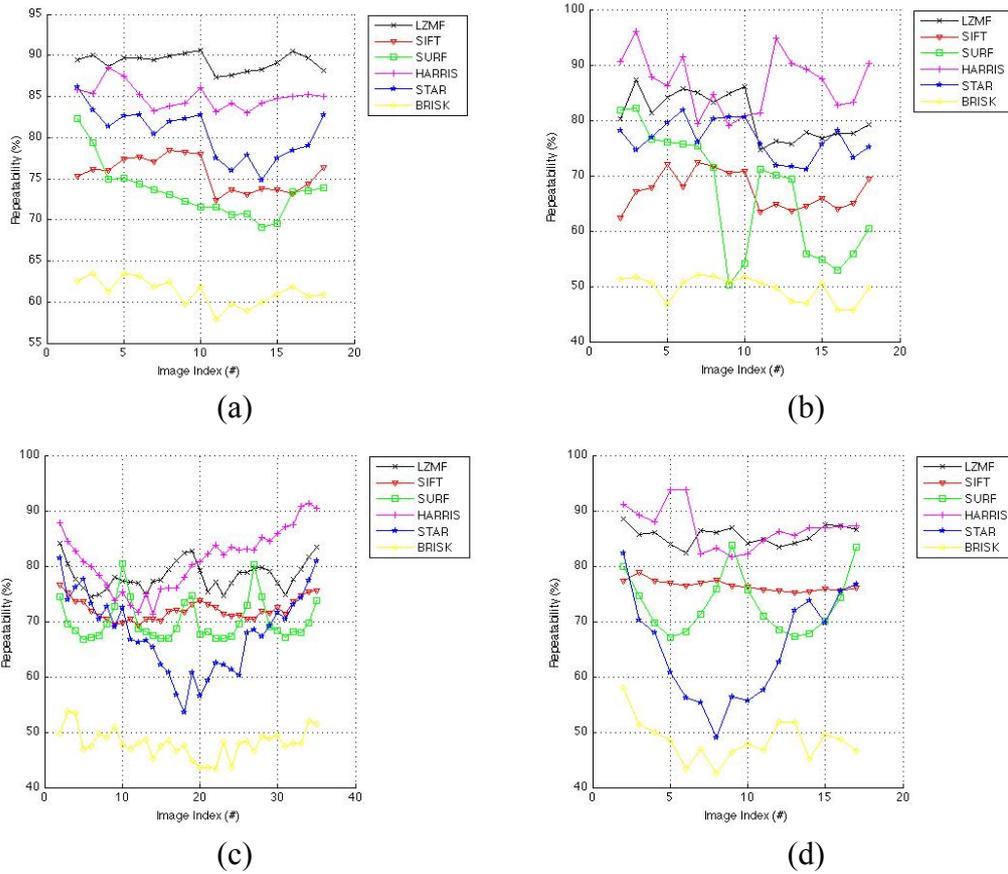


Figure 4.1 : Performance comparison of LZMF with other detectors.

LZMF shows the best performance with Marseille image set and outperforms all other detectors. It hits 90% of repeatability. In Monet image set, LZMF is passed by Harris but outperforms all other detectors. In NewYork image set, LZMF again performs as the second best detector and is only passed by Harris. In VanGogh image set, LZMF competes with Harris and it still outperforms all other detectors. Here, it never decreases below than 80% of repeatability. Average repeatabilities of LZMF and other detectors for Marseille, Monet, NewYork and VanGogh image sets can be examined in Figure 4.2 (a), (b), (c) and (d) respectively.

