

A COMPARATIVE PERFORMANCE EVALUATION OF SCALE INVARIANT
INTEREST POINT DETECTORS FOR INFRARED AND VISUAL IMAGES

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

ERDEM EMİR

IN PARTIAL FULLFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
ELECTRICAL AND ELECTRONICS ENGINEERING

DECEMBER 2008

Approval of the thesis:

**A COMPARATIVE PERFORMANCE EVALUATION OF SCALE
INVARIANT INTEREST POINT DETECTORS FOR INFRARED AND
VISUAL IMAGES**

submitted by **ERDEM EMİR** in partial fulfillment of the requirements for the degree of **Master of Science in Electrical and Electronics Engineering Department, Middle East Technical University** by,

Prof. Dr. Canan Özgen
Dean, Graduate School of **Natural and Applied Sciences** _____

Prof. Dr. İsmet Erkmen
Head of Department, **Electrical and Electronics Engineering** _____

Assoc. Prof. Dr. A.Aydın Alatan
Supervisor, **Electrical and Electronics Engineering Dept., METU** _____

Examining Committee Members:

Prof. Dr. Kemal Leblebicioğlu
Electrical and Electronics Engineering Dept., METU _____

Assoc. Prof. Dr. A.Aydın Alatan
Electrical and Electronics Engineering Dept., METU _____

Prof. Dr. Gözde Bozdağı Akar
Electrical and Electronics Engineering Dept., METU _____

Assoc. Prof. Dr. Tolga Çiloğlu
Electrical and Electronics Engineering Dept., METU _____

Yelda Altuğ (MSc.)
ASELSAN _____

Date: 04.12.2008

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name : Erdem Emir

Signature :

ABSTRACT

A COMPARATIVE PERFORMANCE EVALUATION OF SCALE INVARIANT INTEREST POINT DETECTORS FOR INFRARED AND VISUAL IMAGES

Emir, Erdem

M. Sc., Department of Electrical and Electronics Engineering

Supervisor: Assoc. Prof. Dr. A.Aydın Alatan

December 2008, 206 pages

In this thesis, the performance of four state-of-the-art feature detectors along with SIFT and SURF descriptors in matching object features of mid-wave infrared, long-wave infrared and visual-band images is evaluated across viewpoints and changing distance conditions. The utilized feature detectors are Scale Invariant Feature Transform (SIFT), multiscale Harris-Laplace, multiscale Hessian-Laplace and Speeded Up Robust Features (SURF) detectors, all of which are invariant to image scale and rotation. Features on different blackbodies, human face and vehicle images are extracted and performance of reliable matching is explored between different views of these objects each in their own category. All of these feature detectors provide good matching performance results in infrared-band images compared with visual-band images. The comparison of matching performance for mid-wave and long-wave infrared images is also explored in this study and it is observed that long-wave infrared images provide good matching performance for objects at lower temperatures, whereas mid-wave infrared-band images provide good matching performance for objects at higher temperatures. The matching

performance of SURF detector and descriptor for human face images in long-wave infrared-band is found to be outperforming than other detectors and descriptors.

Keywords: Feature Detectors, Local Descriptors, Interest Regions, Invariance, Infrared, Visual, Matching, Recognition.

ÖZ

BOYUTTAN BAĞIMSIZ ÖZNETELİK/ÖZELLİK ÇIKARIM DETEKTÖRLERİNİN KIZILÖTESİ VE GÖRÜNÜR BANT RESİMLERİ İÇİN KARŞILAŞTIRMALI PERFORMANS DEĞERLENDİRMESİ

Emir, Erdem

Yüksek Lisans, Elektrik Elektronik Mühendisliği Bölümü

Tez Yöneticisi: Doç. Dr. A. Aydın Alatan

Aralık 2008, 206 sayfa

Bu tezde, dört tane popüler öznetelik çıkarım detektörlerinin SIFT ve SURF tanımlayıcıları ile birlikte kullanılarak elde edilen ortadalga, uzundalga kızılötesi ve görünür bant resim nesne özelliklerinin eşleme performansı, değişen açı ve uzaklıklara göre değerlendirilmiştir. Bu çalışmada kullanılan öznetelik çıkarım detektörleri, görüntü boyut ve döndürülmesinden bağımsız olan, ölçek değişmez öznetelik transformasyonu (SIFT), çok ölçekli Harris-Laplace, çok ölçekli Hessian-Laplace ve hızlandırılmış gürbüz özellikler çıkarım (SURF) detektörleridir. Farklı siyahcisimler, insan yüzü ve araba resimlerinin özellikleri çıkarılmış ve bu resimlerin, cisimlerin değişen pozlarına göre özellik eşleme performansları değerlendirilmiştir. Kullanılan bütün öznetelik çıkarım detektörleri, görünür bant resimleri ile karşılaştırıldığında kızılötesi bant resimlerinde iyi eşleme performansı göstermiştir. Ortadalga ve uzundalga kızılötesi resimlerinin eşleme performansı karşılaştırılmış ve uzundalga kızılötesi resimlerin sıcaklığı düşük nesnelere için daha iyi eşleme performansı sağladığı, ortadalga kızılötesi resimlerin ise sıcaklığı daha yüksek cisimler için daha iyi eşleme performansı sağladığı görülmüştür. SURF

detektör ve tanımlayıcıları, uzundalga kızılötesi bantta alınmış insan yüzü resimlerinde, diğer detektör ve tanımlayıcılarına göre daha iyi eşleme performansı sağlamaktadır.

Anahtar Kelimeler: Öznitelik/Özellik Çıkarım Detektörleri, Yerel Tanımlayıcılar, Öznitelik Bölgeleri, Değişmezlik, Kızılötesi, Görünür, Eşleme, Tanıma.

To My Family

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my supervisor Assoc. Prof. Dr. A. Aydın Alatan for his encouragements, guidance, advice, criticism and insight throughout the research.

I would like to thank ASELSAN Inc. for facilities provided for the completion of this thesis.

I would like to forward my appreciation to all my friends and colleagues who contributed to my thesis with their continuous encouragement.

I would also like to express my profound appreciation to my family for their continuous support.

TABLE OF CONTENTS

ABSTRACT	iv
ÖZ	vi
ACKNOWLEDGEMENTS	ix
TABLE OF CONTENTS	x
LIST OF TABLES	xii
LIST OF FIGURES	xiii
LIST OF ABBREVIATIONS	xxi
CHAPTERS	
CHAPTER 1 INTRODUCTION	1
1.1 Background	1
1.2 Motivation and Scope of Thesis	2
CHAPTER 2 RADIOMETRY AND OPTICAL IMAGING SYSTEMS.....	4
2.1 Fundamentals of Radiometry	4
2.2 Optical Imaging Systems	12
2.2.1 Imaging in Infrared Bands	15
2.2.2 Imaging In Visual Bands.....	18
CHAPTER 3 INTEREST POINT DETECTION AND DESCRIPTION.....	21
3.1 Interest Point Detectors in Visual and Infrared Bands : Previous Studies	23
3.1.1 Interest Point Detectors in Visual Bands	23
3.1.2 Interest Point Detectors in Infrared Bands	27
3.2 SIFT Detector	30
3.3 Multiscale Harris-Laplace Detector	52
3.4 Multiscale Hessian-Laplace Detector	69
3.5 SURF Detector	81
3.6 Orientation of Interest Points	92
3.7 Local Region Descriptors.....	99
3.8 Matching Accuracy of Feature Points.....	104
CHAPTER 4 SIMULATIONS AND RESULTS	107
4.1 Infrared Band Images.....	108

4.2 LWIR versus MWIR Infrared Images Experiment.....	109
4.2.1 Experimental Setup	109
4.2.2 Results of LWIR versus MWIR Experiment	117
4.3 LWIR versus Visual Band Images Experiment	123
4.3.1 Experimental Setup	124
4.3.2 Results of LWIR versus Visual Band Images Experiment	130
4.4 Matching Performance for Different Rotation Angles.....	152
4.4.1 Experimental Setup	152
4.4.2 Results of Matching Performance with Different Rotations Experiment	159
4.5 Matching Performance with Changing Distance Experiment.....	183
4.5.1 Experimental Setup	183
4.5.2 Results of Matching Performance with Changing Distance Experiment	186
CHAPTER 5 CONCLUSIONS and DISCUSSIONS	197
REFERENCES.....	201

LIST OF TABLES

TABLES

Table 2-1 Summary for the Radiometric Terms	8
Table 3-1 SIFT Algorithm Steps.....	50
Table 3-2 Harris-Laplace Algorithm Steps	67
Table 3-3 Hessian-Laplace Algorithm Steps	78
Table 3-4 SURF Algorithm Steps.....	90
Table 4-1 Integration Times for MWIR Camera	114
Table 4-2 Matching Accuracy Values, SIFT Detector, Experiment 3.....	172
Table 4-3 Matching Accuracy Values, Harris-Laplace Detector, Experiment 3 ...	176
Table 4-4 Matching Accuracy Values, Hessian-Laplace Detector, Experiment 3	180
Table 4-5 Matching Accuracy Values, SURF Detector, Experiment 3	183
Table 4-6 Matching Accuracy Values, SIFT Detector, Experiment 4.....	195
Table 4-7 Matching Accuracy Values, Harris-Laplace Detector, Experiment 4 ...	195
Table 4-8 Matching Accuracy Values, Hessian-Laplace Detector, Experiment 4	196
Table 4-9 Matching Accuracy Values, SURF Detector, Experiment 4	196

LIST OF FIGURES

FIGURES

Figure 2-1 Electromagnetic Spectrum	6
Figure 2-2 Radiometric Terms-Visual Illustration.....	7
Figure 2-3 Optical Portion of Electromagnetic Radiation Including Infrared Part..	10
Figure 2-4 Spectral Emittance of A Blackbody at 500°C (Typical)	12
Figure 2-5 Simple Model of Image Formation	14
Figure 2-6 “Pinhole” Setup for Image Formation in Visual Band.....	18
Figure 2-7 Nipkow Disk	19
Figure 3-1 Sample Input Image	32
Figure 3-2 Original Sample Image and Double Size Presmoothed Starting Image.	33
Figure 3-3 First Octave First Three Levels (1, 2, 3) of Scale	34
Figure 3-4 First Octave Second Three Levels (4, 5, 6) of Scale.....	34
Figure 3-5 Second Octave First Three Levels (1, 2, 3) of Scale.....	35
Figure 3-6 Second Octave Second Three Levels (4, 5, 6) of Scale	35
Figure 3-7 Third Octave First Three Levels (1, 2, 3) of Scale.....	36
Figure 3-8 Third Octave Second Three Levels (4, 5, 6) of Scale	36
Figure 3-9 Fourth Octave First Three Levels (1, 2, 3) of Scale.....	37
Figure 3-10 Fourth Octave Second Three Levels (4, 5, 6) of Scale.....	37
Figure 3-11 Illustration of Characteristic Scale [2].	38
Figure 3-12 Difference of Gaussian Graphical View-1 for Sample Input Image in Figure 3-1	40
Figure 3-13 Difference of Gaussian Graphical View-2 for Sample Input Image in Figure 3-1	41
Figure 3-14 Graph of 1D Gaussian $G(x)$ and DoG(x)	42

Figure 3-15 Graph of 2D DoG(x,y)	42
Figure 3-16 DoG Images for First Octave First Two DoG Scale Levels.....	43
Figure 3-17 DoG Images for First Octave Second Two DoG Scale Levels	44
Figure 3-18 Local Extremum Point Search Over Scales and Space [1].	45
Figure 3-19 Initial Candidate SIFT Features	46
Figure 3-20 Typical Graph of True Extrema by Localization and Refinement [49].	47
Figure 3-21 Final SIFT Features After Refinement.....	49
Figure 3-22 Diagram of Harris-Laplace Detection [51].	52
Figure 3-23 First Order Gaussian Derivative Images-1, Lx	55
Figure 3-24 First Order Gaussian Derivative Images-2, Lx	56
Figure 3-25 First Order Gaussian Derivative Images-1, Ly	56
Figure 3-26 First Order Gaussian Derivative Images-2, Ly	57
Figure 3-27 Initial Harris-Laplace Feature Points	60
Figure 3-28 3D Graph of Scale-normalized Laplacian of Gaussian	62
Figure 3-29 Side View of Scale-normalized Laplacian of Gaussian	63
Figure 3-30 Scale-normalized Laplacian of Gaussian Filtered Images - 1	64
Figure 3-31 Scale-normalized Laplacian of Gaussian Filtered Images - 2.....	65
Figure 3-32 Final Harris-Laplace Feature Points.....	66
Figure 3-33 Diagram of Hessian-Laplace Detection [51].....	70
Figure 3-34 Second Order Gaussian Derivative Image-1, Lxx	71
Figure 3-35 Second Order Gaussian Derivative Image-2, Lxx	72
Figure 3-36 Second Order Gaussian Derivative Image-3, Lyy	72
Figure 3-37 Second Order Gaussian Derivative Image-4, Lyy	73
Figure 3-38 Second Order Gaussian Derivative Image-5, Lxy	73
Figure 3-39 Second Order Gaussian Derivative Image-6, Lxy	74
Figure 3-40 Initial Hessian-Laplace Feature Points.....	76
Figure 3-41 Final Hessian-Laplace Feature Points	77
Figure 3-42 Rectangular Region for Integral Image.....	82
Figure 3-43 Calculation of an Arbitrary Rectangle Sum [50].....	83
Figure 3-44 Box Filter for Second Order Partial Gaussian Derivative in y-direction [50].	85

Figure 3-45 Box Filter for Second Order Partial Gaussian Derivative in xy-direction [50].	85
Figure 3-46 Scale Space Structure of SURF Feature Detector [50].	87
Figure 3-47 Final SURF Feature Points.....	89
Figure 3-48 Typical Orientation Histogram Diagram [49]......	94
Figure 3-49 SIFT Features With Orientations	95
Figure 3-50 Harris-Laplace Feature Points with Orientations	96
Figure 3-51 Hessian-Laplace Feature Points with Orientations	97
Figure 3-52 Haar Wavelet Types [50].	98
Figure 3-53 Fractional Displacements and Orientation Invariance [47]......	101
Figure 3-54 4x4x8 Typical SIFT Descriptor [49]......	102
Figure 3-55 Diagram of SURF Descriptor [50]......	103
Figure 4-1 White Cartoon Mask of Experiment 1	111
Figure 4-2 Blackbody Device without Mask (Experiment 1).....	111
Figure 4-3 Blackbody Device with Mask (Experiment 1).....	112
Figure 4-4 Typical Infrared Image of Blackbody Device with Mask (Experiment 1)	112
Figure 4-5 LWIR Band Infrared Images, 32°C Blackbody, 4m.	115
Figure 4-6 MWIR Band Infrared Images, 32°C Blackbody, 4m.	115
Figure 4-7 LWIR Band Infrared Images, 112°C Blackbody, 4m.	116
Figure 4-8 MWIR Band Infrared Images, 112°C Blackbody, 4m.	116
Figure 4-9 LWIR 112°C Infrared Image Feature Points Extracted by SIFT Detector.	117
Figure 4-10 LWIR 112°C Infrared Image Feature Points Extracted by Harris- Laplace Detector.....	118
Figure 4-11 LWIR 112°C Infrared Image Feature Points Extracted by Hessian- Laplace Detector.....	118
Figure 4-12 LWIR 112°C Infrared Image Feature Points Extracted by SURF Detector.	119
Figure 4-13 Matched Feature Points of LWIR 112°C Infrared Images, SIFT detector.	120

Figure 4-14 Matched Feature Points of LWIR 112°C Infrared Images, Harris-Laplace detector.	120
Figure 4-15 Matched Feature Points of LWIR 112°C Infrared Images, Hessian-Laplace detector.	121
Figure 4-16 Matched Feature Points of LWIR 112°C Infrared Images, SURF detector.	121
Figure 4-17 Matching Accuracy with Changing Temperature-LWIR Band	122
Figure 4-18 Matching Accuracy with Changing Temperature-MWIR Band	123
Figure 4-19 LWIR Band Infrared Images of Face, 4m, Slided.....	125
Figure 4-20 LWIR Band Infrared Images of Face, 4m, Turned to left.....	125
Figure 4-21 Visual Band Images of Face, 4m, Slided	126
Figure 4-22 Visual Band Images of Face, 4m, Turned to right	126
Figure 4-23 LWIR Band Infrared Images of Face for 4m and 8m.	127
Figure 4-24 LWIR Band Infrared Images of Face for 4m and 12m.	127
Figure 4-25 Visual Band Images of Face for 4m and 8m.	128
Figure 4-26 Visual Band Images of Face for 4m and 12m.	128
Figure 4-27 LWIR Band Infrared Images of Vehicle, 0 degree and 5 degree.	129
Figure 4-28 LWIR Band Infrared Images of Vehicle, 0 degree and 45 degree.	129
Figure 4-29 Visual Band Images of Vehicle, 0 degree and 5 degree.....	130
Figure 4-30 Visual Band Images of Vehicle, 0 degree and 45 degree.....	130
Figure 4-31 LWIR Band Face Image Feature Points Extracted by SIFT Detector.	131
Figure 4-32 LWIR Band Face Image Feature Points Extracted by Harris-Laplace Detector.	132
Figure 4-33 LWIR Band Face Image Feature Points Extracted by Hessian-Laplace Detector.	133
Figure 4-34 LWIR Band Face Image Feature Points Extracted by SURF Detector.	133
Figure 4-35 Visual Band Face Image Feature Points Extracted by SIFT Detector.	134
Figure 4-36 Visual Band Face Image Feature Points Extracted by Harris-Laplace Detector.	135

Figure 4-37 Visual Band Face Image Feature Points Extracted by Hessian-Laplace Detector.....	135
Figure 4-38 Visual Band Face Image Feature Points Extracted by SURF Detector.....	136
Figure 4-39 LWIR Band Vehicle Image Feature Points Extracted by SIFT Detector.....	136
Figure 4-40 LWIR Band Vehicle Image Feature Points Extracted by Harris-Laplace Detector.....	137
Figure 4-41 LWIR Band Vehicle Image Feature Points Extracted by Hessian-Laplace Detector.....	137
Figure 4-42 LWIR Band Vehicle Image Feature Points Extracted by SURF Detector.....	138
Figure 4-43 Matched Feature Points of LWIR Band Face Images, SIFT Detector.....	139
Figure 4-44 Matched Feature Points of LWIR Band Face Images, Harris-Laplace Detector.....	139
Figure 4-45 Matched Feature Points of LWIR Band Face Images, Hessian-Laplace Detector.....	140
Figure 4-46 Matched Feature Points of LWIR Band Face Images, SURF Detector.....	140
Figure 4-47 False Matches of LWIR Band Face Images, SIFT Detector.....	141
Figure 4-48 False Matches of LWIR Band Face Images, Harris-Laplace Detector.....	142
Figure 4-49 False Matches of LWIR Band Face Images, Hessian-Laplace Detector.....	142
Figure 4-50 False Matches of LWIR Band Face Images, SURF Detector.....	143
Figure 4-51 Matching Accuracy of Face Images with Changing Rotation Angle-LWIR Band	144
Figure 4-52 Matching Accuracy of Face Images with Changing Rotation Angle-Visual Band	144
Figure 4-53 Matching Accuracy of Face Images with Changing Distance-LWIR Band.....	146

Figure 4-54 Matching Accuracy of Face Images with Changing Distance-Visual Band.....	147
Figure 4-55 Matched Features of LWIR Band Vehicle Images-SIFT Detector. ...	148
Figure 4-56 Matched Features of LWIR Band Vehicle Images-Harris Laplace Detector.	148
Figure 4-57 Matched Features of LWIR Band Vehicle Images-Hessian Laplace Detector.	149
Figure 4-58 Matched Features of LWIR Band Vehicle Images-SURF Detector. .	149
Figure 4-59 Matching Accuracy of Vehicle Images with Changing Rotation Angle-LWIR Band	150
Figure 4-60 Matching Accuracy of Vehicle Images with Changing Rotation Angle-Visual Band	151
Figure 4-61 White Cartoon Mask of Experiment 3	153
Figure 4-62 Blackbody Device without Mask (Experiment 3).....	154
Figure 4-63 Blackbody Device with Mask (Experiment 3).....	154
Figure 4-64 Typical Infrared Images of Blackbody Device with Mask (Experiment 3).....	155
Figure 4-65 LWIR Band Infrared Images with 150°C Blackbody at 8m.	156
Figure 4-66 LWIR Band Infrared Images with 150°C Blackbody at 43m.	156
Figure 4-67 MWIR Band Infrared Images with 150°C Blackbody at 8m.	157
Figure 4-68 MWIR Band Infrared Images with 150°C Blackbody at 43m.	157
Figure 4-69 Visual Band Infrared Images with 150°C Blackbody at 8m.	158
Figure 4-70 Visual Band Infrared Images with 150°C Blackbody at 43m.	158
Figure 4-71 MWIR Band Image Feature Points Extracted by SIFT Detector, 150°C Blackbody, 8m.....	159
Figure 4-72 MWIR Band Image Feature Points Extracted by Harris-Laplace Detector, 150°C Blackbody, 8m.	160
Figure 4-73 MWIR Band Image Feature Points Extracted by Hessian-Laplace Detector, 150°C Blackbody, 8m.	160
Figure 4-74 MWIR Band Image Feature Points Extracted by SURF Detector, 150°C Blackbody, 8m.....	161

Figure 4-75 Visual Band Image Feature Points Extracted by SIFT Detector, 150°C Blackbody, 8m.....	162
Figure 4-76 Visual Band Image Feature Points Extracted by Harris-Laplace Detector, 150°C Blackbody, 8m.	163
Figure 4-77 Visual Band Image Feature Points Extracted by Hessian-Laplace Detector, 150°C Blackbody, 8m.	164
Figure 4-78 Visual Band Image Feature Points Extracted by SURF Detector, 150°C Blackbody, 8m.....	165
Figure 4-79 Matched Feature Points of MWIR Band Images, SIFT detector, 150°C Blackbody, 8m.....	166
Figure 4-80 Matched Feature Points of MWIR Band Images, Harris-Laplace Detector, 150°C Blackbody, 8m.	167
Figure 4-81 Matched Feature Points of MWIR Band Images, Hessian-Laplace Detector, 150°C Blackbody, 8m.	168
Figure 4-82 Matched Feature Points of MWIR Band Images, SURF Detector, 150°C Blackbody, 8m.	169
Figure 4-83 Matching Accuracy with Changing Rotation : MWIR Band, SIFT...	170
Figure 4-84 Matching Accuracy with Changing Rotation : LWIR Band, SIFT	170
Figure 4-85 Matching Accuracy with Changing Rotation : Visual Band, SIFT....	171
Figure 4-86 Matching Accuracy with Changing Rotation : MWIR Band, Harris-Laplace	173
Figure 4-87 Matching Accuracy with Changing Rotation : LWIR Band, Harris-Laplace	174
Figure 4-88 Matching Accuracy with Changing Rotation : Visual Band, Harris-Laplace	175
Figure 4-89 Matching Accuracy with Changing Rotation : MWIR Band, Hessian-Laplace	177
Figure 4-90 Matching Accuracy with Changing Rotation : LWIR Band, Hessian-Laplace	178
Figure 4-91 Matching Accuracy with Changing Rotation : Visual Band, Hessian-Laplace	179
Figure 4-92 Matching Accuracy with Changing Rotation : MWIR Band, SURF.	181

Figure 4-93 Matching Accuracy with Changing Rotation : LWIR Band, SURF ..	181
Figure 4-94 Matching Accuracy with Changing Rotation : Visual Band, SURF..	182
Figure 4-95 LWIR Band Images with 150°C Blackbody at 8m and 22.5m.	184
Figure 4-96 MWIR Band Images with 150°C Blackbody at 8m and 22.5m.	185
Figure 4-97 Visual Band Images with 150°C Blackbody at 8m and 22.5m.	185
Figure 4-98 MWIR Band Image Feature Points Extracted by SIFT Detector, 150°C Blackbody.....	186
Figure 4-99 MWIR Band Image Feature Points Extracted by Harris-Laplace Detector, 150°C Blackbody.....	187
Figure 4-100 MWIR Band Image Feature Points Extracted by Hessian-Laplace Detector, 150°C Blackbody.....	188
Figure 4-101 MWIR Band Image Feature Points Extracted by SURF Detector, 150°C Blackbody.	189
Figure 4-102 Visual Band Image Feature Points Extracted by SIFT Detector.....	190
Figure 4-103 Visual Band Image Feature Points Extracted by Harris-Laplace Detector.	190
Figure 4-104 Visual Band Image Feature Points Extracted by Hessian-Laplace Detector.	191
Figure 4-105 Visual Band Image Feature Points Extracted by SURF Detector....	191
Figure 4-106 Matching Accuracy with Changing Distance : SIFT	192
Figure 4-107 Matching Accuracy with Changing Distance : Harris-Laplace.....	193
Figure 4-108 Matching Accuracy with Changing Distance : Hessian-Laplace.....	193
Figure 4-109 Matching Accuracy with Changing Distance : SURF.....	194

LIST OF ABBREVIATIONS

AGC	: Automatic Gain Control
bmp	: Bitmap
CCD	: Charge Coupled Device
DCR	: Discriminative Cluster Refinement
DoG	: Difference of Gaussian
FPA	: Focal Plane Array
GHz	: Giga Hertz
GUI	: Graphical User Interface
InGaAs	: Indium Galium Arsenide
InSb	: Indium Antimonide
IR	: Infrared
LoG	: Laplacian of Gaussian
LRF	: Laser Range Finder
LWIR	: Long-Wave Infrared Band
MCT	: Mercury Cadmium Telluride
MWIR	: Mid-Wave Infrared Band
NIR	: Near Infrared
Ref	: Reference
RF	: Radio Frequency
SAR	: Synthetic Aperture Radar
SIFT	: Scale Invariant Feature Transform
Si	: Silicon
SPIE	: International Society for Optical Engineering
SURF	: Speeded Up Robust Features
Tr	: Transformed
UAV	: Unmanned Air Vehicle
VOC	: Visual Object Classes

CHAPTER 1

INTRODUCTION

1.1 Background

Image features extracted from different images have been popular recently in image matching and object recognition applications. The features that are invariant to scale, rotation and viewpoint changes in images are widely used in these applications. It is usually difficult to obtain scale and rotation invariant features under different viewpoint conditions with high matching accuracy.

The scale invariant feature extraction and matching operation consists of four distinct steps listed below.

- a) Candidate feature locations are determined by feature or keypoint detectors utilized. Scale space representations of the images are obtained. Keypoint detectors are operated over the spatial location of the images and also over different scales of the image in order to find candidate feature locations consisting of spatial coordinates and corresponding scales. The final feature or keypoints are obtained by applying special conditions of the detector used, such as stability under scale, local threshold of the detector, etc.
- b) Orientations of the feature points are calculated by obtaining local image gradient directions.

- c) Descriptor vectors are obtained for each feature points in the image. The descriptor vector characterizes the corresponding feature point with respect to its local region in the image.
- d) The matching feature points between two images, one of which is reference image and the other one being training (transformed) image, are obtained. The matching feature points with high accuracy are selected among the ones found by the matching algorithm.

1.2 Motivation and Scope of Thesis

The efficiency of the scale and rotation invariant features and matching accuracy of these feature points obtained are evaluated in this study for different detectors. The detectors that are utilized in this study are listed below.

1. Scale Invariant Feature Transform (SIFT) Detector (Difference of Gaussian) [1]
2. Multiscale Harris-Laplace Detector [2]
3. Multiscale Hessian-Laplace Detector [2]
4. Speeded Up Robust Features (SURF) Detector [3]

SIFT detection Scale Invariant Feature Transform (SIFT) technique is developed by Lowe [1], which is widely used to obtain scale invariant features.

Multiscale Harris detector is an adapted version of Harris detector, whereas multiscale Hessian detector is an adapted version of Hessian detector. These detectors are used with Laplace detector one after the other. In general, these detectors, developed by Mikolajczyk and Schmid [2], are named as multiscale Harris-Laplace and multiscale Hessian-Laplace detectors.

SURF feature detector is another scale and rotation invariant detector, developed by Bay, Tuytelaars and Gool [3]. This method provides very fast and efficient results when compared with other detectors.

The performance and matching accuracy of each of these detectors are evaluated in this study by using images taken in the following listed bands.

1. Mid-Wave Infrared Band (MWIR)
2. Long-Wave Infrared Band (LWIR)
3. Visual Band

Fundamentals of Radiometry and Optical imaging system theory are explained in Chapter 2 for each of the infrared-bands and visual-band in detail. The details of Radiometry Theory are summarized for MWIR and LWIR bands in this chapter.

Mathematics of the feature detectors are explained in Chapter 3 in detail. The orientation and descriptor algorithms used in simulations are explained and matching algorithm is presented in this section.

Experimental setups, simulations performed in this study, and the results of the experiments are explained in Chapter 4 along with the way infrared images are obtained and represented in simulations. Firstly, the evaluation of matching of MWIR and LWIR images among objects at different temperatures is discussed in this chapter. Then, matching performance for images of real objects, face and a vehicle, in LWIR and visual-bands are evaluated by using each feature detector. Thirdly, matching performance of each detector is evaluated for images in each band with different rotations. Finally matching performance of each detector is evaluated for images in each band with different distances.

Conclusions and their discussion are presented in Chapter 5.

CHAPTER 2

RADIOMETRY AND OPTICAL IMAGING SYSTEMS

2.1 Fundamentals of Radiometry

Radiometry is the name of the field dealing with detection and measurement of electromagnetic radiation in ultraviolet, visible and infrared-bands [4]. Radiometric parameters, units and terminology are described in this section.

Radiation radiates from objects in the form of electromagnetic waves flying through space. Light is the visible portion of the electromagnetic spectrum that we see in real life. This light is termed as visual-band of this spectrum in this study.

Electromagnetic radiation is the general terminology of the radiation radiating from objects and it is an electromagnetic wave that propagates in vacuum in constant speed expressed by the letter c , (speed of light $c=3 \times 10^8$ m/sec). Electromagnetic radiation from objects is characterized by the wavelength of light coming from these objects. It is defined by the following relation (2-1) :

$$\nu = \frac{c}{\lambda} \quad (2-1)$$

where c is the speed of light ($c=3 \times 10^8$ m/sec), λ is the wavelength of radiation (light) in meters and ν is the frequency of the electromagnetic radiation in Hertz.

Electromagnetic radiation exhibits wave and quantum properties as it is explained in detail [5]. Electromagnetic wave radiation is expressed by the relation (2-2) and electromagnetic wave (light) is carried by particles, each of which is called photons explained by Planck and Einstein [5].

$$E = h \frac{c}{\lambda} \quad (2-2)$$

where E is the photon energy in joules, h is Planck's constant ($h=6.623 \times 10^{-34}$ J s), c is the speed of light ($c=3 \times 10^8$ m/sec) and λ is the wavelength of radiation (light) in meters.

Radiation energy of photons increase for shorter wavelengths or for higher frequencies. Electromagnetic wave radiation is represented in terms of a spectrum, called Electromagnetic spectrum. The term spectrum or spectral means that the change of the dependent variable, here the energy, is given in terms of wavelength. The electromagnetic spectrum is shown in Figure 2-1.

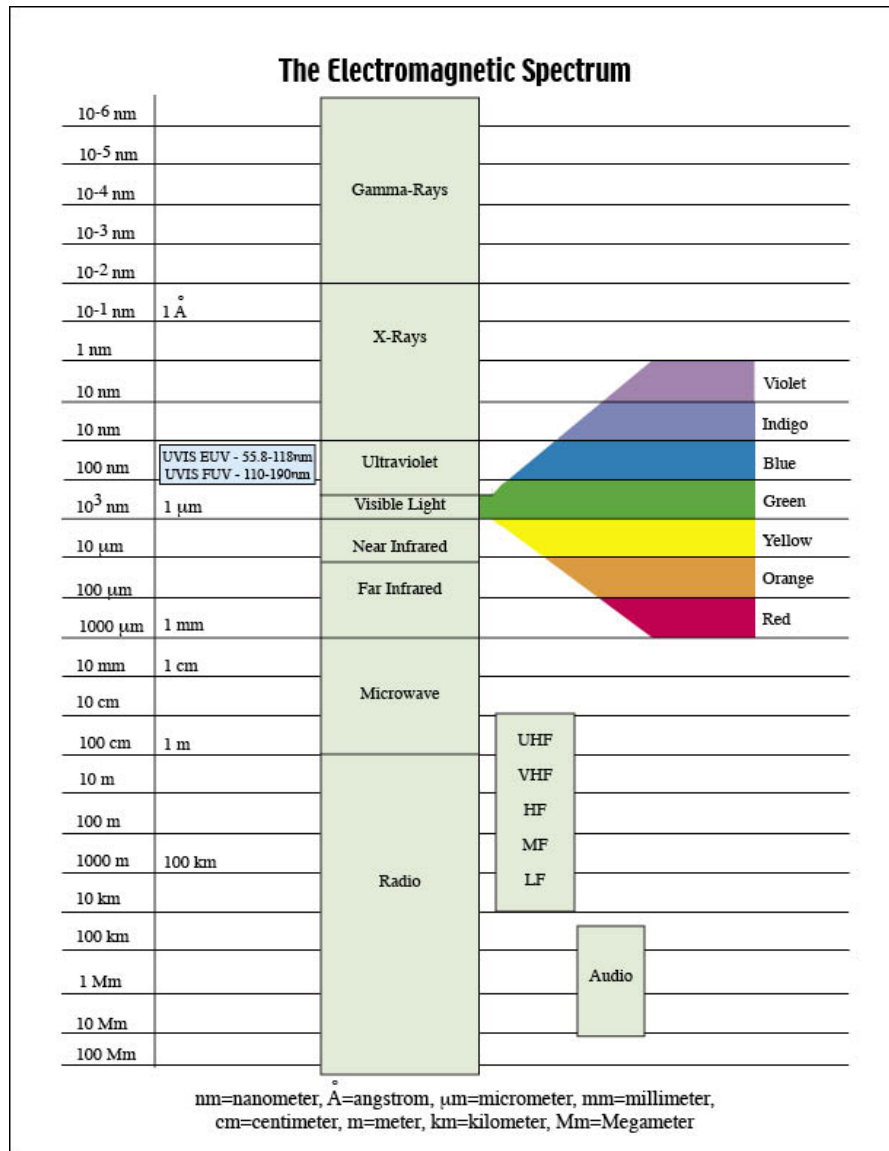


Figure 2-1 Electromagnetic Spectrum

Figure 2-1 shows the portion of the electromagnetic spectrum of interest. Frequency of light is typically used in Radio portions (RF) of the spectrum. However, wavelength is more common to use at higher frequencies such as frequencies higher than 300 GHz [6]. The unit of wavelength is in micron, 10^{-6} meters, abbreviated as μm [6].

Radiometric terms are divided into two categories based on the radiation whether it is coming from a source or it is radiating and coming into a surface [4], [5]. Three radiometric terms are used to describe the Radiant Flux from a source and they are listed below.

1. Radiance (L)
2. Radiant Emittance (M)
3. Radiant Intensity (I)

The following term is used to describe the Radiant Flux into/on a surface listed below.

1. Irradiance (E)

These radiometric terms are illustrated with a typical visual representation as it is shown below.

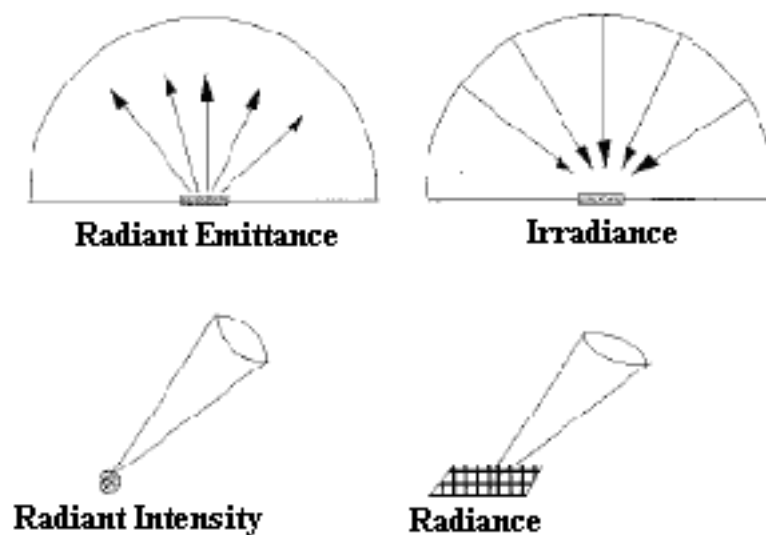


Figure 2-2 Radiometric Terms-Visual Illustration

It is important to understand a key concept while considering the geometry of radiometric terms, which is solid angle. The unit of solid angle is in steradians. A sphere has 4π steradians. A steradian is defined as “the solid angle which, having its vertex at the center of the sphere, cuts off a spherical surface area equal to the square of the radius of the sphere” in [5]. The solid angle of a surface can be calculated by using the mathematical relation (2-3).

$$\Omega = \frac{A}{r^2} \quad (2-3)$$

where Ω is the solid angle in steradians, A is the area of the surface area in meter-square and r is the distance in meters from the radiating point source.

All of the radiometric terms are summarized in Table 2-1 presented below.

Table 2-1 Summary for the Radiometric Terms

No	Symbol	Term	Expression	Unit
1.	Q (U)	Radiant Energy	Q	Joule, J
2.	Φ (P)	Radiant flux	$\frac{\partial\Phi}{\partial t}$	Watt
3.	M (W)	Radiant Emittance	$\frac{\partial\Phi}{\partial A}$	Watt/cm ²
4.	M_λ (W_λ)	Spectral Radiant Emittance	$\frac{\partial\Phi}{\partial A \cdot \partial\lambda}$	Watt/cm ² /μm
5.	L (N)	Radiance	$\frac{\partial\Phi}{\partial A \cos\theta \cdot \partial\Omega}$	Watt/cm ² /str
6.	L_{λ} (N_λ)	Spectral Radiance	$\frac{\partial\Phi}{\partial\lambda \cdot \partial A \cos\theta \cdot \partial\Omega}$	Watt/cm ² /str/μm
7.	E (H)	Irradiance	$\frac{\partial\Phi}{\partial A}$	Watt/cm ²

Table 2-1 Summary for the Radiometric Terms (Continued)

No	Symbol	Term	Expression	Unit
8.	E_{λ} (H_{λ})	Spectral Irradiance	$\frac{\partial\Phi}{\partial A \cdot \partial\lambda}$	Watt/cm ² /μm
9.	I	Radiant Intensity	$\frac{\partial\Phi}{\partial\Omega}$	Watt/str
10.	I_{λ}	Spectral Radiant Intensity	$\frac{\partial\Phi}{\partial\Omega \cdot \partial\lambda}$	Watt/str/μm

Each radiometric term has its spectral and total (cumulative in spectral domain) counterpart. Spectral terms give the corresponding quantity in terms of wavelength of the electromagnetic radiation whereas the other counterpart gives the total value of energy or energy density. Explanations for these terms can be found in [4] and [5] in detail.

Infrared portion of the electromagnetic radiation covers a large part of the electromagnetic spectrum starting from ultra-violet and visible part until microwave portion. Infrared spectrum starts from 770-800 nm of wavelength and covers the part of the radiation until 1000 μm. The optical part of the electromagnetic radiation is presented in Figure 2-3 below [5].