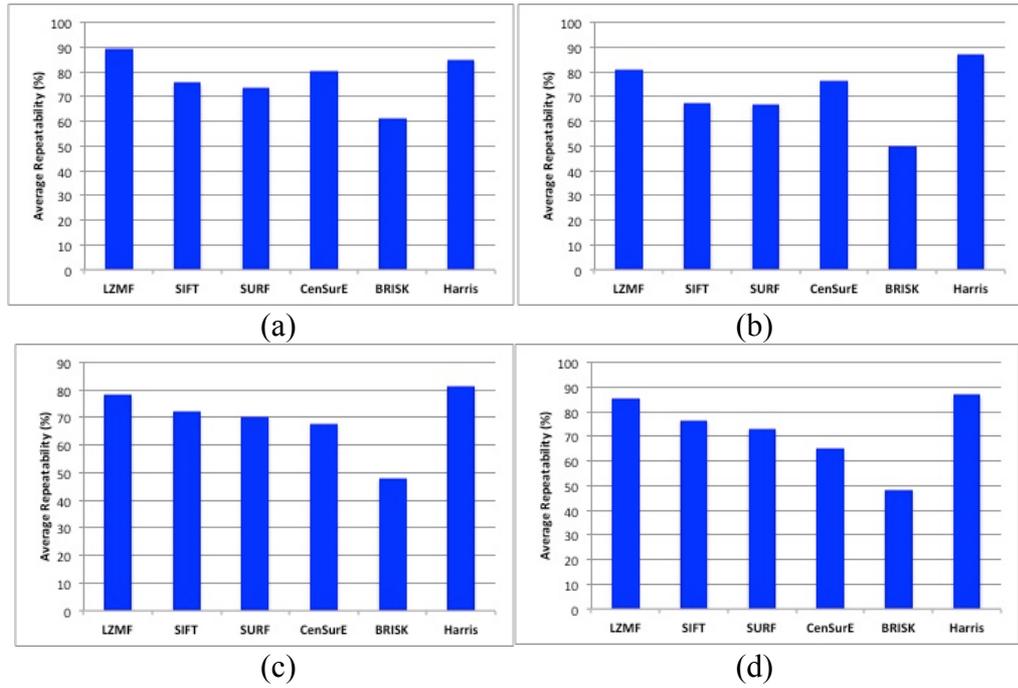


Figure 4.2 : Average repeatabilities of LZMF and other detectors.

LZMF detects large number of distinctive interest points and thus increases the chance of detecting objects in interest. For Marseil image set, it detects about 3000 keypoints. For Monet image set, this number decreases to about 500 because the scene has so much flat regions such as sky and frame. The number of detection for NewYork image set is about 1300 and the detection rate is about 1000 for VanGogh image set. In Table 4.4, the number of keypoints detected by interest point detectors in comparison is shown. Some detection results on images in Monet set can be examined in Figure 4.3. As a note, detected interest points are indicated as red circles for the images in Figure 4.3.

Table 4.4 : Numbers of keypoints detected by LZMF and others in Marseil set.

Detector	#kp in Base Image	#kp in Transformed image	Correspondence
LZMF	2915	2888	2573
SIFT	6130	5991	4487
SURF	4635	4446	3644
Harris	1000	1000	852
CenSurE	1386	1326	1142
BRISK	1514	1429	894