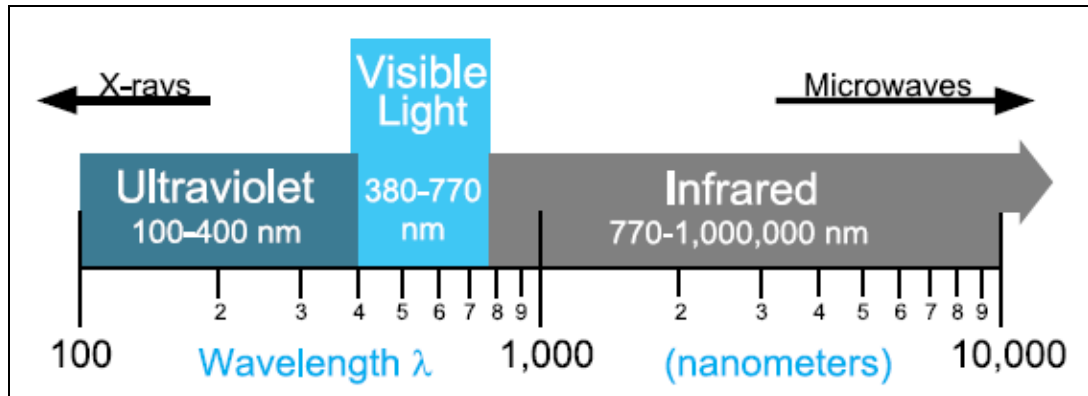


Figure 2-3 Optical Portion of Electromagnetic Radiation Including Infrared Part

Infrared radiation can be divided into three sub regions that are listed below.

1. Near Infrared (NIR, 770-1400 nm)
2. Far Infrared (1.4-1000 μm)

Far infrared region can be divided into two regions as well, as it is listed below [5].

1. Mid Wave Infrared (MWIR, 1.5-5.0 μm)
2. Long Wave Infrared (LWIR, 7.0-13.0 μm)

Images obtained in mid-wave and long-wave infrared-bands are examined in this study. Special infrared cameras are used during controlled experiments.

Infrared portion of the electromagnetic spectrum contains less amount of radiation when it is compared with other parts of the spectrum.

Blackbody concept is important while considering infrared radiation. A blackbody is defined as an ideal body, which absorbs all incident energy at all wavelengths and at all angles. Blackbody reflects none of the incident energy at all wavelengths. Electronic devices close to blackbody behavior can be found in real life, which provides and emits uniform radiation from their surfaces. These devices called

blackbodies are often used as reference sources for infrared cameras and some blackbody reference sources are used in controlled experiments in this study.

Blackbody radiation can be calculated by using Planck's Radiation Law, as it is explained in [4] and [5]. This law gives the relationship between the Spectral Radiance of a blackbody and the temperature of that blackbody. The mathematical expression is presented with (2-4) below.

$$L(\lambda, T) = \frac{c_1}{\lambda^5 (\exp(c_2 / \lambda T) - 1)} \quad (2-4)$$

where $L(\lambda, T)$ is the spectral radiance of the blackbody, λ is the wavelength in μm , T is the temperature of the blackbody in K, c_1 and c_2 are constants as listed below.

$$c_1 = 1.19 \times 10^4 \text{ Watts}/(\mu^4 \text{ steradian cm}^2)$$

$$c_2 = 1.438 \times 10^4 \mu^\circ\text{K}$$

Spectral emittance of a blackbody can be calculated by the relation (2-5)

shown below.

$$M(\lambda, T) = L(\lambda, T) \cdot \pi \quad (2-5)$$

where M is the spectral emittance and L is the spectral radiance of the blackbody.

Spectral emittance curve of a blackbody at 500°C temperature is presented in Figure 2-4 below.

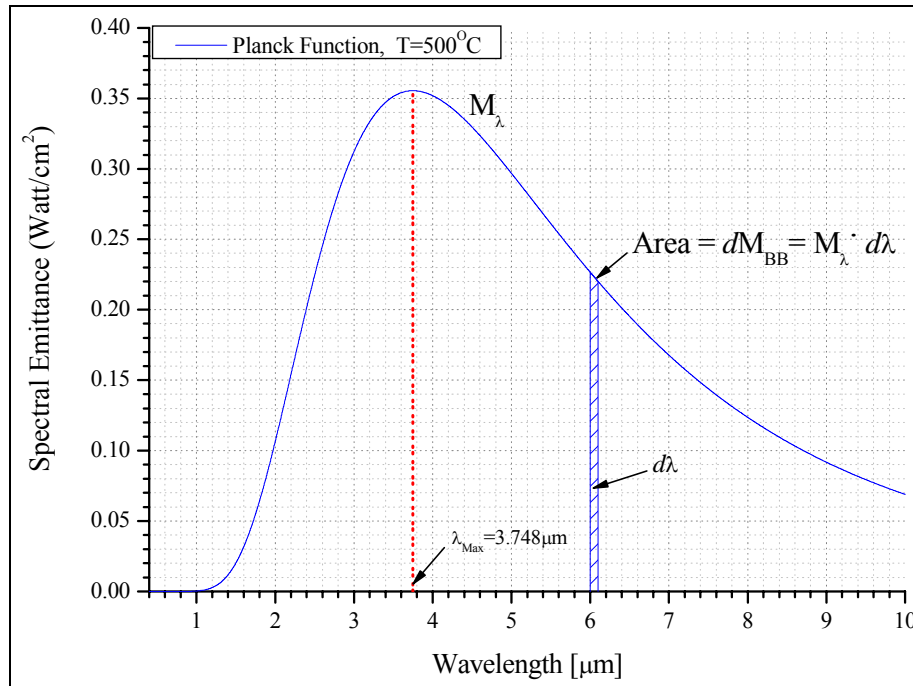


Figure 2-4 Spectral Emittance of A Blackbody at 500°C (Typical)

Area under spectral emittance curve gives total emittance of the blackbody between specified wavelengths.

2.2 Optical Imaging Systems

Objects are observed and detected by optical and electronical systems that have been developed and designed in this century. Detection and observation process of an object is based on several physical phenomena, such as “target-to-background contrast” and “sensor performance” [4].

Engineers and experts design systems and devices, so called detection systems, that provide the “highest possible detection probability” to detect special features, such as “color, temperature, sound, smell, shape” of a target or an object [4]. Detection systems are divided into three categories as it is explained in [4] and they are listed below.

1. Active Sensor Systems
2. Passive Sensor Systems
3. Semiactive Sensor Systems

Active sensor type of systems consists of an active source and a detector integrated in one system. Active source creates the signal and the detector detects the reflected form of that signal from the target. The idea is that active source “illuminates” the target and the reflected radiation from the target is detected [4]. Laser Range Finder (LRF) and Synthetic Aperture Radar (SAR) systems are typical examples for these types of systems. Active sensor systems are explained in [4] in detail.

Passive sensor systems do not include an active source, but consist of only a sensor or detector instead. The idea of detection is such that these types of systems detect the signal or radiation emitted from an object [4]. Thermometers and Thermal Imagers are typical examples for these types of systems. Active Sensor Systems are explained in [4] in detail.

Semiactive sensor systems are systems that detect active sources such as Laser Range Finder, Radar and Electronic Support measures [4]. Counter-Counter Measure systems of Electronic Warfare are examples for these types of systems.

Detection systems are designed in either form of imaging or nonimaging types [4]. Imaging sensor systems or optical imaging systems are the main concern of this chapter.

Imaging concept requires the capture of light or radiation from a scene, the detection of the feature to be displayed and the form of display of a “two-dimensional picture” of that scene [6]. This imaging or capture of light or radiation can be in either the visible band or a non-visible band, such as in one of the infrared-bands.

Image formation process depends on two factors as it is explained in [7]. One of them is the geometry of the image formation. The projection of a point of the scene is located in the image plane. The spatial location of this point in the image is determined by the geometry of the image formation, in other words the viewpoint of the scene, the rotation of the object, the object is either moving or static, etc. The second factor is related with the physics of light. The brightness of a point of an object scene in the image plane changes as a function of illumination and surface properties. This property is especially important in visual-band.

A simple model for the image formation process is presented in Figure 2-5 below [7].

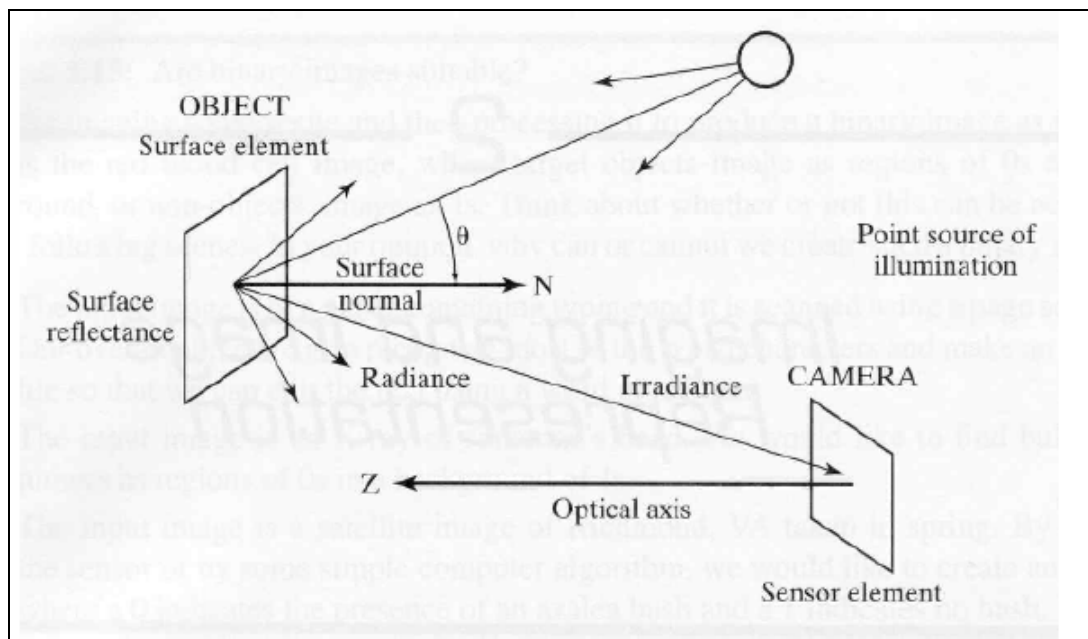


Figure 2-5 Simple Model of Image Formation

The scene of the object is illuminated by an outside source such as the sun or a light bulb simply. The object reflects the coming radiation towards the detection system,

usually a type of passive sensor system, working in either visual or one of the infrared-bands. The detection system “senses” this radiation via the detectors on it and displays the corresponding image in a visual format to the users [7].

Passive sensor imaging systems are used in this study, which operate in visual and infrared-bands. The optical components that capture the light in the corresponding band, the detection and visualization components differ for each band. The detection, visualization and capture mechanisms are explained in Section 2.2.1 and Section 2.2.2 for infrared-band and visual-band passive sensor imaging systems respectively.

2.2.1 Imaging in Infrared Bands

Imaging in infrared-bands is closely related with the thermal radiation emitted from the objects and “hot” sources. Imaging in infrared-bands is also called Thermal Imaging for this purpose. Systems that are used for imaging in infrared-bands are generally called Thermal Imaging Systems and they are passive sensor systems operating in infrared wavelengths between $3\mu\text{m}$ and $14\mu\text{m}$, emitted by objects at around 300K or 27°C [8].

Thermal imaging systems provide the users, the “temperature gradient” of the object that is viewed. The temperature gradient of the object is viewed by television, such as images at night, as well as during the day [8]. The visual output of infrared imagery changes with respect to the imaging system’s output. The visual side of the infrared imaging systems is quite much like the monitoring of visual-band imaging systems, which is explained in Section 2.2.2 in detail. The displayed infrared-band image or picture is divided into pixels like for all electronically implemented imagery. A pixel is defined as “a spot” on the screen. The brightness (gray value) or color is stored in digital format for each pixel and this information is displayed at each pixel location on the screen [6].

Infrared-bands contain less energy when compared with other bands, which necessitates the use of special type of detectors to capture these low energy levels [5]. Detectors translate light energy into electrical current, in other words, photons coming through the semiconductor material of the detector excites the electrons in the semiconductor material and creates an electrical current through the output of the detector. The value of the current induced is linear with the energy of the photon or incident light on the detector [5].

Detectors of infrared or thermal imaging systems are very important in such a way that it affects system performance directly. Thus, detectors should be chosen carefully before system design by considering some parameters. Firstly, sensitivity of the detector under concern to the required specific infrared-band is an important parameter. The detector to be used should have adequate strong signal in the required infrared-band. Secondly, linearity of the detector is another parameter. Detector should have a linear response in the active band with the incident light upon it. Thirdly, some detector types used in infrared-bands should be cooled to have a stable operation, which is the stability parameter, since they suffer from temperature sensitivity. Finally, durability of the detector material is important, due to the fact that detector materials production is a difficult process and in order to maintain the detector material working properly in the system, one should not put the system optics in excessive energetic light or one should not allow the detector material to be subject to external effects such as touching, cleaning its surface.

Number of detector types suitable for infrared-band radiation detection has increased in number recently as a result of technological advances. However, it is still difficult to produce detector materials. Thermopile detectors, InGaAs and Si quantum type detectors provide excellent, very flat responsivity (sensitivity) in Near infrared-band regions [5]. The other type of detectors, such as Indium Antimonide (InSb) detectors suitable for mid-wave infrared-band detection and Mercury Cadmium Telluride (MCT) detectors suitable for long-wave infrared-band detection should be cooled to very low temperatures (e.g.77K) [8]. In order to maintain such

low temperatures, detectors are put inside vacuum housing called Dewar and cooling is achieved by the use of Liquid Nitrogen or Stirling cooler technology.

Filters are used in infrared or thermal imaging systems in order to measure only a small wavelength band, or sometimes to attenuate the coming very strong signal. Infrared-band radiation coming from the source passes through this stage just before the detector component. Spectral or narrow band-pass filters are used for measuring only within a small wavelength band. This approach is sometimes required for some applications. If transmission band of the filter is very narrow, the output of the system gives us only the energy at that specific wavelength [5]. Neutral density type filters are examples to the filters that attenuate the coming signal into the system. These types of filters are sometimes required, if the incoming signal strength from the source is very strong and if it is out of detection range in terms of signal strength. There are also other types of special filter examples, such as long-pass filters, interference filters and polarization filters.

Lenses are used in infrared or thermal imaging systems' input optics part in order to focus and redirect the incoming light in infrared-band, such that the maximum amount of light or radiation coming from the target can be collected and concentrated in the detection process [5]. Infrared-band radiation coming from the source passes firstly this stage just before the filtering. The input optics and lenses have an important role in the geometry of the imaging scenario. They should be chosen appropriately by considering the size of the target and the viewing angle of the lens to be used. If the target to be measured is very far away from the imaging system, then lenses with narrow viewing angle and large focal length should be selected. On the contrary, if the target to be measured is close enough to the imaging system and field of view is required to be completely filled, then lenses with wide viewing angle and small focal length should be used.

2.2.2 Imaging In Visual Bands

Visual-band imaging systems rely on reflected radiation from the objects illuminated by visual light sources, such as the sun and light bulb. Visual-band imaging devices usually cover the short wavelengths between $0.45\mu\text{m}$ and $1\mu\text{m}$. Visual-band imaging is closely related with the human eye vision, where human eye can only see between $0.46\mu\text{m}$ and $0.65\mu\text{m}$ [9]. Visual-band imaging systems operation is based on the simple model of image formation explained in Section 2.2 and shown in Figure 2-5.

The simplest setup to form an image in Visual-band of a three dimensional scene on a two dimensional surface is so called “pinhole” camera. Reflected light from the object pass through the pinhole and form an inverted image of the object on the image plane. The schematic of this setup is shown in Figure 2-6 below.

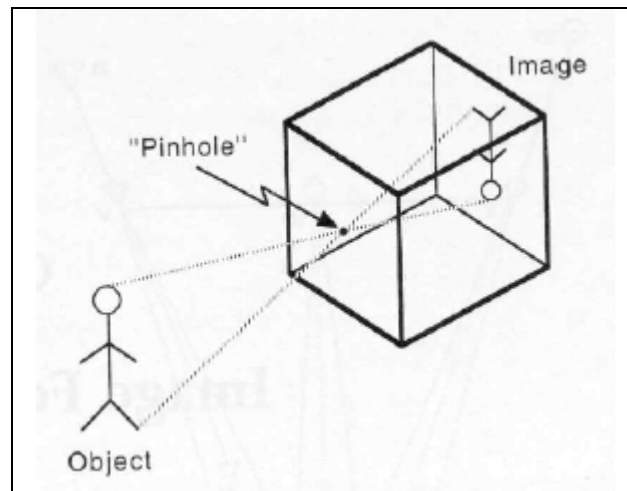


Figure 2-6 “Pinhole” Setup for Image Formation in Visual Band

Visual-band camera input optics are designed in such a way that the diameter of the aperture is large enough to admit incoming visual light from the scene and lenses are placed appropriately to focus the image of the scene without any loss of light.

Visual-band imaging systems have evolved from the mechanical scanning mechanisms historically such as “Nipkow Disk”. As the disk rotates, the scene is scanned horizontally one line at a time and the whole area of the image is obtained [10]. “Nipkow Disk” is shown in Figure 2-7 below.

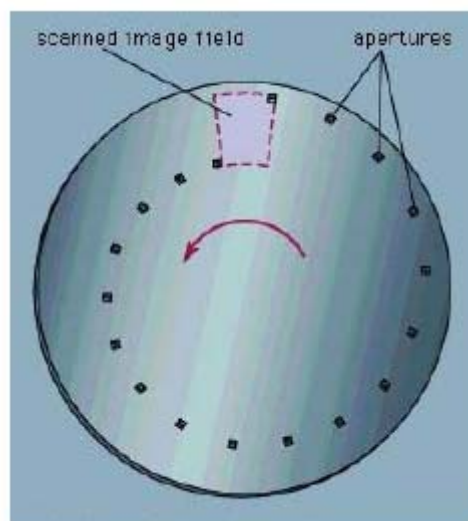


Figure 2-7 Nipkow Disk

After the foundation of mechanical scanning mechanism for optical capture of the scene, an important discovery of cathode ray tubes has been done and its application with line by line scanning. Photoelectric elements, converted after the scanning mechanism, are discharged sequentially by the cathode ray beam. These were the beginning forms of modern television, which were developed further by the change of mechanical scanning with electronic methods [10]. Cathode ray tubes are early examples of Image Intensifiers, which are developed further with newer

generations. The intensified electrons coming out of the Image Intensifier fall over the screen and visualized as the image of the scene. Image Intensifier is the counterpart of the detector component of the infrared imaging systems in visual-band.

The Charge Coupled Device (CCD) is one of the important systems used in visual imaging area. They are widely used in most of the imaging systems and they can be easily integrated into most of the systems as their display component. CCD is an array of tiny solid state cells that convert visual light energy into electrical charges. They are manufactured on small size chips typically 1cm by 1cm in size. This is a typical size for an 512 by 512 array output image. The output of a CCD array is a continuous video signal generated by scanned input scene line by line with photo-sensors [7].

CHAPTER 3

INTEREST POINT DETECTION AND DESCRIPTION

Finding corresponding points between two images of the same scene from different viewpoints or having some type of transformations is a challenging task in Computer Vision and Pattern Recognition area. Correspondence problem and the mathematical insights for this subject are explained in [11].

Matching process of local regions or feature points between two images starts with interest point or feature point extraction. “Edges” and “corners” are the first candidate points that are taken as interest points traditionally [11]. However, edges are not distinctive in the sense that a horizontal or vertical edge of an object may not be distinguished from its local neighborhood. Corners are better candidates compared to edges in that sense. Despite this fact, most of the images in real life have round and curved structures along with blobs inside their local regions. Then, the question of “how can one identify those local regions in images and also how can a person find good correspondences of local regions or feature points between two such images?” comes to the mind.

Several local measures are defined in literature in order to find interest points and local regions especially for images having round, curved structures [1], [2], [3]. Moreover, local cornerness measure is defined by Multiscale Harris detector in [2]. Interest points are selected on the basis of finding the local maximum of these measures with respect to some threshold values applied. These local measures are explained in this chapter in detail.