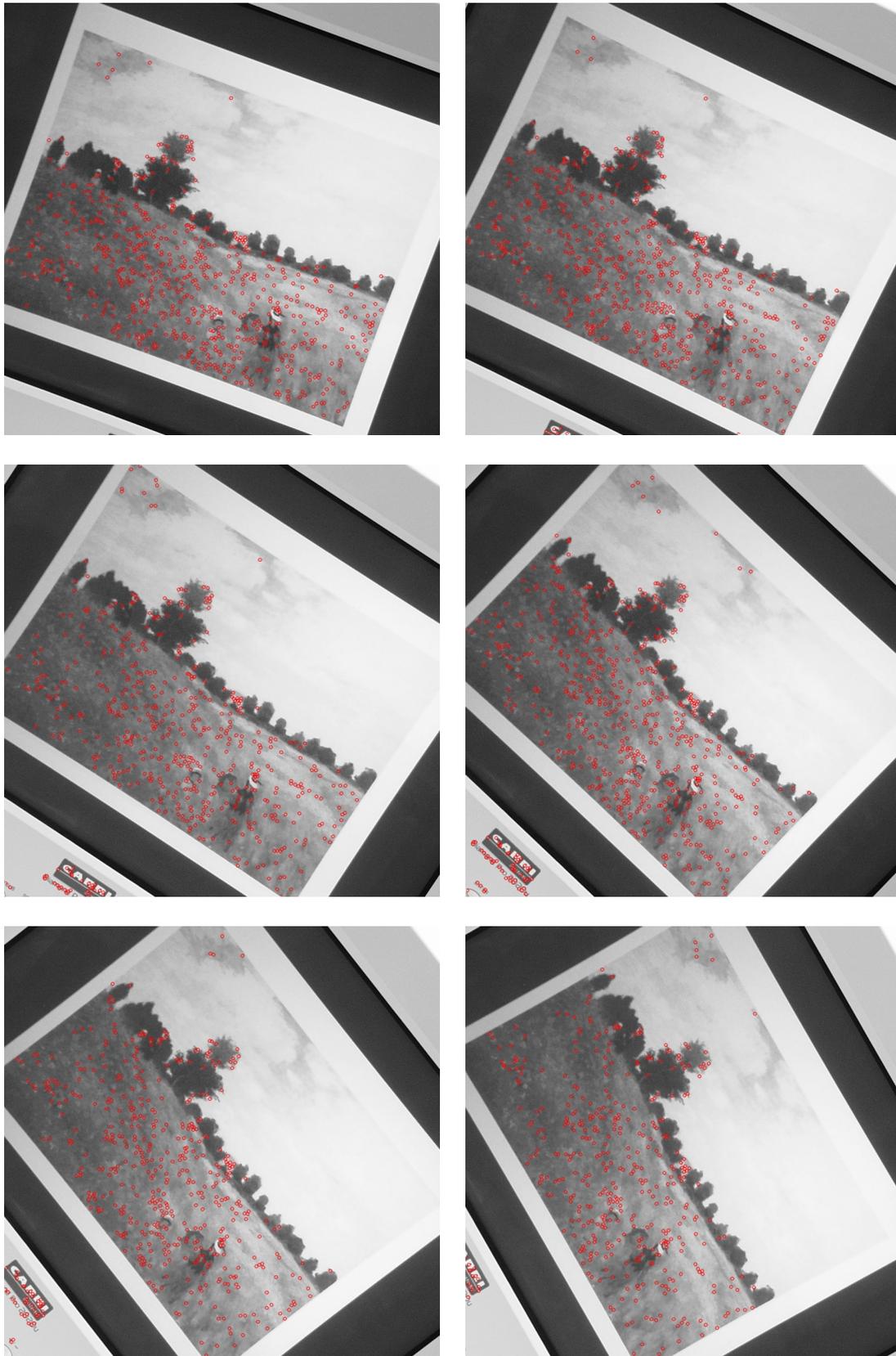


Figure 4.3 : Some detection results of LZMF on Monet image set.

The performance of R-LZMF detector was evaluated on Asterix, Crolles and VanGogh image sets from "Zoom" sequence and East_park, Laptop and Resid image sets from "Zoom&Rotation" sequence. R-LZMF was compared to SIFT, SURF, CenSurE (STAR), BRISK and ORB by using OpenCV v2.4.8 with default parameter settings again. Detection results for Asterix, Crolles, VanGogh, East_park, Laptop and Resid are shown in Figure 4.4 (a), (b), (c), (d), (e) and (f) respectively.

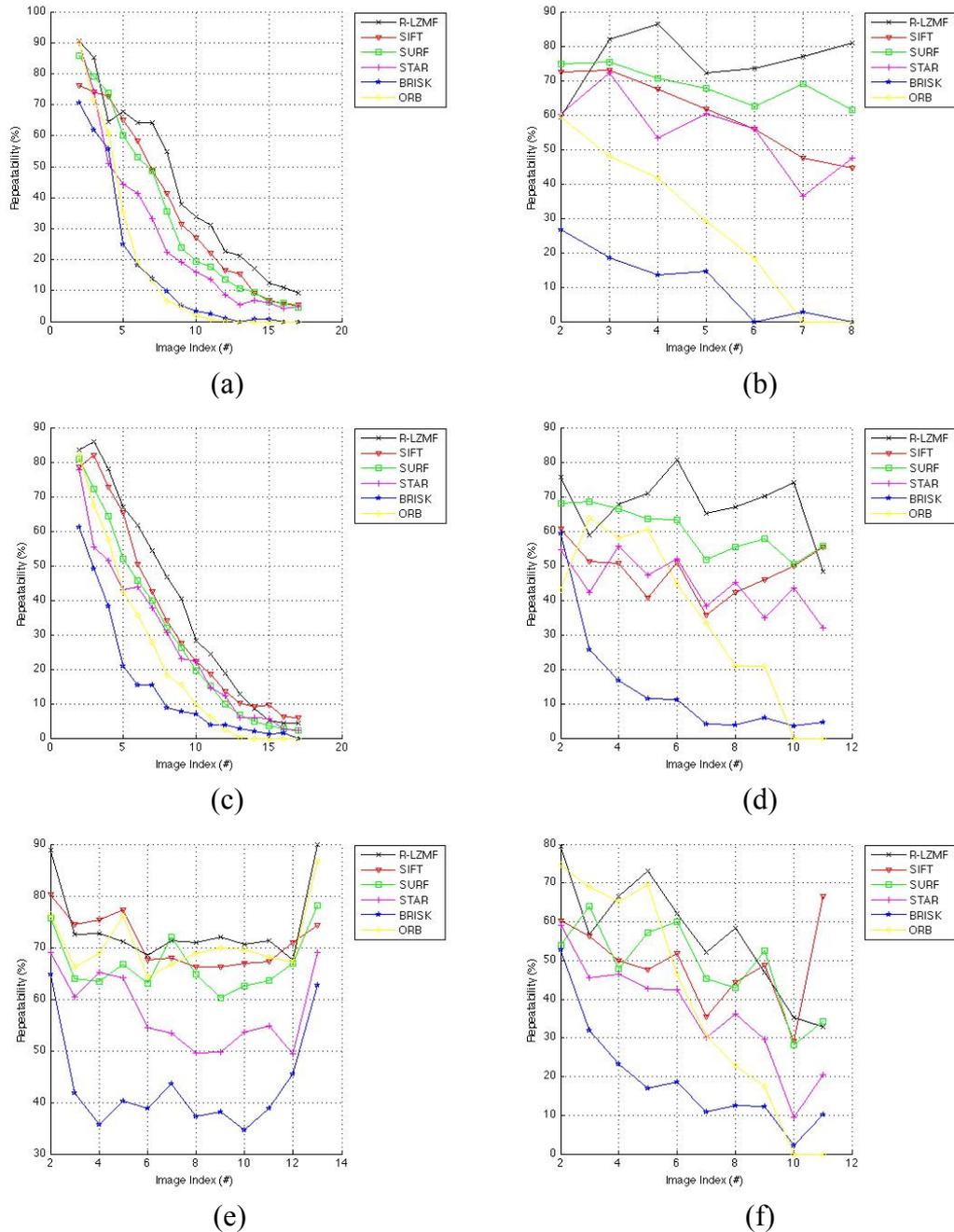


Figure 4.4 : Performance comparison of R-LZMF with other detectors.

R-LZMF outperforms all other detectors for all image sets as seen from Figure 4.4. This can also be seen from the bar charts in Figure 4.5 (a), (b), (c), (d), (e) and (f) for Asterix, Crolles, VanGogh, East_park, Laptop and Resid respectively.

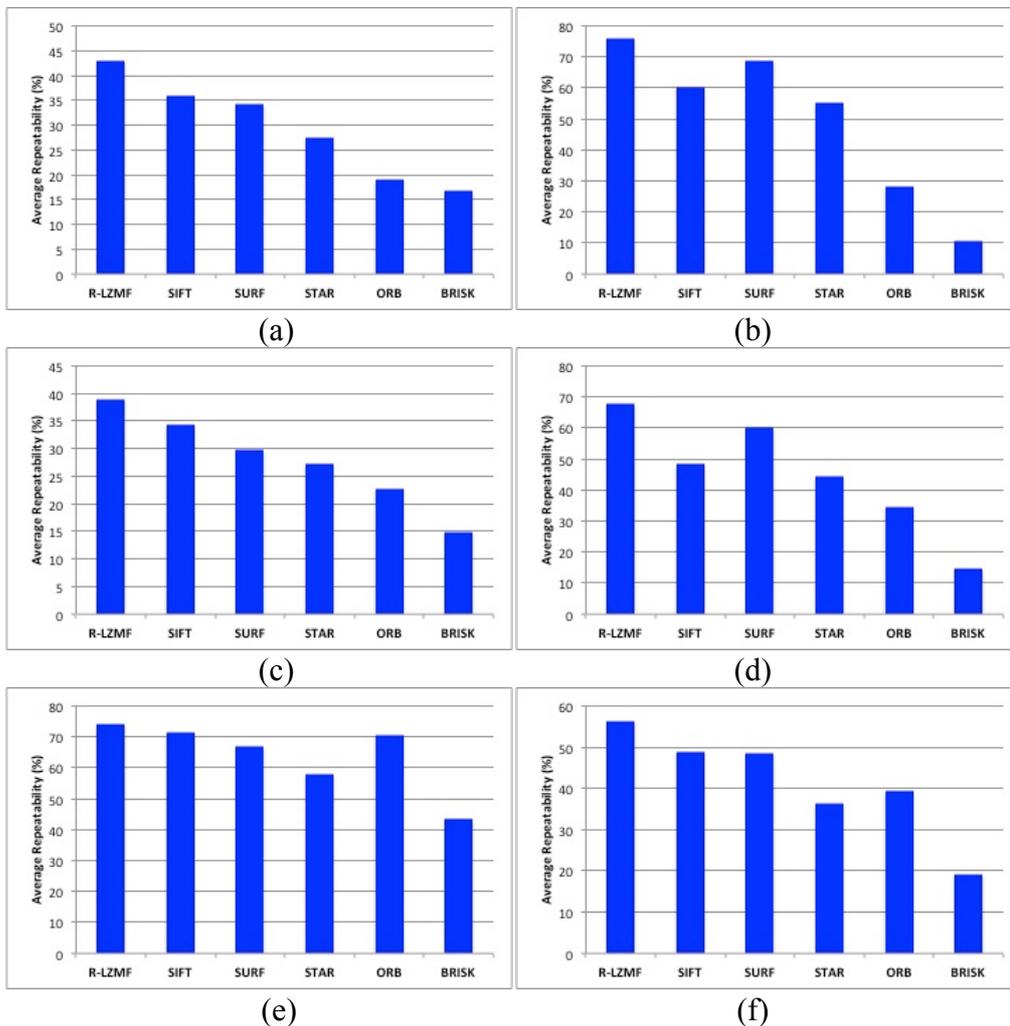


Figure 4.5 : Average repeatabilities of R-LZMF and other detectors.