The local measures used in order to find candidate interest points provide spatial coordinates of these interest points. When there is a scale change of the same scene, the interest point belonging to the reference image can be a “dull” feature in the transformed image, located at the same spatial corresponding coordinate. Thus, the features obtained need be scale invariant at the same time [12]. This is not the only problem of scale change between two images. One of the corresponding features in one image may be a “sharp” (or strong) one whereas the other corresponding feature in the second image may be a “diffuse” (or weak) one, which results in errors of matching and correspondence because of the lack of multiscale representation of the images [12]. Thus, the local measures used in finding the interest points need to be in multiscale format, yielding scale invariance. Also, the images have to be evaluated with their corresponding scale space representations. The scale space representation of discrete real world images or signals is developed along with the detailed mathematical theory insight in [13]. Scale dimension of feature points along with their spatial coordinates provide high matching accuracy and correspondence, if there is less affine deformation and transformations between two images of the same scene. (i.e. Rotation, illumination change, etc.)

Scale space representation is the representation of an image as a one-parameter family, set of smoothed images. The parameters used in construction of scale space representation are size of smoothing kernel (2x2, 9x9, etc.), base standard deviation of smoothing (σ), and level of the smoothing at which, standard deviation is a constant multiple k of base standard deviation, $\sigma' = k\sigma$. The term “scale” is defined as this level of smoothing [46].

Scale selection stage of interest points are performed by finding the scale level at which derivative of some local measures takes local maximum in scale space, based on Lindeberg’s ideas [12]. Difference of Gaussian (DoG) is used in scale selection of SIFT features in Lowe’s study [1], whereas Laplacian of Gaussian (LoG) is used in scale selection of multiscale Harris, multiscale Hessian and SURF features in Mikolajczyk-Schmid and Bay-Tuytelaars-Gool’s studies respectively [2], [3].

SIFT, multiscale Harris-Laplace, multiscale Hessian-Laplace and SURF (Speeded Up Robust Feature, Fast Hessian Detector) detectors are used in interest point extraction in this study. The utilization of these detectors of previous studies in visual and infrared-bands, are explained in the next section. Next, the mathematical formulation of each detector and main ideas behind them are discussed. Orientation computation, descriptor calculation and matching accuracy calculation details are discussed at the end of this chapter.

3.1 Interest Point Detectors in Visual and Infrared Bands : Previous Studies

Scale and affine invariant interest point detectors and descriptors have been popular research topics in recent years in Computer Vision and Pattern Recognition areas. Insights behind developing such scale and affine invariant feature extractors and descriptors are discussed in the introduction of this chapter. Usage of these detectors for visual and infrared-band images is discussed and examples from previous studies are also presented in this section.

3.1.1 Interest Point Detectors in Visual Bands

Harris corner detector was developed by Harris and Stephens (1988), that has been widely used in tracking and recognition studies [35]. The original Harris corner detector is not scale invariant and the scale invariant (multiscale) and affine invariant versions of this detector was developed by Mikolajczyk and Schmid (2004) later in [2]. The scale invariance was achieved by the “automatic scale selection” principle based on Lindeberg’s (1998) study in [12]. Multiscale Harris-Laplace detector is the combined name of scale invariant version of this detector.

Zhang et.al (2007) used multiscale Harris-Laplace detector and only Laplace detector with different descriptors in their study, to conduct an evaluation for their recognition methods on several textures and object visual-band image databases

[19]. They have observed that overall performance of Harris-Laplace and Laplace only feature detectors with SIFT descriptors in terms of “classification accuracy” is better and preferable among the other alternatives [19]. This observation shows that multiscale Harris-Laplace detector provides good matching and classification performance, when corners are dominant in the visual-band images, such as the texture images used in this study.

Yang et.al compared the usage of Harris-Laplace features and Laplacian features with SIFT descriptors on the classification performance of their “Discriminative Cluster Refinement (DCR)” algorithm [20]. They performed their experiments with “PASCAL VOC Challenge 2006” dataset of visual-band images, containing 5304 images with 9507 objects, which include bicycle, bus, motorbike, cat, cow, dog, horse, sheep and person object classes. Both Harris-Laplace and Laplacian feature detectors with SIFT description provided good classification performance with their DCR algorithm [20]. This shows that multiscale Harris-Laplace feature detector provides good matching and classification performance for different types of object images in visual-band as in the study of [19].

Marszalek and Schmid performed a similar study to the ones, Zhang et.al and Yang et.al, to evaluate the classification performance of their “Semantic Hierarchic Classifier” by using Harris-Laplace and Laplacian feature detectors with SIFT description [21]. They used “PASCAL VOC'06” dataset of visual-band images as in [20] of which 1277 training and 1341 validation images are taken in their experiments. Harris-Laplace features with SIFT descriptions provide successful matching and classification performance among classes of different object image classes, which is an evidence for the successful usage of Harris-Laplace feature detector in visual-band.

Hessian detector was developed by Beaudet [23], which was not as widely accepted as traditional Harris corner detector. Hessian detector is naturally scale invariant itself. Knowing this property, Lindeberg (1998) evaluated the performance of the determinant and trace of the Hessian matrix in his famous study [12]. Trace of the

Hessian matrix is actually called as the Laplacian detector in literature. Lindeberg showed the scale invariance property of Hessian detector. Mikolajczyk and Schmid refined Lindeberg's study by defining Hessian-Laplace detector in their study [24]. Initial feature points are detected as local maxima of Hessian matrix determinant in space and characteristic scale of each interest point is obtained at the local extrema of Laplacian of Gaussian (LoG) [24].

Mikolajczyk and Schmid compared the performance of Hessian-Laplace detector among other interest point detectors used with several descriptor combinations [24]. They compared SIFT descriptor in their study as well. They used real images with lots of different geometric and photometric transformations in their experiments including rotation, scale changes and other types of affine changes like viewpoint change. The results showed that Hessian type detectors (Hessian-Laplace and Hessian-affine) provided better performance in terms of accuracy for visual images with blob-like structures [24].

Scale Invariant Feature Transform (SIFT) detector was initially developed by Lowe in [14]. Lowe proposed improvements in "stability" and "invariance" for SIFT detector and descriptor finding in his later study [1]. SIFT feature extraction and description has become very popular in Computer Vision and most of the related areas and found extensive usage with visual-band images.

SIFT feature extraction and description method was used successfully in matching different images of the same scene in visual-band and stitching these images to form panoramic views of the image in [15]. SIFT features of all of the images are extracted and then matched. Images with different zoom and orientations are matched successfully, since SIFT features are invariant to scale and rotation changes. Despite changes in illumination in some images of the same scene, SIFT interest points are matched robustly and very good panoramic view of the scene is obtained [15].

SIFT feature extraction method is also used in another similar study by Snavely et.al in [16], in order to “geometrically register large photo collections from the Internet and other sources” and to use this information in other applications. SIFT features of all the images in their database are extracted. Local SIFT descriptors are obtained for each keypoint and matches are obtained for image pairs. They have applied this process to set of photo collections and the scene is identified correctly, according to which category or set of photos it belongs to [16].

SIFT feature extraction is further utilized in feature-based face recognition study of [17] in visual-band. Two dimensional face recognition is performed by extracting SIFT features and descriptors. These feature information are fused with three dimensional feature data extracted by different algorithms. Two dimensional SIFT face features show successful results in scale and rotation invariance whereas moderate results in matching performance [17].

Moreels and Perona evaluated popular feature detectors along with descriptors for comparing their matching performance in [18]. They set up controlled experiments where they obtained “100 objects viewed from 144 calibrated viewpoints under three different lighting conditions”. According to their results, SIFT feature detector performed best among other compared descriptors [18].

Fraundorfer and Bischof also compared SIFT feature detector among other popular feature detectors used with SIFT descriptors [22]. They compared the repeatability scores of detectors with increasing viewpoint change. They used two types of scenes in visual-band one of which is an image of two boxes, the other one is a scene, showing different objects in part of a room. SIFT feature detector provided one of the high repeatability and “number of correspondences” performances for both scenes [22]. The repeatability and number of correspondences decreased as viewpoints of images increased, as expected [22].

Speeded Up Robust Feature (SURF) detector and descriptor was recently developed by Bay et.al [3]. This detector and descriptor algorithm has a fast execution time

and efficient match finding performance, when compared with other feature detectors and descriptors. Fast operation is achieved by the use of integral images for image convolutions, which is firstly used in the study of Viola and Jones [25]. SURF feature detector is based on the Hessian detector, but a different version of it, used with box-type convolution filters with pyramidal structure. Authors call it “Fast Hessian” detector for this reason. SURF descriptor is a distribution of “Haar wavelet” responses in the neighbourhood of each feature point [3].

3.1.2 Interest Point Detectors in Infrared Bands

Infrared-band images have been used widely in Computer Vision and Pattern Recognition areas recently. There are a lot of algorithms developed in literature for the tracking, recognition and vision analysis of patterns of infrared images belonging to humans especially. Most of these algorithms are out of scope of this study. The use of infrared images, especially thermal sensor images or long-wave infrared-band images, have been widely used in face recognition [26], [27], [28], pedestrian detection [29], [30], human or people tracking [31], background modeling [32], aircraft recognition [33], and geographic image retrieval [34] studies.

Infrared-band images have significant advantages over visual-band images in terms of detection of interest points. Firstly, infrared-band images, especially thermal or long-wave infrared-band images, are independent from illumination changes. In day time, visual-band images are the results of reflection of sunlight from the objects we see or reflection of external illumination. When there is shadow or dim light on objects, visual-band images’ quality suffer, and less interest points can be obtained. In night time with no external illumination, visual-band images do not give any information about the viewed target, so feature points can not be extracted. The images of objects at night or under low illumination can be obtained with quite good quality in infrared bands and features of these images can be extracted easily for other applications [27], [30], [31]. This property of infrared or thermal image is

used especially in aircraft detection, face recognition and pedestrian detection studies.

Second advantage of infrared-band images is that they are less affected by scattering and absorption by smoke and dust in the environment [27]. This advantage provides good visibility of the target observed far away and while visible image quality suffers under smoke and dust conditions, far away target image quality is good for infrared-bands. Hence, under smoke and dust conditions, sufficient feature points can be extracted from infrared-band images of objects for special applications.

Thirdly, infrared-band images of three dimensional objects from closer viewpoints provide better quality features and feature regions when compared with visual-band images of the same objects or targets [27]. This advantage is especially important, when the target is a human face or part of a human body.

As a result of the abovementioned facts, the motivation of this study is towards infrared-band images in order to verify the effectiveness of feature extraction and their performances in such cases or conditions. These verifications will be discussed in Chapter 5 of this thesis.

Scale Invariant Feature Transform (SIFT) detector was used successfully in various studies for infrared images, especially in object detection at moving platforms. Pai et.al used SIFT feature detector for moving object detection from an infrared camera on an aircraft [33]. They presented a method that detects moving objects from the aircraft just before landing. Reference frames from the video are chosen automatically by SIFT features in the neighborhood of the landing runway. The ground is assumed to be a planar surface in infrared video frames. Harris corner features [35], which is not the multiscale and scale invariant version, is also examined during the experiments, giving unstable results, since there are viewpoint and scale changes in the video frames compared. SIFT feature provided much better results as a result of the fact that it is scale invariant. In this study, SIFT feature

points of images of runway in consecutive infrared video frames are extracted and matches are found between these pairs. These matches are used to create a “background model” to be compared later for moving object detection. Their system detected a large percentage of vehicles moving in the tests. The system is also unaffected by the illumination, since infrared images are used [33].

SIFT features were used successfully in infrared images by Wang et.al [36], where mosaicking and stitching of infrared video frames were obtained. The infrared video frames and images used in their study [36], were taken from unmanned air vehicle airborne infrared imaging system. The SIFT feature points are extracted from these infrared images and matching pairs are obtained. Then, they use a novel method, namely Random M-Least Square and adopt “Multi-Resolution blending” in order to stitch the video frames. The results of their tests showed that their proposed methods work quite well with SIFT features to obtain high quality panoramic views of infrared images from UAV infrared video frames [36]. They propose to implement new studies by using new feature detection methods such as SURF feature detector [3], which is evaluated in this study for infrared images.

In a different application, Newsam, and Yang performed a successful study of geographic image retrieval from large internet dataset of infrared images by using SIFT feature points and SIFT descriptors [34]. They used 1-meter resolution, panchromatic IKONOS satellite images to evaluate their method of retrieval. IKONOS satellite images were chosen from ten sets of 64 by 64 pixel images with classes covering aqueduct, commercial, dense residential, desert chaparral, forest, freeway, intersection, parking lot, road, and rural residential [34]. The images chosen included infrared images as well taken at near-infrared-band of $0.757 \mu\text{m} - 0.853 \mu\text{m}$ [37]. SIFT feature points and corresponding SIFT descriptors of the images are extracted. They used standard Euclidean distance measure to compute similarity measures for images. Then, the test images are compared on the basis of the similarity measure and class categories of the test images are obtained [34]. The

results showed high accuracy retrieval of test images with SIFT features and descriptors.

The theoretical foundations behind these feature points for each type of detectors, SIFT, multiscale Harris-Laplace, multiscale Hessian-Laplace, and SURF detectors, are explained in the following sections. Scale space representations of the input image are obtained for each of the detection processes.

3.2 SIFT Detector

SIFT detector is based on an approximation of scale-normalized Laplacian operator. SIFT detection process uses Difference of Gaussian (DoG) scale space in order to obtain features that have local extremums over spatial coordinates and scales. Before DoG scale space, Gaussian scale space representation of the input image is constructed.

The scale space of an image is defined by $L(x,y,\sigma)$ function and it is a smoothed version of the input image $I(x,y)$ by the variable scale Gaussian kernel, $G(x,y,\sigma)$. $L(x,y,\sigma)$ is obtained by the convolution operation as follows :

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (3-1)$$

where $G(x,y,\sigma)$ is the two dimensional Gaussian kernel and $I(x,y)$ is the input image.

Gaussian kernel can be represented by the following equation.

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \quad (3-2)$$

Input image $I(x,y)$ is sampled appropriately in image pyramids consisting of certain number of octaves. Each input image is doubled in size by interpolation initially before building its scale space. The reason for this image size doubling is to make

full use of the input (assumed to be presmoothed) in other words, the image can be expanded to create more sample points than were present in the original.

The scale space is divided into octaves after this step, that is, scale space domain is discretized in logarithmic steps. Before each successive scale space octave, the image data is interpolated in size by half. The smoothed image at the mid scale level of the corresponding octave is interpolated being the next octave input image. The reason for this interpolation of the image in size by half is that it is sampled in a particular way to reduce redundancy [47]. This way of interpolation leads to a compact and efficient multi-scale representation. The application of octaves and image pyramids were introduced initially by Burt, and Adelson (1983) in their study of [38].

The scale resolution and base scale offset are input parameters for each octave of scale space. Scale resolution (S) is the number of sublevels used in building each octave of scale space, whereas base scale offset is the base smoothing standard deviation, σ in $G(x,y,\sigma)$.

At each octave of corresponding image pyramid, there are S+3 number of sublevels, in which the input image of that octave is smoothed by a two dimensional Gaussian filter $G(x,y,\sigma)$. Smoothing is a kind of low-pass filter. The Gaussian filter uses a statistical weighting function as the filter kernel. Pixels closer to the center pixel are given more weight than those at the extremities.

Each input data in each of the octave of scale space is smoothed by a constant multiplicative factor k (scale factor), where the smoothing kernel is equal to $G(x,y,k\sigma)$ and $k = 2^{1/S}$.

A visual band input image is used as a sample input for evaluating the performance of the SIFT detector algorithm at the end of major steps. The input image used as the sample to the SIFT detector is shown in Figure 3-1 below.

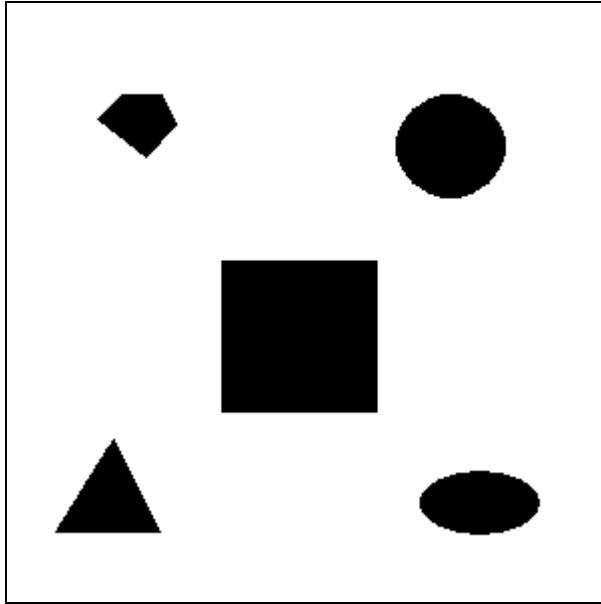


Figure 3-1 Sample Input Image

The original sample input image and the presmoothed, double sized starting image of scale space is shown in Figure 3-2 below.

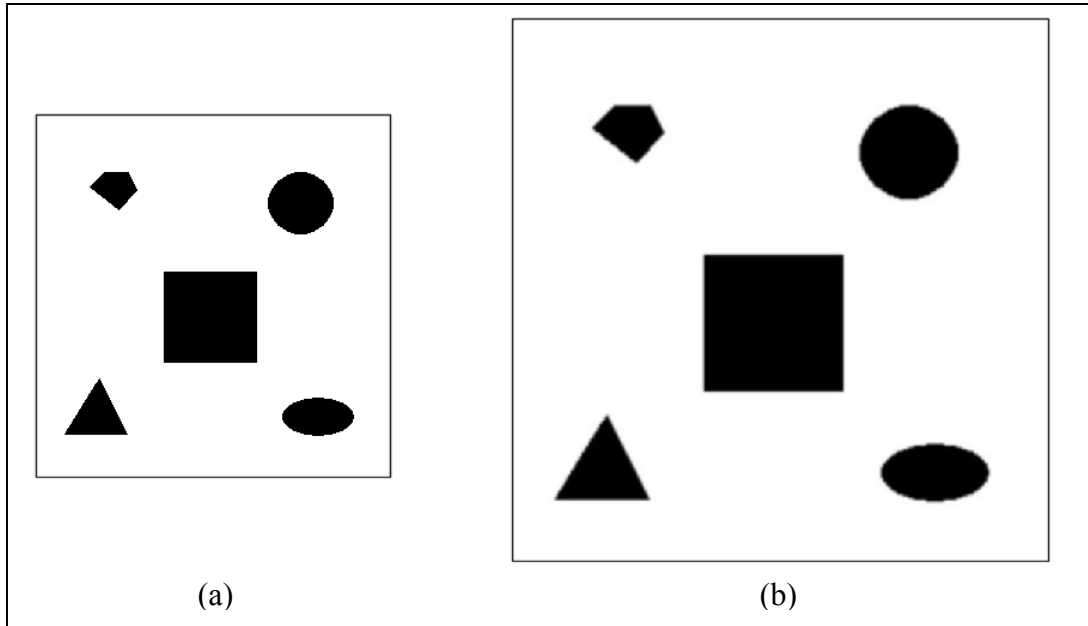


Figure 3-2 Original Sample Image and Double Size Presmoothed Starting Image

(a) Original Image, (b) Interpolated Image (Double Sized)

The outputs of Gaussian scale space of the sample input image for first four octaves are presented in Figure 3-3, Figure 3-4, Figure 3-5, Figure 3-6, Figure 3-7, Figure 3-8, Figure 3-9, and Figure 3-10 below. Two figures for each octave consisting of six scale levels are presented. The octaves after fourth octave are not shown, because the images belonging to these octaves are so small that they are not discernible.