The numbers of detected interest points by R-LZMF and other detectors are shown in Table 4.5. As seen from Table 4.5, R-LZMF generates enough number of interest points for Asterix set and R-LZMF's keypoints are more repeatable although they're fewer than the numbers of interest points detected by SIFT and SURF. In Figure 4.6 and Figure 4.7, some detection results on Crolles and Laptop image sets can be examined as red circles.

Table 4.5 : Numbers of keypoints detected by R-LZMF and others in Asterix set.

Detector	#kp in Base Image	#kp in Transformed image	Correspondence
R-LZMF	361	370	322
SIFT	603	601	457
SURF	643	650	552
STAR	251	258	226
BRISK	259	263	183
ORB	500	500	451

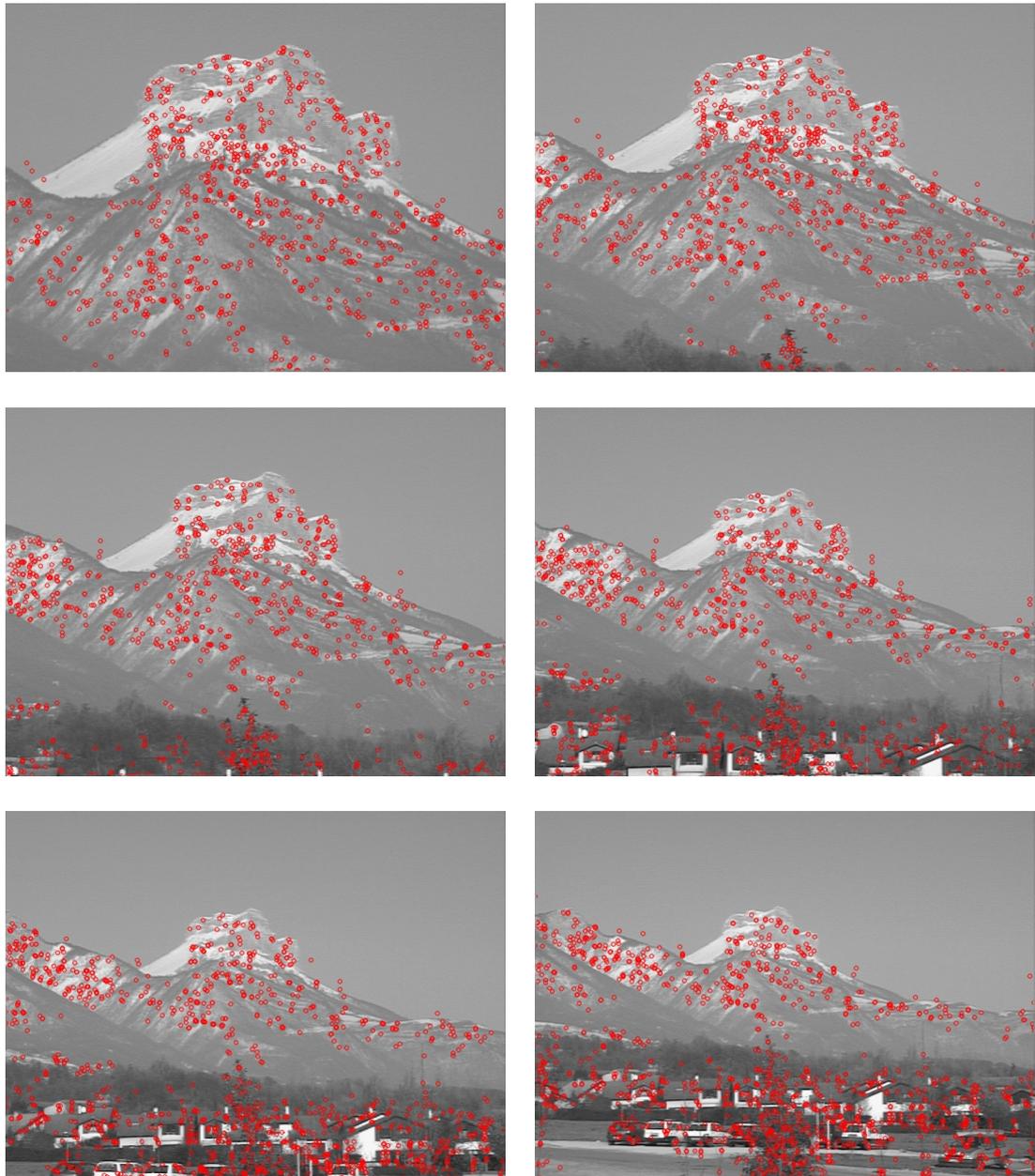


Figure 4.6 : Some detection results of R-LZMF on Crolles image set.