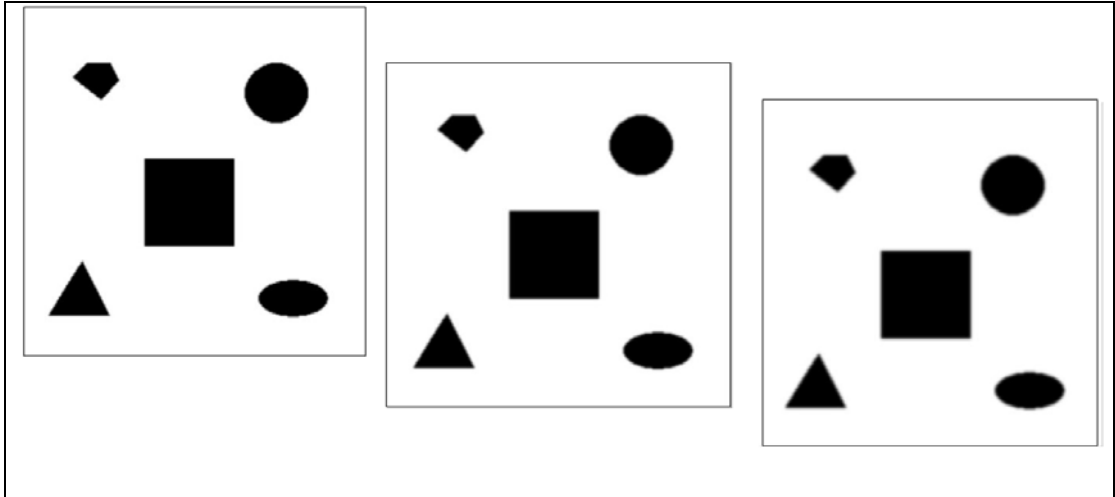


Figure 3-3 First Octave First Three Levels (1, 2, 3) of Scale

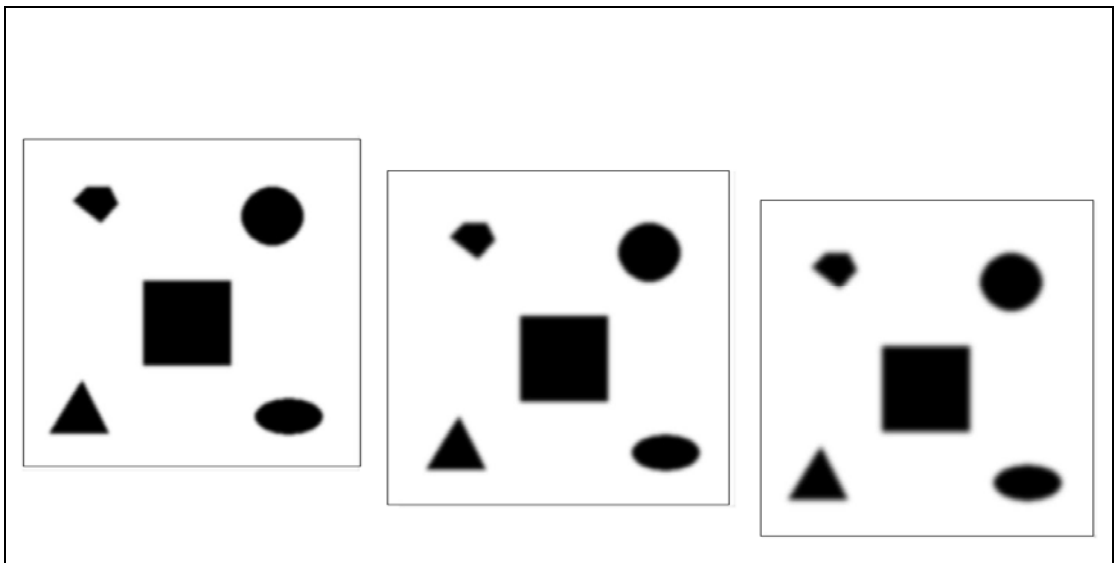


Figure 3-4 First Octave Second Three Levels (4, 5, 6) of Scale

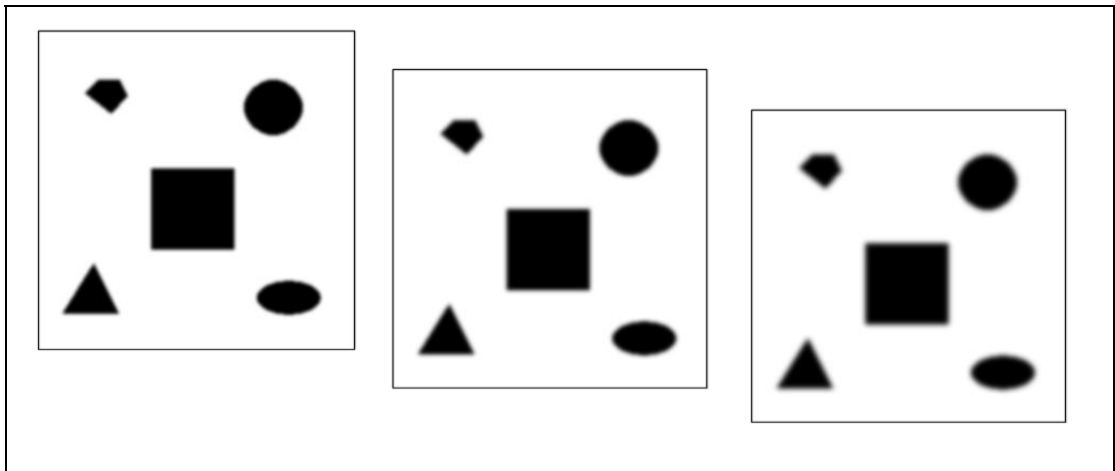


Figure 3-5 Second Octave First Three Levels (1, 2, 3) of Scale

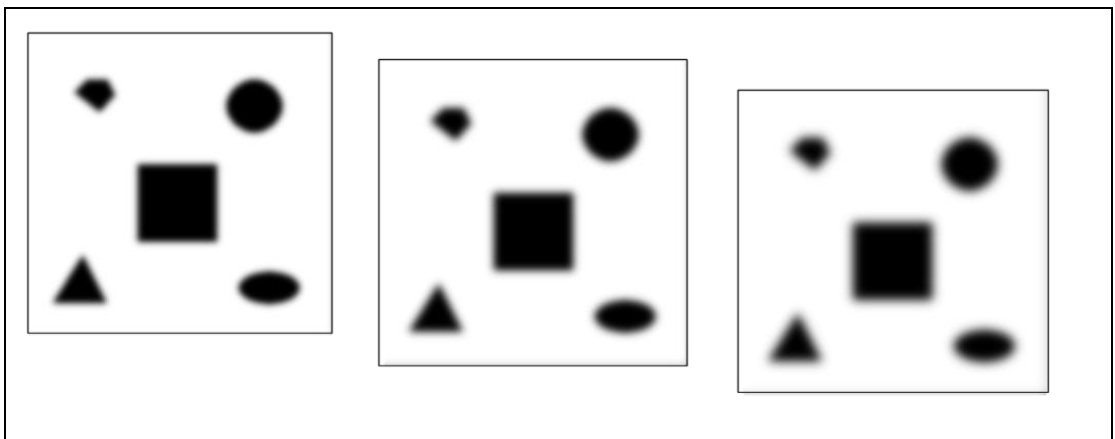


Figure 3-6 Second Octave Second Three Levels (4, 5, 6) of Scale

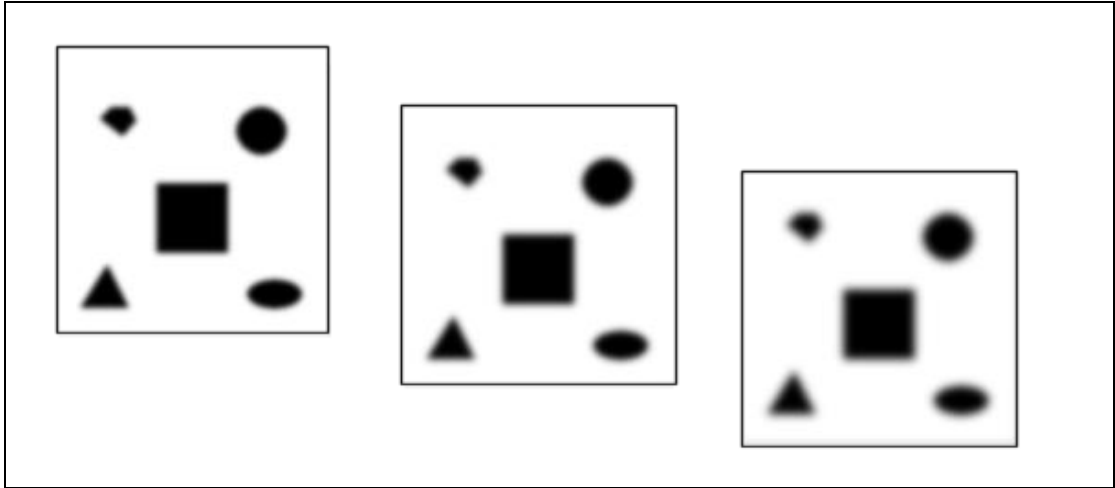


Figure 3-7 Third Octave First Three Levels (1, 2, 3) of Scale

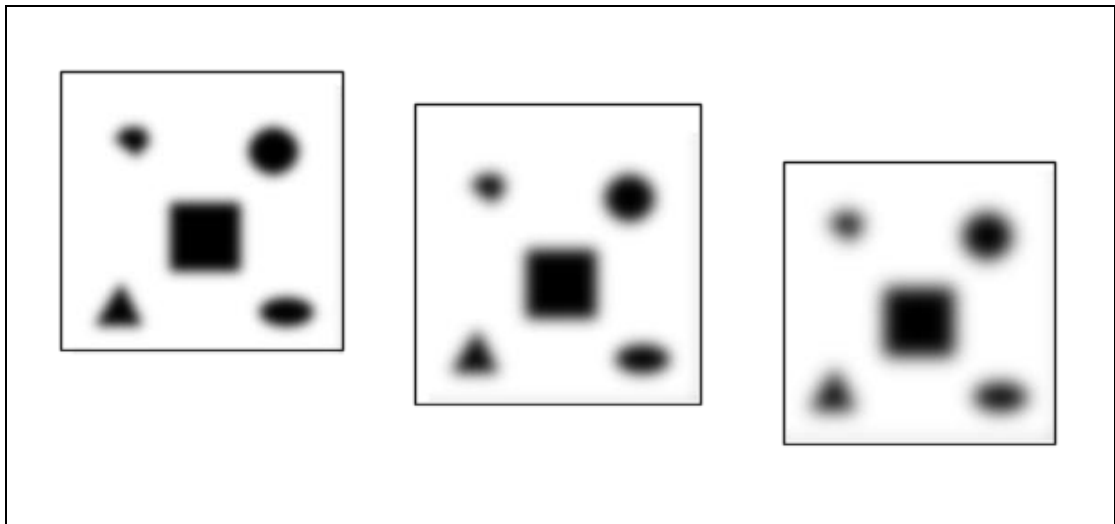


Figure 3-8 Third Octave Second Three Levels (4, 5, 6) of Scale

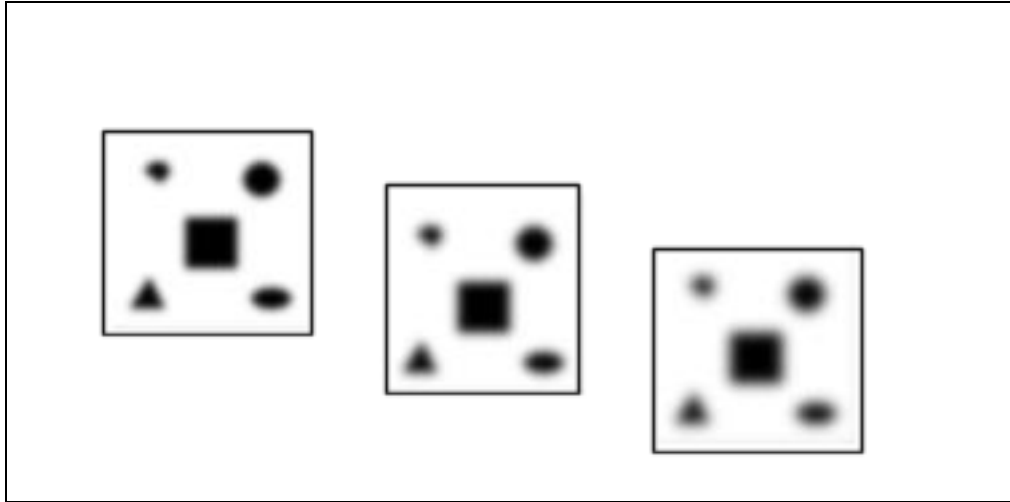


Figure 3-9 Fourth Octave First Three Levels (1, 2, 3) of Scale

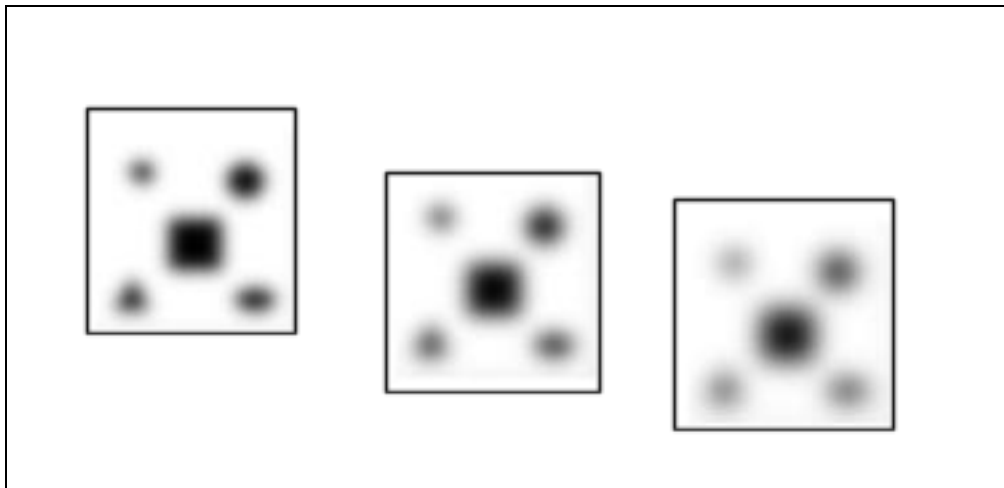


Figure 3-10 Fourth Octave Second Three Levels (4, 5, 6) of Scale

The Gaussian scale space of the input image is obtained by the aforementioned steps. Then, Difference of Gaussian (DoG) scale space is obtained by using this Gaussian scale space. The reason for this stage is that DoG is an approximation for the scale-normalized Laplacian of Gaussian operation.

Scale space is a continuous function of standard deviation, σ of the corresponding scale level. The characteristic scale of the local feature points should be found for best notion of scale. Experimentally scale-normalized LoG is used in finding the characteristic scale by obtaining the local extremum where LoG value is located [24]. An illustration figure, showing the way characteristic scale of a building scenery is obtained, is presented in Figure 3-11 below.

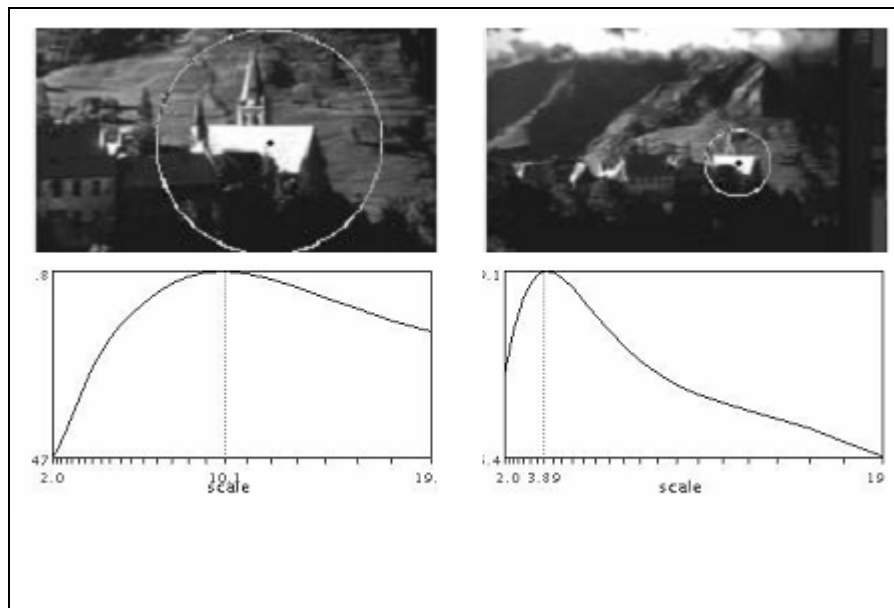


Figure 3-11 Illustration of Characteristic Scale [2].

Scale-normalized Laplacian of Gaussian detector is shown to be an accurate detector in order to find characteristic scales over the scale space of the image in Lindeberg, [12] and Mikolajczyk, and Schmid, [2]. The DoG detector is shown to be a close approximation to the scale-normalized Laplacian of Gaussian (LoG) in Lowe by the use of the Heat Diffusion equation (3-3) shown below [1] :

$$\frac{\partial G}{\partial \sigma} = \sigma \nabla^2 G \quad (3-3)$$

When $\frac{\partial G}{\partial \sigma}$ is written in its open form, Heat Diffusion equation becomes as it is shown below.

$$\sigma \nabla^2 G = \frac{\partial G}{\partial \sigma} \approx \frac{G(k\sigma) - G(\sigma)}{k\sigma - \sigma} \quad (3-4)$$

Scale-normalized Laplacian of Gaussian is $\sigma^2 \nabla^2 G$.

By multiplying $(k-1)\sigma$ of both sides of the equation (3-4), we obtain DoG operator on the right, approximately equal to scale-normalized LoG.

$$(k-1)\sigma^2 \nabla^2 G \approx G(k\sigma) - G(\sigma) \quad (3-5)$$

So, Difference of Gaussian is defined as it is shown in equation (3-6) below.

$$D(\sigma) \equiv (G(k\sigma) - G(\sigma)) * I \quad (3-6)$$

DoG can be represented by the following equation (3-7).

$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma) \quad (3-7)$$

The Difference of Gaussian scale space can be obtained straightforward after the Gaussian scale space is calculated. Due to this reason, SIFT detector is also named as DoG detector. The Difference of Gaussian scale space is calculated by subtracting two Gaussian smoothed images. Two graphical views of DoG output of the first octave first difference scale level image are presented in Figure 3-12 and Figure 3-13 below.

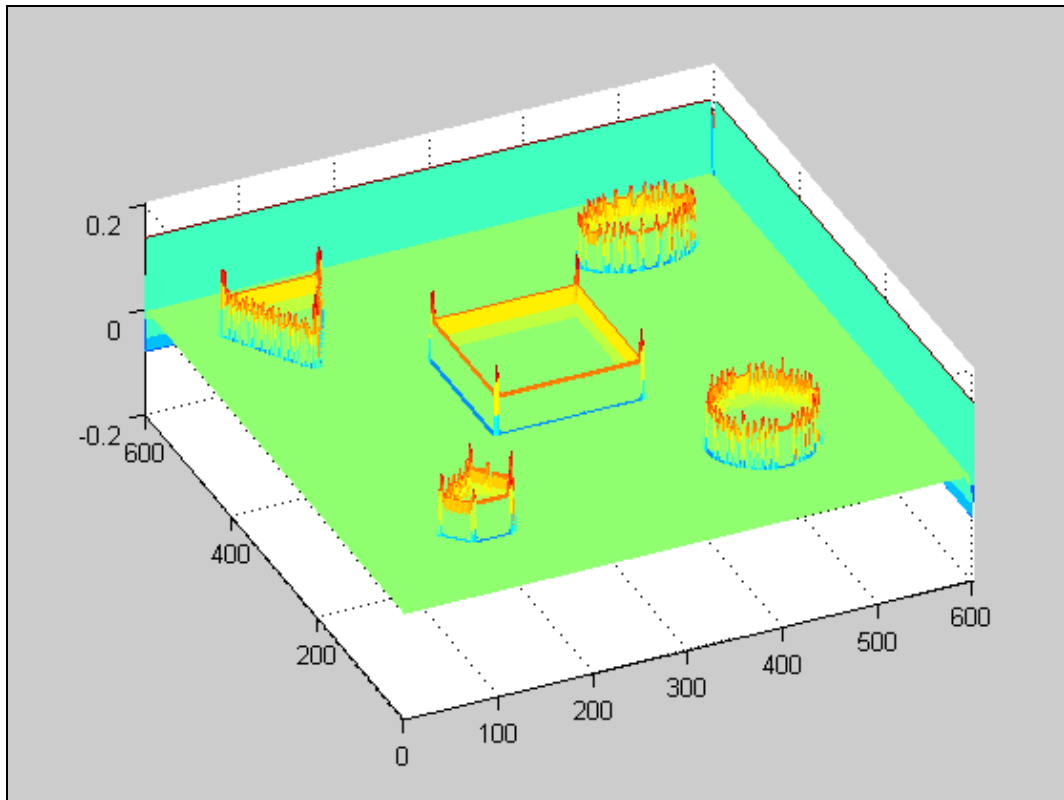


Figure 3-12 Difference of Gaussian Graphical View-1 for Sample Input Image in Figure 3-1

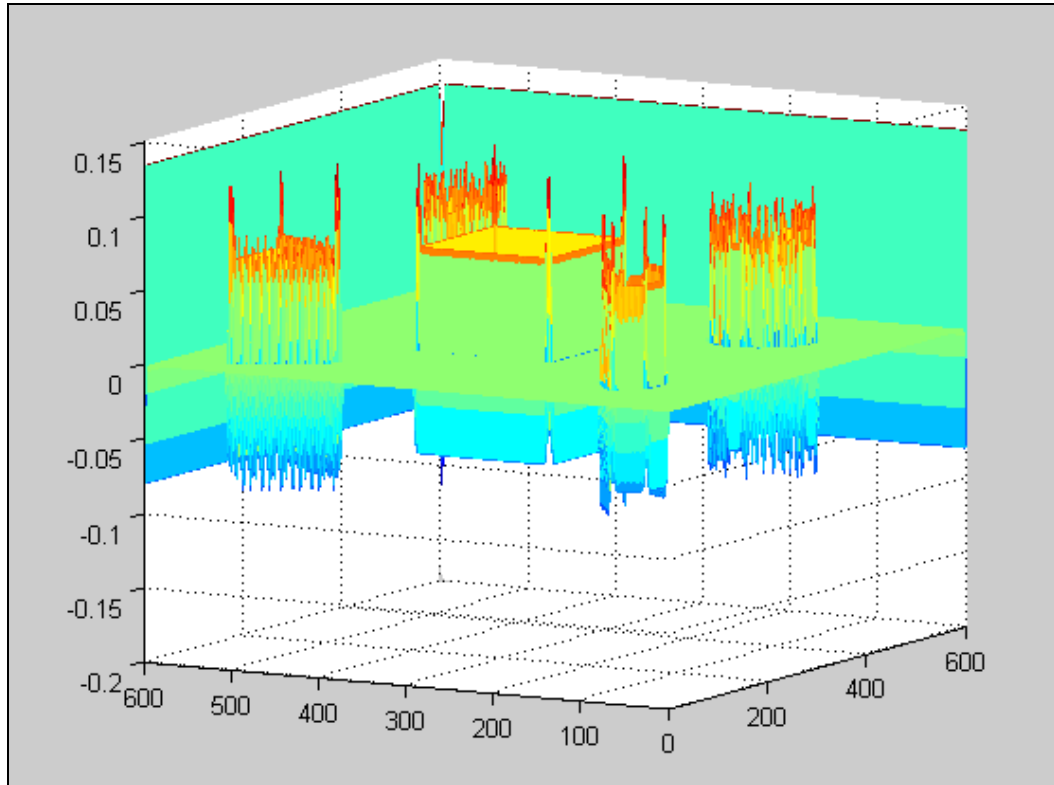


Figure 3-13 Difference of Gaussian Graphical View-2 for Sample Input Image in Figure 3-1

Difference of Gaussian (DoG) can be written in its open form as it is shown mathematically below.

$$DoG = G_{\sigma_1} - G_{\sigma_2} = \frac{1}{2\pi} \left[\frac{1}{\sigma_1} e^{-((x^2+y^2)/2\sigma_1^2)} - \frac{1}{\sigma_2} e^{-((x^2+y^2)/2\sigma_2^2)} \right] \quad (3-8)$$

Graphs belonging to both 1D and 2D functions of $G_{\sigma_1}(x, y)$ and $G_{\sigma_2}(x, y)$ and their difference are shown in Figure 3-14 and Figure 3-15 below.

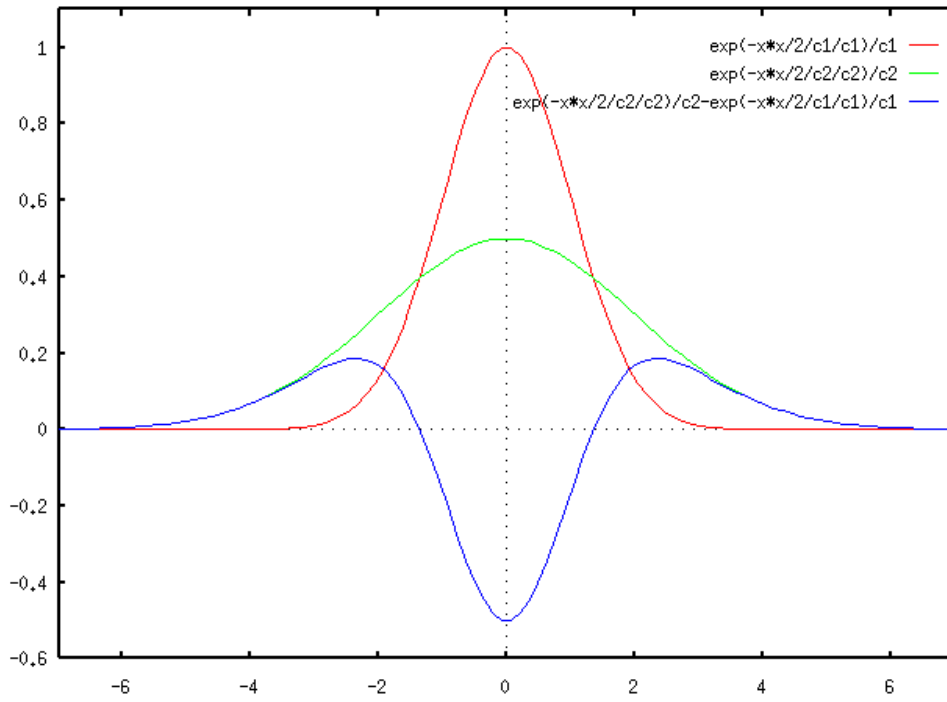


Figure 3-14 Graph of 1D Gaussian $G(x)$ and DoG(x)

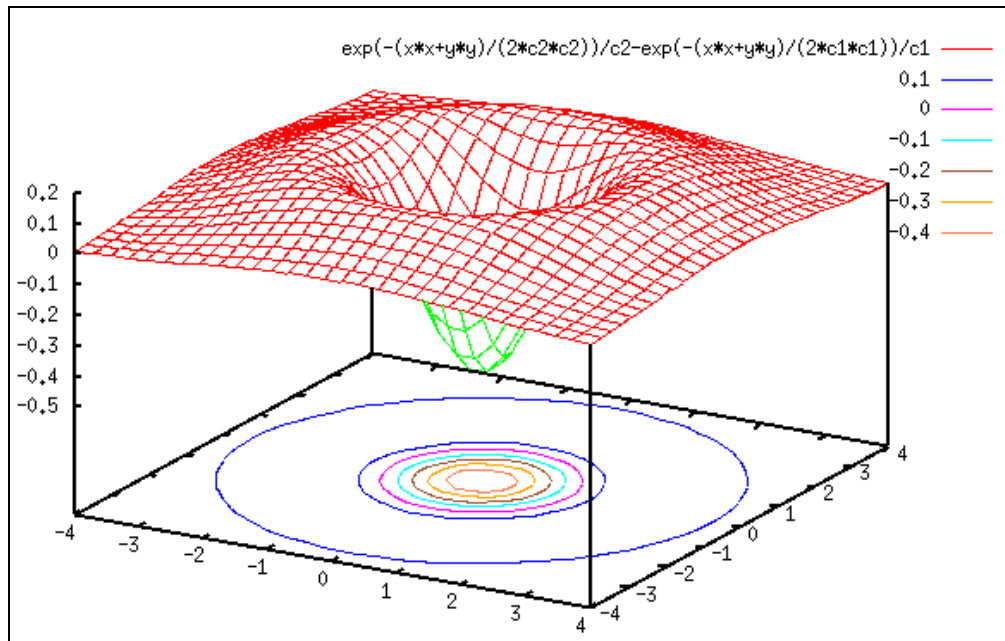


Figure 3-15 Graph of 2D DoG(x,y)

The two-dimensional Difference of Gaussian operator is similar in shape to an inverted Mexican hat. When this filter kernel is applied, the edges of the input image can be obtained similar to scale-normalized LoG operator.

The DoG output images of the sample input image for first octave and first four difference scale levels are presented in Figure 3-16 and Figure 3-17 below. As it can be seen from the images, the edges of the shapes in the input image are clear.

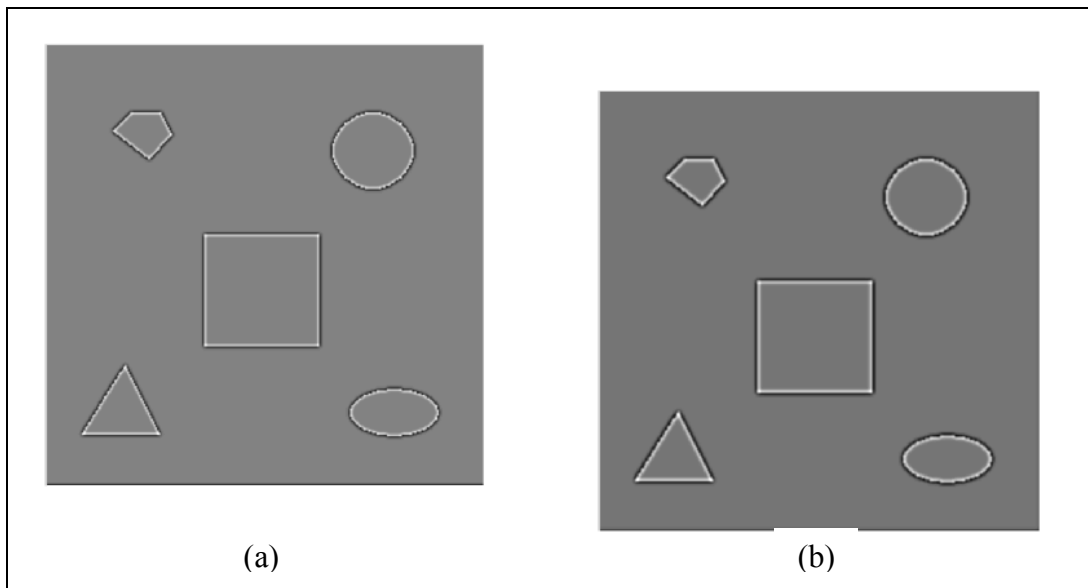


Figure 3-16 DoG Images for First Octave First Two DoG Scale Levels

(a) First Scale Difference, (b) Second Scale Difference

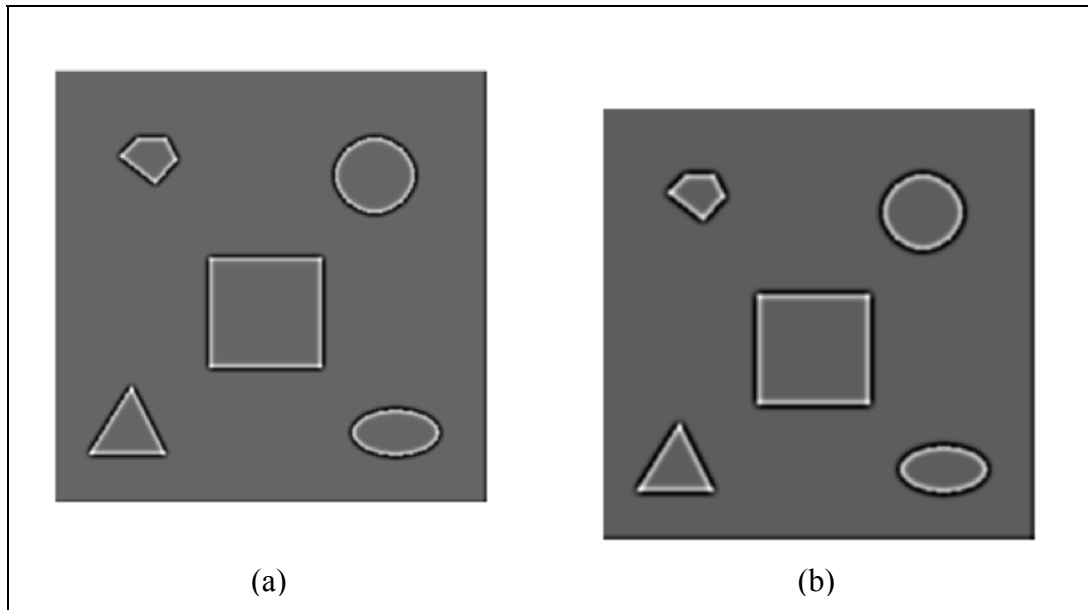


Figure 3-17 DoG Images for First Octave Second Two DoG Scale Levels

(a) Third Scale Difference, (b) Fourth Scale Difference

The local extremum points for Difference of Gaussian scale space values are actually the candidate feature points of the analyzed image. Local extremum points of DoG are the local maxima and local minima of $D(x,y,\sigma)$. The values of DoG scale space are compared with each point of the image both in spatial coordinates and in scale dimension. Each point is compared by examining its 9 neighborhood points in the current scale and 9 neighborhood points in one scale above and below. (9x9x9) The comparison of neighborhood points is shown in Figure 3-18 below.

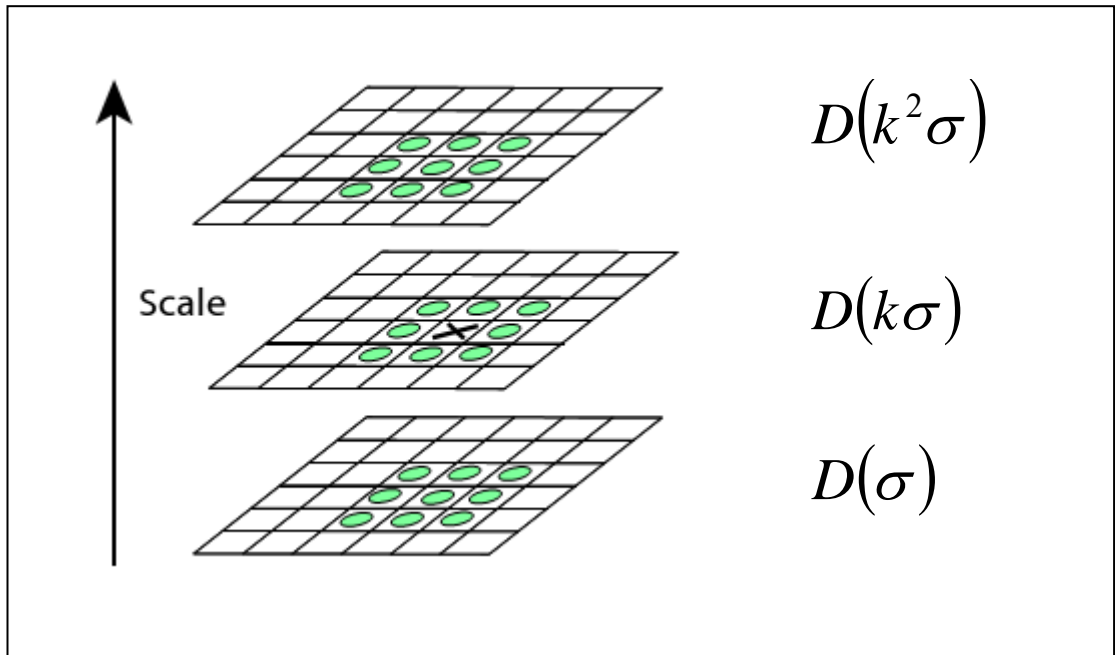


Figure 3-18 Local Extremum Point Search Over Scales and Space [1].

The initial (candidate) SIFT features are presented in Figure 3-19 below for the sample input image. Note that, there are no feature points close to the boundary of the image. Feature points close to the boundary are eliminated.

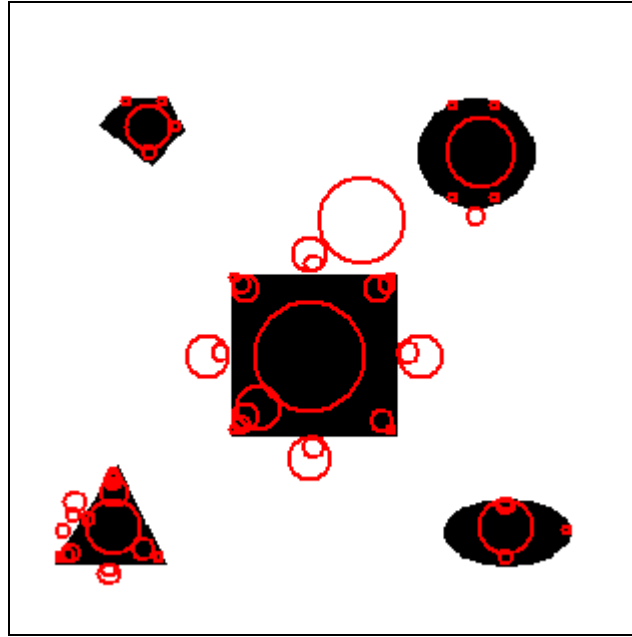


Figure 3-19 Initial Candidate SIFT Features

From difference-of-Gaussian local extrema detection, we obtain approximate values for keypoints. After initial SIFT features are obtained, we can still have undesirable feature points at especially higher scales. True extrema can be obtained by subpixel localization and refinement in this step. The typical graph showing the localization of initial keypoints is shown in Figure 3-20 below.

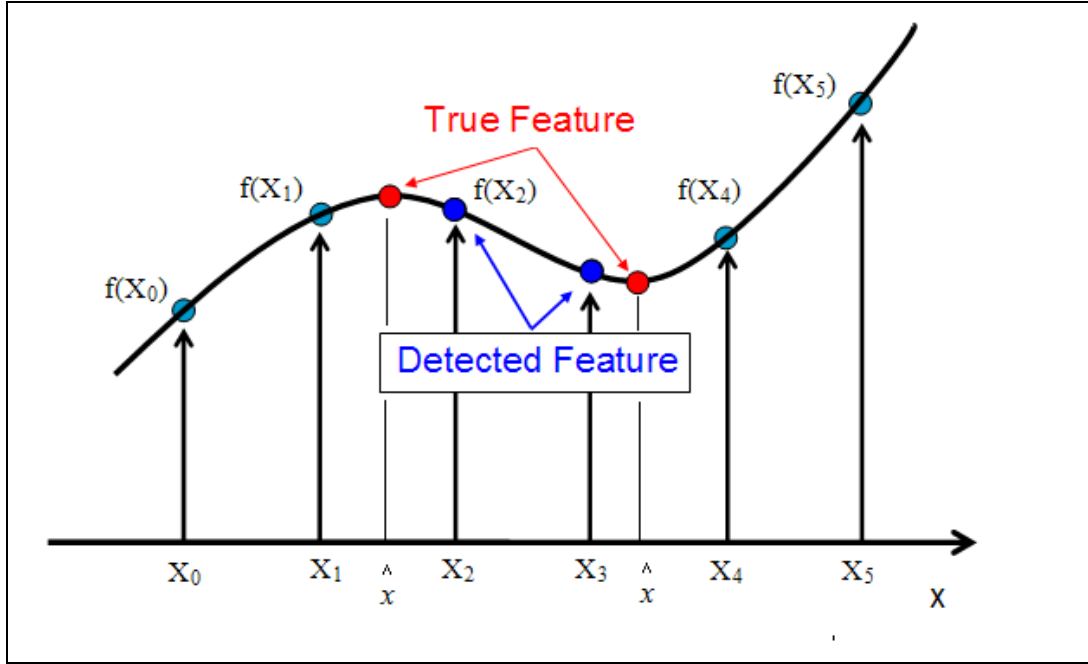


Figure 3-20 Typical Graph of True Extrema by Localization and Refinement [49].