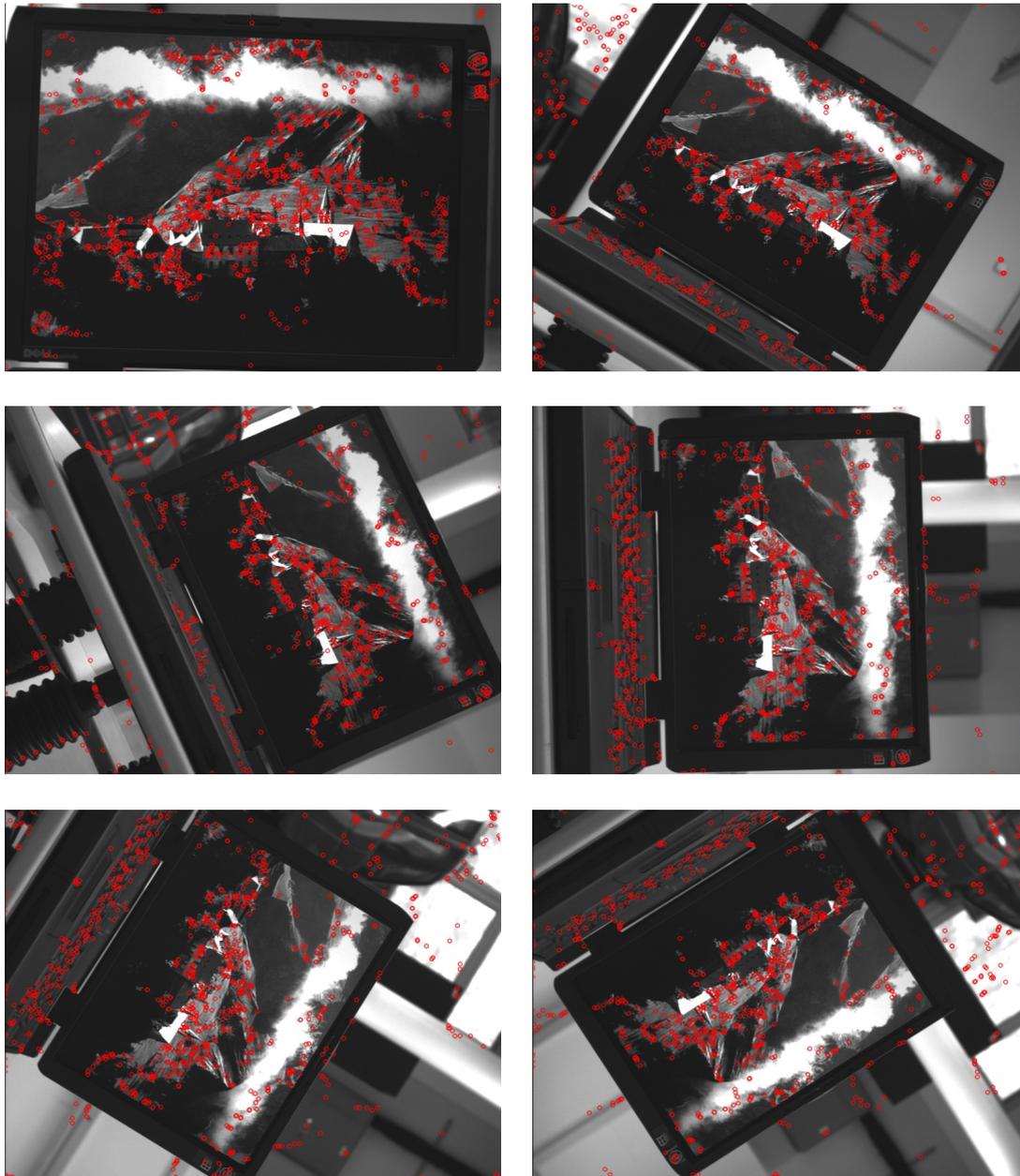


Figure 4.7 : Some detection results of R-LZMF on Laptop image set.