The method and approach developed by Brown and Lowe is used for subpixel localization [48]. Taylor series expansion of Difference of Gaussian is derived up to quadratic terms as it is shown below.

$$D(\vec{x}) = D + \frac{\partial D^T}{\partial \vec{x}} \vec{x} + \frac{1}{2} \vec{x}^T \frac{\partial^2 D^T}{\partial \vec{x}^2} \vec{x} \quad (3-9)$$

In the derivation of Taylor series expansion of DoG, D is shifted so that the origin is at the sample, initial feature point. The location of the true extremum \hat{x} (offset) is determined by taking the derivative of DoG with respect to x and setting it to zero. \hat{x} is given by the following equation.

$$\hat{x} = - \frac{\partial^2 D^{-1}}{\partial x^2} \frac{\partial D}{\partial x} \quad (3-10)$$

Hessian and gradient of DoG around each initial feature point is calculated by approximation of differences of neighboring sample points. If the offset \hat{x} is larger than 0.5 in any dimension, then final offset is added to the location of initial feature point to get the interpolated estimate for the true location of the extremum. The function value at the extremum point $D(\hat{x})$ is determined and it is used to reject local extrema with low contrast, below a threshold value of 0.03 [1].

DoG has strong response along edges. Harris-Stephens corner detection algorithm is applied at this stage in order to eliminate edges and get feature points with strong cornerness values or low ratio of principle curvature values. Ratio of principle curvatures is defined as it is shown below.

$$\text{Principle_curvature} = \frac{\text{trace}^2(H)}{\det(H)} \quad (3-11)$$

where H is the Hessian matrix of each feature point.

Difference of Gaussian values will have large principle curvatures across the edges, but they will have small values when two edges combine vertically, which is actually a corner. In order to eliminate edge responses of initial feature points, Hessian matrix of each feature point is calculated around these feature points. Then, the criterion shown in equation (3-12) below is checked. This criterion is used to reject points with strong edge responses. (i.e. accept the ones with high cornerness such as the ones in Harris)

$$\frac{\text{trace}^2(H)}{\det(H)} < \frac{(r+1)^2}{r} \quad (3-12)$$

where r is the ratio of largest magnitude eigenvalue α to the smallest magnitude eigenvalue β , belonging to the trace and determinant of Hessian of DoG matrix [1]. r is also called as edge threshold value and its value is taken as 10 [39].

This criterion satisfies the same condition that the following cornerness measure in (3-13) be maximum at the candidate feature point.

$$\text{cornerness} = \det(H) - \alpha \text{trace}^2(H) \quad (3-13)$$

where α is a constant different from the eigenvalue of Hessian of DoG

(with $\text{cornerness} > 0.274$ (when $\alpha=0.06, r=10$)).

Final SIFT feature points are extracted with their spatial coordinates x, y and scale parameter, at the end of this refinement step. Refined SIFT features are shown in Figure 3-21 below for the sample input image.

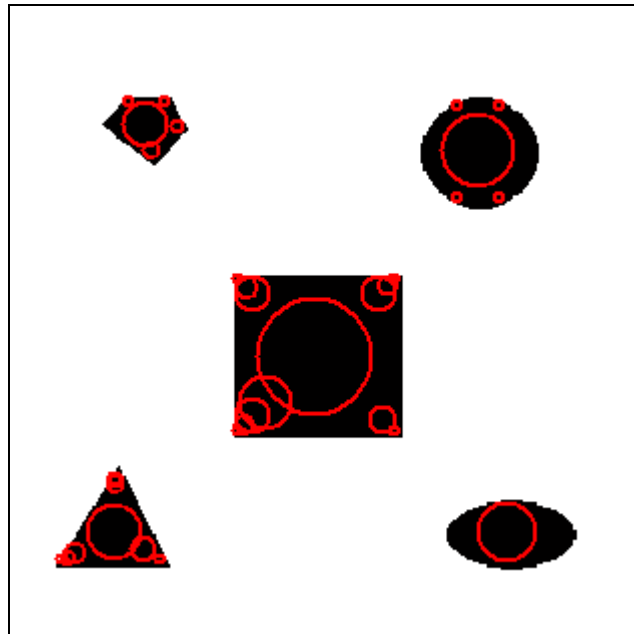


Figure 3-21 Final SIFT Features After Refinement

The summary of the steps used in SIFT feature detector are presented as a pseudocode in Table 3-1, SIFT algorithm shown below.

Table 3-1 SIFT Algorithm Steps

SIFT Algorithm

- Gaussian Scale Space of input image is constructed.
 - Interpolation of input image in double size initially
 - Smooth double sized input image with Gaussian filter

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}$$
 - Continue smoothing incrementally with $G(x,y,k\sigma)$ where $k = \sqrt{2}$ for first octave of scale space
 - Interpolation of the smoothed image in halve size for next octave input
 - Smooth interpolated image with Gaussian filter

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}$$
 - Continue smoothing incrementally with $G(x,y,k\sigma)$ where $k = \sqrt{2}$ for second octave of scale space
 - Continue until number of octaves corresponding to the size of interpolated images is less than 16x16 pixels
 - End, Gaussian scale space $L(x,y,\sigma)$ is obtained.
- Difference of Gaussian Scale Space of input image is constructed.
 - Difference of Gaussian smoothed images between next scale level and the current scale level is calculated.

$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma)$$
 - Continue this operation for all octave of scale space.
 - End, Difference of Gaussian scale space $D(x,y,\sigma)$ is obtained.

SIFT Algorithm (Continued)

- Finding local extrema of DoG scale space
 - Local maxima of DoG image points are found in 9 neighborhood points at the current scale, 9 neighborhood points in one scale above and below.
 - Local minima of DoG image points are found in 9 neighborhood points at the current scale, 9 neighborhood points in one scale above and below.
 - Feature points close to the boundary are eliminated.
- Subpixel localization and refinement
 - Location of true extremum \hat{x} (offset) is determined by

$$\hat{x} = -\frac{\partial^2 D^{-1}}{\partial x^2} \frac{\partial D}{\partial x}$$
 - If the offset \hat{x} is larger than 0.5 in any dimension, then final offset is added to the location of initial feature point to get the interpolated estimate for the true location of the extremum.
 - Local extrema with low contrast $D(\hat{x})$ is rejected.
 - Feature points with strong edge responses are eliminated with the following criterion.

$$\frac{\text{trace}^2(H)}{\det(H)} < \frac{(r+1)^2}{r}$$
 - H is the Hessian matrix of feature point around its neighborhood, r is the edge threshold value.
 - End, final SIFT feature points are obtained.

SIFT scale space representation and detector implementation of Vedaldi, is used for the extraction of SIFT feature points and SIFT descriptors in this thesis [39].

3.3 Multiscale Harris-Laplace Detector

Multiscale Harris-Laplace detector consists of two detectors which are multiscale Harris detector and Laplace detector combined together. Traditional Harris detector developed by Harris and Stephens (1988), extracts corner points of an image [35]. Multiscale version of Harris detector provides scale invariance to traditional Harris detector and corner detection becomes scale invariant at the same time. Laplace detector component of the process provides “Automatic Scale Selection” as described in Mikolajczyk and Schmid [2].

The diagram showing how Harris-Laplace feature points are detected is presented in Figure 3-22 below.

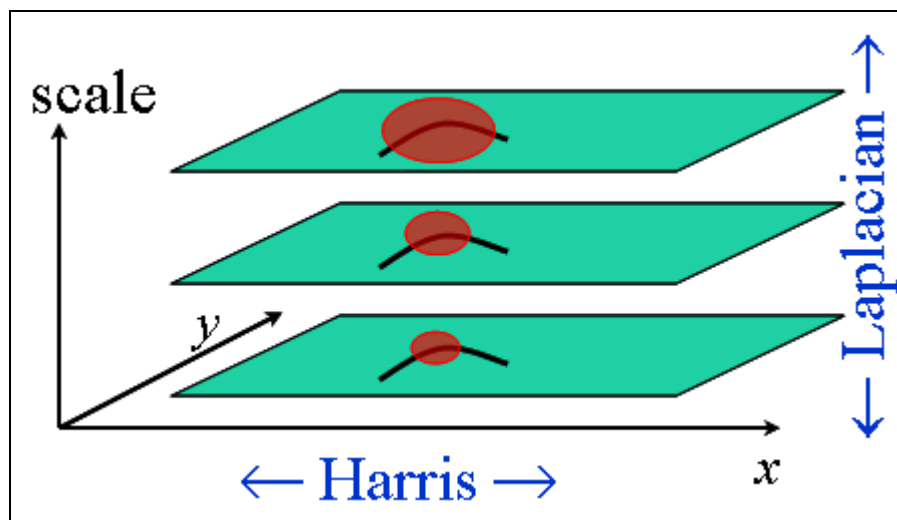


Figure 3-22 Diagram of Harris-Laplace Detection [51].

Firstly, the Harris corners are detected in space and scale coordinates by the multiscale Harris corner detector. Then, the scale of the initial Harris feature point is checked whether or not it is the characteristic scale by the Laplace detector.

The scale space representation of the input image is built, before multiscale Harris detector is applied. First order Gaussian differentials are needed in scale space representation for the multiscale Harris detector. So, these first order Gaussian derivatives of the input image are constructed in scale space. One of the most common technique to obtain these type of differentials is to use basic derivative kernels and then apply smoothing with Gaussian derivative function of required standard deviation at these filter kernels.

The first order basic differential filter kernels used are shown below.

$$h_x = [1 \quad 0 \quad -1] \quad h_y = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \quad (3-14)$$

The following first order Gaussian derivative function is applied at these filter kernels.

$$\frac{\partial G}{\partial x} = -\frac{x}{\sigma^3 \sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} \quad (3-15)$$

As a result, the filter kernels $\frac{\partial G}{\partial x}$ and $\frac{\partial G}{\partial y}$ are obtained for x and y direction

respectively. $\frac{\partial G}{\partial y}$ filter is the transpose of $\frac{\partial G}{\partial x}$ filter kernel.

First order partial derivatives of $L(x,y,\sigma)$ in x and y directions are calculated at each scale. Theoretically, these partial differentials are shown in the equations below.

$$L_x = \frac{\partial L}{\partial x} \quad L_y = \frac{\partial L}{\partial y} \quad (3-16)$$

When these partial derivatives of $L(x,y,\sigma)$ are written in their open form, the following equations 3-17 and 3-18 are obtained for these first order partial derivatives of the input image L_x and L_y .

$$L_x = \frac{\partial}{\partial x}(G(x, y, \sigma) * I(x, y)) \quad (3-17)$$

$$L_y = \frac{\partial}{\partial y}(G(x, y, \sigma) * I(x, y)) \quad (3-18)$$

Convolution operator has the mathematical differentiation property which is shown below.

$$\frac{\partial}{\partial x}(f * g) = \frac{\partial f}{\partial x} * g(x) \quad (3-19)$$

L_x can be written as follows, by using this property :

$$L_x = \frac{\partial G}{\partial x} * I(x, y) \quad (3-20)$$

The first order partial Gaussian derivatives of the input image L_x and L_y are obtained by taking the convolutions of $\frac{\partial G}{\partial x}$ and $\frac{\partial G}{\partial y}$ with input image directly.

Octaves and image pyramids are not used for this detector. Hence, octave parameters and interpolation of input image are not applied in this differential Harris scale space representation.

Same input image is used as a sample input for evaluating the performance of the multiscale Harris-Laplace detector algorithm at the end of major steps. The input image used as the sample to this detector is shown in Figure 3-1 before.

The outputs, differential Harris scale space of the sample input image are presented in Figure 3-23, Figure 3-24, Figure 3-25, and Figure 3-26 below. First order Gaussian derivative images L_x and L_y are shown for only first six levels of scale.

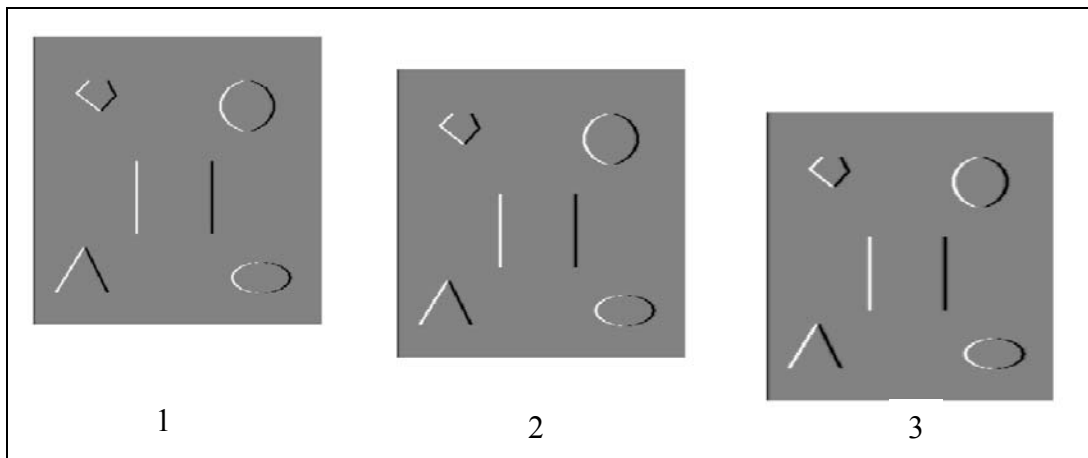


Figure 3-23 First Order Gaussian Derivative Images-1, L_x

(First Three Scale Levels, 1, 2, 3)

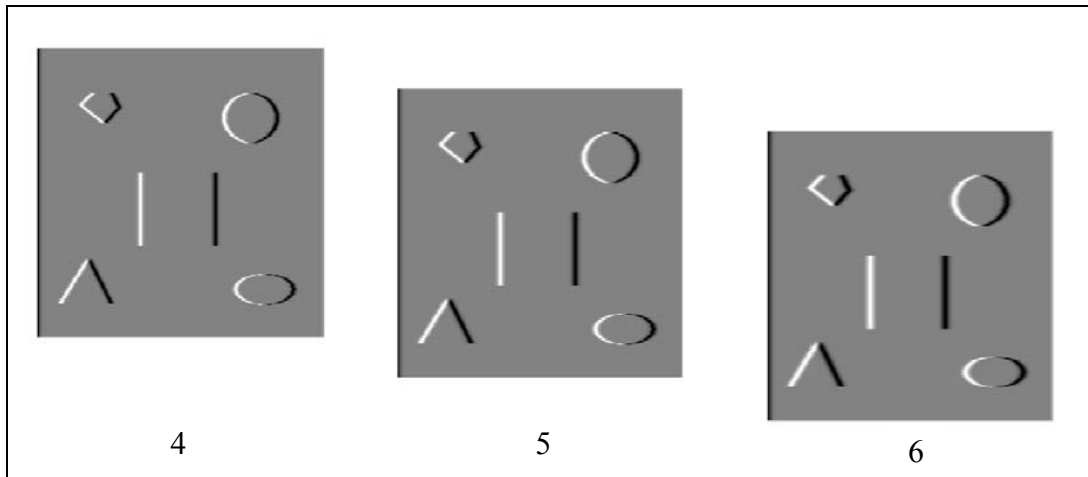


Figure 3-24 First Order Gaussian Derivative Images-2, L_x

(Second Three Scale Levels, 4, 5, 6)

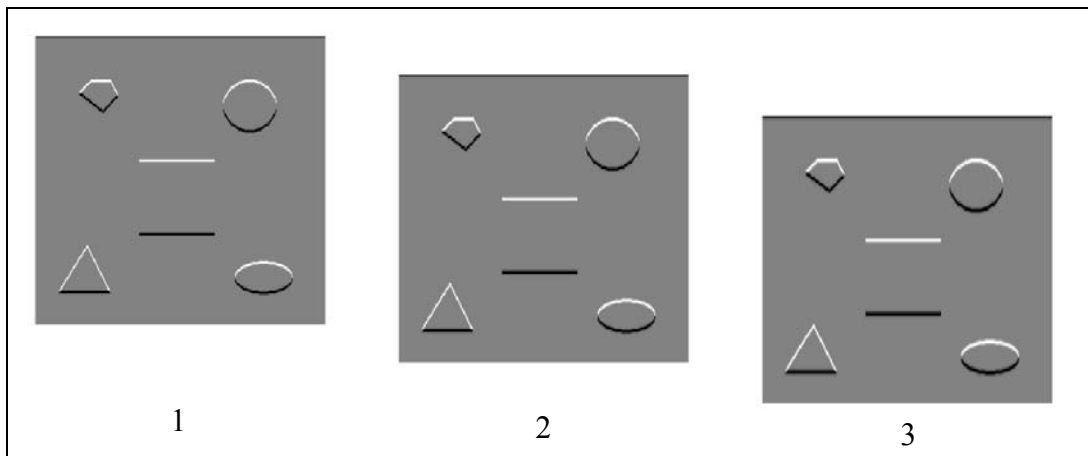


Figure 3-25 First Order Gaussian Derivative Images-1, L_y

(First Three Scale Levels, 1, 2, 3)

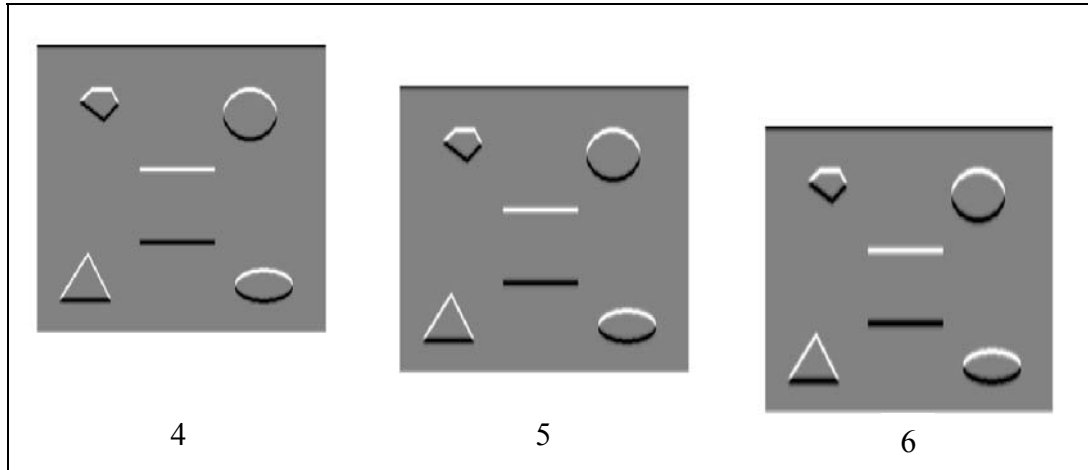


Figure 3-26 First Order Gaussian Derivative Images-2, Ly

(Second Three Scale Levels, 4, 5, 6)

The local measure used in Harris corner detection is based on the distinctness measure. This measure is represented mathematically shown below.

$$m(x, y)_{(\Delta x, \Delta y)} = \left(\int_R (I(p, q) - I(p + \Delta x, q + \Delta y))^2 dR \right) \quad (3-21)$$

where R represents a region around the point considered. This measure is actually the correlation value of a patch of pixels with its neighboring pixels. It gives the change in the value of the function in the corresponding region.