5. CONCLUSIONS AND FUTURE WORK

In this chapter, the thesis is concluded by summarizing the research conducted in order to design robust interest point detection methods and highlighting experimental results of proposed interest point detection algorithms. Future directions are also discussed in order to improve and extend proposed algorithms.

5.1 Conclusions

In this thesis, two novel interest point detection algorithms, Local Zernike Moment based Features (LZMF) and Robust Local Zernike Moment based Features (R-LZMF), are proposed by focusing on image matching problem that is widely considered in computer vision applications such as object recognition, stereo vision and motion tracking. Proposed interest point detection algorithms realize corner-based detections in order to detect interesting structures in images and apply some other mechanisms before and after corner detection such as global/local normalization and non-maximum suppression. An image pyramid is also constructed in order to sample the image in scale-space and expose the structures in different scales.

Robust interest point detection is very important when dealing with image matching problem in computer vision. Corresponding points should be detected as much as possible with minimum missing rate in order to be successful in further processing. In practice, because of geometric and photometric transformations occurred in the images taken from same scene under different conditions, detection of such corresponding points with high rates is not easy. At this point, robustness of interest point detector against these transformations comes into play. An interest point, which handles such transformations with special mechanisms, increases the chance of detecting interesting points as much as possible and thus contributes to the problem in interest significantly.

Proposed rotation and translation-invariant interest point detection algorithm, LZMF, utilizes from discriminative power of Zernike moments and searches for corners with these moments in the image. Proposed scale, rotation and translation-invariant interest point detection algorithm, R-LZMF, uses LZMF in order to detect keypoints in spatial-space and analyzes these points in scale-space for determining characteristic scales. Interest points detected by R-LZMF are local features that can be further described in size of characteristic scale. Description of the detected interest points, however, is out of scope in this thesis and will be further examined in details. LZMF and R-LZMF interest point detection algorithms are the main contributions of this thesis.

The repeatability performance, which is the main criterion for evaluating interest point detectors, of proposed algorithms shows their superiority on keypoint detection. On the Inria Dataset, LZMF outperforms SIFT, SURF, CenSurE and BRISK, and it also competes with Harris whereas R-LZMF outperforms SIFT, SURF, CenSurE, BRISK and ORB in every case based on repeatability. The numbers of interest points detected by LZMF and R-LZMF are also well enough in order to tackle background clutter and occlusion problems.

5.2 Future Work

The repeatability performance of R-LZMF interest point detection algorithm should be analyzed for affine transformation as well in order to say that it's affine-invariant. Theoretically, R-LZMF should be partially affine-invariant because it's invariant to scale, rotation and translation, and an affine transformation is combination of these transformations. However, this fact should be verified by evaluating repeatability performance of R-LZMF with affine specific datasets.

Proposed interest point detection algorithms, LZMF and R-LZMF, have computational load especially during getting Zernike moment representations. As a result of this, LZMF and R-LZMF are not as fast as SIFT and SURF. Quantized Zernike moments can be applied in local sense to boost the detection performance in speed. LZMF and R-LZMF algorithms can also be parallelized in specific Graphics Processing Unit (GPU) platforms such as Compute Unified Device Architecture (CUDA).

LZMF and R-LZMF algorithms can also be extended to have a descriptor so that they present a complete schema, which has both detector and descriptor, as in SIFT and SURF. For description, the descriptive power of local Zernike moment representation can be utilized or some well-known descriptors such as Histogram of Oriented Gradients (HoG) can be used. Thus, by combining high repeatability of proposed interest point detectors with discriminative description power, a complete and robust local feature extraction schema may be proposed.

REFERENCES

- [1] **Harris, C. and Stephens, M.** (1988). A combined corner and edge detector. *Alvey Vision Conference*, pp. 147-151.
- [2] **Witkin, A.P.** (1983). Scale-space filtering. *International Joint Conference on Artificial Intelligence*, Karlsruhe, Germany, pp. 1019–1022.
- [3] **Koenderink, J.J.** (1984). The structure of images. *Biological Cybernetics*, 50:363–396.
- [4] **Lindeberg, T.** (1998). Feature detection with automatic scale selection. *International Journal of Computer Vision*, 30(2):79-116.
- [5] **Lowe, D.G.** (1999). Object recognition from local scale-invariant features. *International Conference on Computer Vision*, Corfu, Greece, pp. 1150-1157.
- [6] **Mikolajczyk, K. and Schmid, C.** (2001). Indexing based on scale invariant interest points. *International Conference on Computer Vision*, pp. 525-531.
- [7] **Schmid, C., Mohr, R., Bauckhage, C.** (1998). Comparing and evaluating interest points. *International Conference on Computer Vision*, pp. 230-235.
- [8] **Bay, H., Ess, A., Tuytelaars, T., Gool, L.V.** (2008). SURF: Speeded Up Robust Features. *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346- 359.
- [9] **Rosten, E., Drummond, T.** (2006). Machine learning for high-speed corner detection. *European Conference on Computer Vision*, pp. 430-443.
- [10] **Rublee, E., Rabaud, V., Konolige, K., Bradski, G.** (2011). ORB: an efficient alternative to SIFT or SURF. *International Conference on Computer Vision*, pp. 2564-2571.
- [11] **Calonder, M., Lepetit V., Strecha, C., Fua, P.** (2010). BRIEF: Binary Robust Independent Elementary Features. *European Conference on Computer Vision*, pp. 778-792.
- [12] **Agrawal, M., Konolige, K., Blas, M. R.** (2008). CenSurE: Center surround extremas for realtime feature detection and matching. *European Conference on Computer Vision*, pp. 102-115.
- [13] **Leutenegger, S., Chli, M., Siegwart R.** (2011). BRISK: Binary Robust Invariant Scalable Keypoints. *International Conference on Computer Vision*, pp. 2548-2555.

- [14] **Khotanzad, A., Hong, Y. H.** (1990). Invariant image recognition by Zernike moments. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 12, pp. 489-497.
- [15] **Ghosal, S., Mehrotra, R.** (1997). A moment-based unified approach to image feature detection. *IEEE Trans. Image Processing*, vol. 6, no. 6, pp. 781-793.
- [16] **Sariyanidi, E., Dagli, V., Tek, S.C., Tunc, B., Gokmen, M.** (2012). Local Zernike Moments: A new representation for face recognition. *International Conference on Image Processing*, pp. 585-588.
- [17] **Url-1**, <http://lear.inrialpes.fr/people/mikolajczyk/Database/>, date retrieved 19.11.2014.
- [18] **Flusser J., Suk T., Zitova B.** (2009). Moments and Moment Invariants in Pattern Recognition. Wiley Press.
- [19] **Liao S. X.** (1993). Image Analysis by Moments, PhD dissertation, University of Manitoba.
- [20] **Zernike F.** (1934). *Physica*, vol. 1.
- [21] **Teague, M.R.** (1980). Image analysis via the general theory of moments, *J. Optical Soc. Am.*, Vol. 70, pp.920-930.
- [22] **Url-2**, http://docs.opencv.org/trunk/doc/py_tutorials/py_feature2d/py_sift_intro/py_sift_intro.html, date retrieved 13.12.2014.
- [23] **Url-3**, http://en.wikipedia.org/wiki/Aberrations_of_the_eye, date retrieved 14.12.2014.
- [24] **Lowe, D.G.** (2004). Distinctive image features from scale invariant keypoints. *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91-110.
- [25] **Mikolajczyk, K. and Schmid, C.** (2002). An affine invariant interest point detector. *European Conference on Computer Vision*, pp. 128-142.
- [26] **Mikolajczyk, K. and Schmid, C.** (2004). Scale and affine invariant interest point detectors. *International Journal of Computer Vision*, vol. 60, no. 2, pp. 63-86.

CURRICULUM VITAE

Name Surname: Gökhan Özbulak

Place and Date of Birth: Istanbul, 1984

E-Mail: gokhanozbulak@gmail.com

EDUCATION:

- **B.Sc.:** 2010, Dokuz Eylul University, Engineering Faculty, Computer Engineering

PROFESSIONAL EXPERIENCE AND REWARDS:

- Researcher, The Scientific and Technological Research Council of Turkey (TUBITAK)
- DB2 System Tester, IBM Toronto Software Laboratory (IBM Canada Ltd.)
- Best 2nd project, Izmir EMO Competition, 2010

PUBLICATIONS, PRESENTATIONS AND PATENTS ON THE THESIS:

- **Özbulak G.**, Gökmen M. (2014). A Rotation Invariant Local Zernike Moment based Interest Point Detector. *Proc. SPIE of 7th International Conference on Machine Vision (ICMV)*, November 19-21, Milano, Italy.
- **Özbulak G.**, Gökmen M. (2015). Robust Interest Point Detection by Local Zernike Moments. *10th International Conference on Computer Vision Theory and Applications (VISAPP)*, March 11-14, Berlin, Germany.

OTHER PUBLICATIONS, PRESENTATIONS AND PATENTS:

- Birant D., **Özbulak G.**, Teyran I., Serin T. (2010). Dynamic Bus Scheduling with FP-Growth. *International Symposium on INnovations in Intelligent SysTems and Applications (INISTA)*, pp. 491-495, June 21-24, Kayseri, Turkey.