If the value of $m(x, y)$ is high, then the point is said to be significantly distinct. The second term inside the integral can be written approximately by using Taylor series expansion as it is shown below.

$$I(x + \Delta x, y + \Delta y) \approx I(x, y) + \Delta x I_x(x, y) + \Delta y I_y(x, y) \quad (3-22)$$

When this expression is substituted in the expression 3-21 of distinctness measure, the first terms disappear. The integral term can be expressed with a summation

since the image intensity values are discrete. Then, the distinctness measure can be written as it is shown below.

$$m(x, y)_{(\Delta x, \Delta y)} \approx \sum_R \left(\begin{bmatrix} I_x & I_y \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} \right)^2 = \begin{bmatrix} \Delta x & \Delta y \end{bmatrix} \underbrace{\begin{bmatrix} \sum_R I_x I_x & \sum_R I_x I_y \\ \sum_R I_x I_y & \sum_R I_y I_y \end{bmatrix}}_{C'} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} \quad (3-23)$$

If matrix C' has two significant high eigenvalues, then distinctness measure will also be large indicating that the point is distinct. This matrix is therefore taken as the corresponding measure for Harris corner detection. It is modified by Gaussian weighing as it is shown below.

$$C(x, y) = \begin{bmatrix} \sum_R w_R I_x^2 & \sum_R w_R I_x I_y \\ \sum_R w_R I_x I_y & \sum_R w_R I_y^2 \end{bmatrix} \quad (3-24)$$

where R is the region of interest, w_R is the Gaussian weighing function in this region.

The matrix $C(x, y)$ is modified further to obtain multiscale version of Harris corner detector. Scale-adapted second moment matrix is obtained by using the mathematical representation 3-25 shown below.

$$\mu(x, \sigma_I, \sigma_D) = \sigma_D^2 G(\sigma_I) * \begin{bmatrix} L_x^2(x, \sigma_D) & L_x L_y(x, \sigma_D) \\ L_x L_y(x, \sigma_D) & L_y^2(x, \sigma_D) \end{bmatrix} \quad (3-25)$$

It is difficult to calculate the eigenvalues of this matrix. Therefore, the product of eigenvalues is computed, which is called cornerness measure. This value is used as the local measure for detecting Harris corners.

Cornerness (Harris) measure of the Scale-adapted second moment matrix (Autocorrelation matrix) is calculated directly by using the equation shown below :

$$\text{cornerness} = \det(\mu(x, \sigma_I, \sigma_D)) - \alpha \text{trace}^2(\mu(x, \sigma_I, \sigma_D)) \quad (3-26)$$

Cornerness measure consists of an α constant which is taken as 0.06 as in traditional Harris detectors. Cornerness is also known as corner response function.

The scale resolution and integration scale parameters are input parameters for building Harris scale space. Scale resolution (S) is the number of sublevels used in building the scale space. The local scale parameter σ_D is used in calculation of local derivatives of $L(x,y,\sigma)$. Local scale σ_D is taken as 0.7 times integration scale σ_I as suggested in Mikolajczyk and Schmid [2]. Simplified Harris-Laplace approach is used in finding the Harris-Laplace feature points. Therefore, the input image in scale space is smoothed by a constant multiplicative factor k , namely, scale factor, where the smoothing kernel is equal to derivative of $G(x,y,k\sigma)$ kernel and $k=1.2$. (Small scale factor of 1.2 is suggested for simplified Harris-Laplace in Mikolajczyk and Schmid [2].)

The components of the autocorrelation matrix are computed at this stage. After multiplication values are calculated, the values are averaged by a rotationally symmetric Gaussian filter with standard deviation σ_I (integration scale).

Cornerness values are calculated for each point of the image. Initial (candidate) Harris-Laplace feature points are extracted by finding the local maxima of cornerness measure in a circular region of 6 times σ_I around each point of the image. Local maximum points are stored and the initial feature points are obtained, whose cornerness values are greater than some local maximum threshold (with $\text{cornerness} > \text{threshold}$).

The initial Harris-Laplace feature points of the sample input image are shown in Figure 3-27 below. Note that, all of the corners belonging to each scale levels are found.

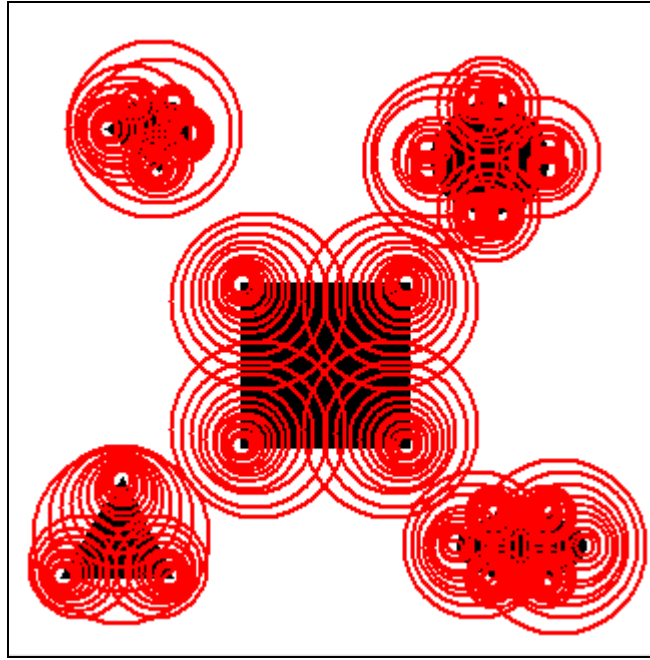


Figure 3-27 Initial Harris-Laplace Feature Points

Laplace detector utilizes one of the most common measures, scale-normalized Laplacian of Gaussian. It is commonly used in order to detect blobs in an image. Laplacian of Gaussian consist of two parts, one of which is Laplacian operator ∇^2 and the other one Gaussian filter.

Laplacian operator is represented mathematically as it is shown below.

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \quad (3-27)$$

When Laplacian operator is applied in Gaussian filtered image, the following result is obtained.

$$h(x, y) = \sigma^2 \cdot \nabla^2 (G(x, y, \sigma) * I(x, y)) \quad (3-28)$$

The σ^2 factor is used at the beginning of this expression 3-28, since $h(x,y)$ is scale-normalized Laplacian of Gaussian.

Convolution operator has the following property.

$$\nabla^2(f * g) = (\nabla^2 f) * g(x) \quad (3-29)$$

By using this property in 3-28, following scale-normalized result is obtained.

$$h(x, y) = (\sigma^2 \nabla^2 G(x, y, \sigma)) * I(x, y) \quad (3-30)$$

The scale-normalized Laplacian of Gaussian (LoG) filter kernel can be written as it is shown below.

$$\sigma^2 \nabla^2 G(x, y, \sigma) = \left(\frac{x^2 + y^2 - 2\sigma^2}{\sigma^4} \right) e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (3-31)$$

The three dimensional graph and side view of scale-normalized LoG filter kernel is shown in Figure 3-28 and Figure 3-29 below.

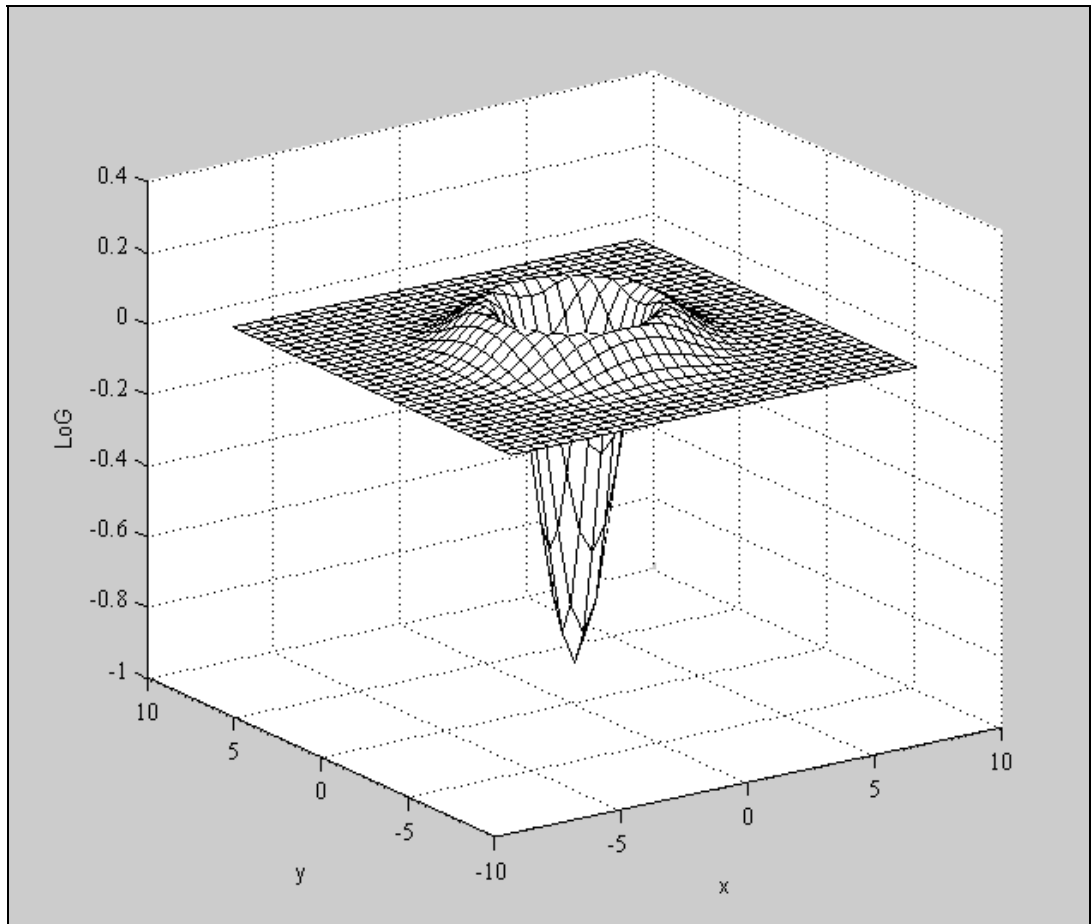


Figure 3-28 3D Graph of Scale-normalized Laplacian of Gaussian

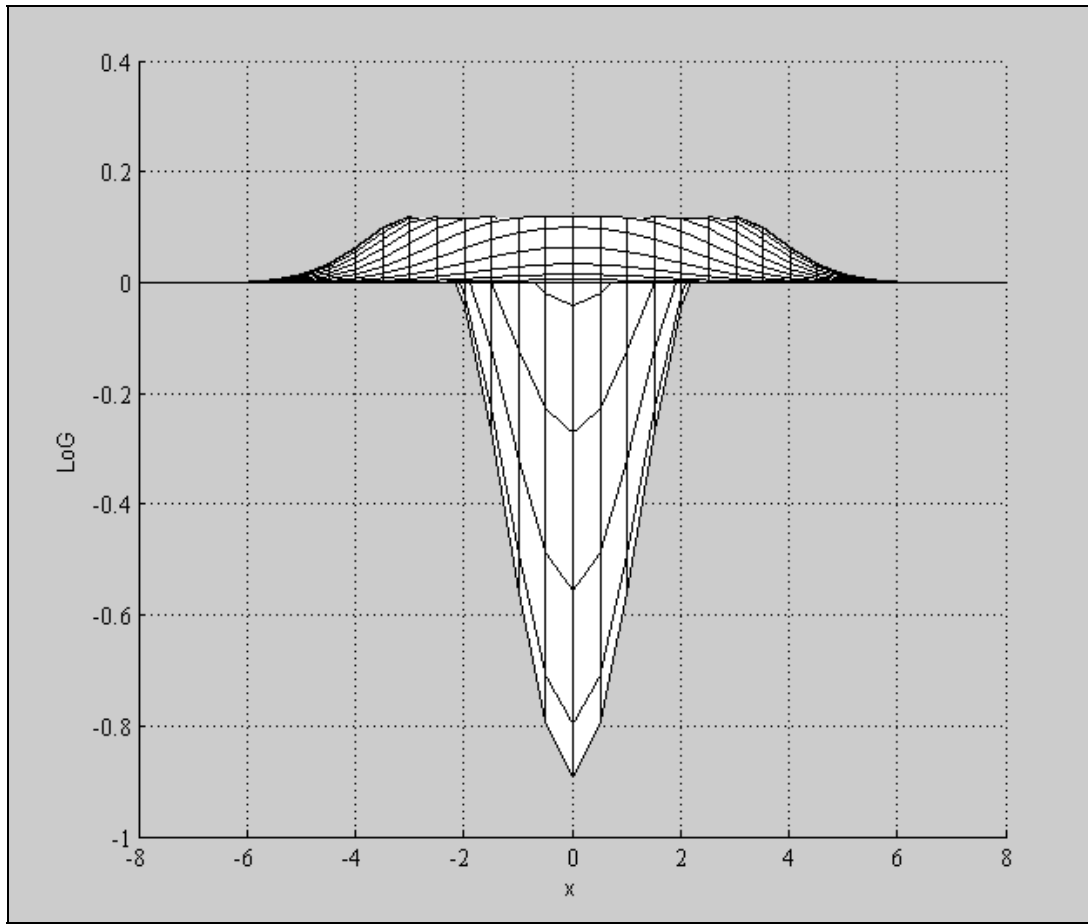


Figure 3-29 Side View of Scale-normalized Laplacian of Gaussian

When scale-normalized Laplacian of Gaussian filter is convolved with input image, it usually results in strong positive responses for dark blobs and strong negative responses for bright blobs of similar size. Therefore, either of the local extrema gives dark blobs (local maxima) and bright blobs (local minima) of the image. The scale-normalized LoG filter is similar in shape to an inverted Mexican hat.

Scale-normalized Laplacian of Gaussian is also represented mathematically as shown below.

$$LoG(x, \sigma_1) = \sigma_1^2 (L_{xx}(x, \sigma_1) + L_{yy}(x, \sigma_1)) \quad (3-32)$$

The Laplacian of Gaussian of the input image is obtained for each integration scale level, σ_1 . The scale-normalized Laplacian of Gaussian filtered images are shown in Figure 3-30 and Figure 3-31 below for the sample input image.

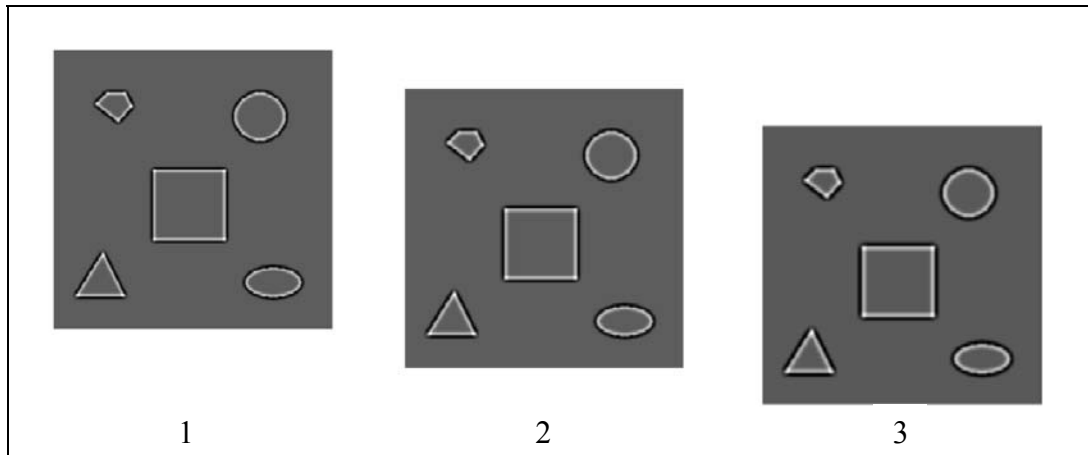


Figure 3-30 Scale-normalized Laplacian of Gaussian Filtered Images - 1

(First Three Scale Levels, 1, 2, 3)

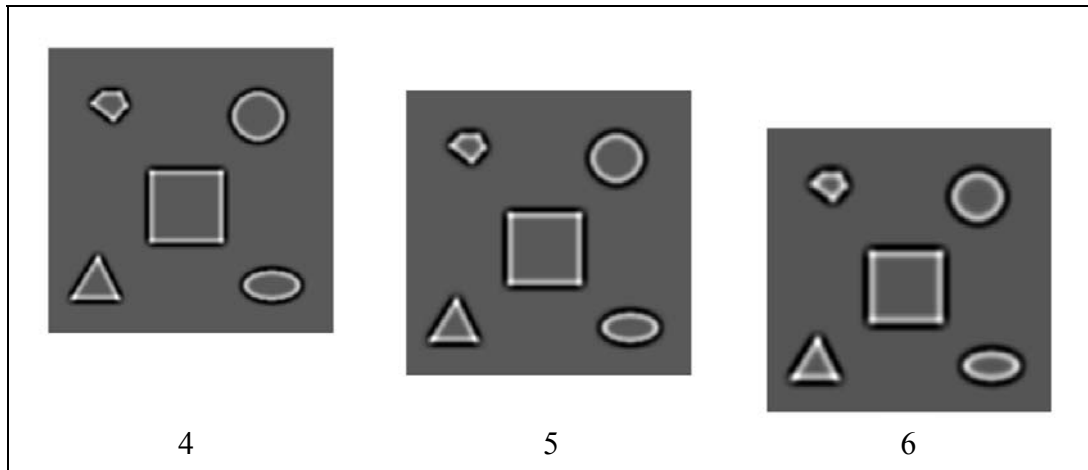


Figure 3-31 Scale-normalized Laplacian of Gaussian Filtered Images - 2

(Second Three Scale Levels, 4, 5, 6)

The local extrema of the scale-normalized Laplacian of Gaussian filtered image values are searched at the initial Harris-Laplace feature points. If the scale of the initial feature point is the local extremum for the LoG value, it is accepted as final Harris-Laplace feature point. As a result, final multiscale Harris-Laplace feature points are extracted with their spatial coordinates x , y and scale parameter. The final Harris-Laplace feature points are shown in Figure 3-32 below for the sample input image.

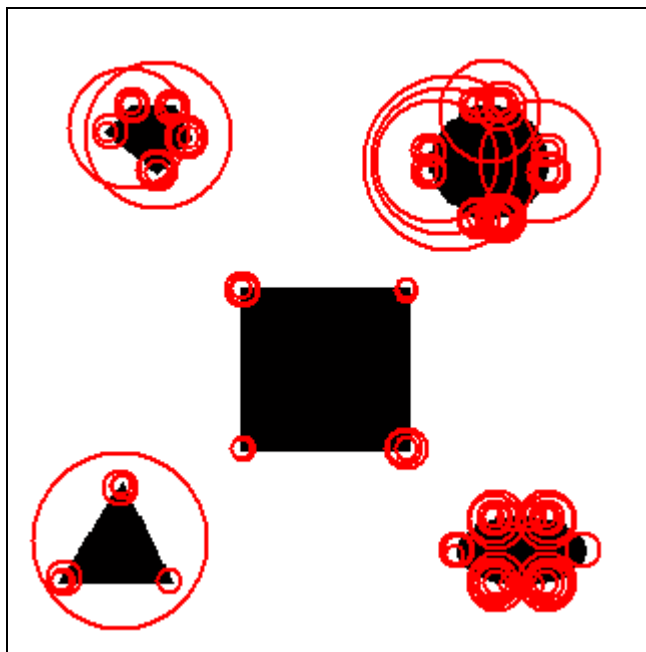


Figure 3-32 Final Harris-Laplace Feature Points

The summary of the steps used in Harris-Laplace feature detector are presented as a pseudocode in Table 3-2, Harris-Laplace algorithm shown below.

Table 3-2 Harris-Laplace Algorithm Steps

Harris-Laplace Detector Algorithm

- Differential Harris Scale Space of input image is constructed.
 - First order basic differential filter kernels are initialized in x and y direction.

$$hx = [1 \quad 0 \quad -1] \quad hy = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$$

- First order Gaussian derivative function is applied at these filter kernels.

$$\frac{\partial G}{\partial x} = -\frac{x}{\sigma^3 \sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}$$

- First order partial derivatives of $L(x,y,\sigma)$ in x and y directions are calculated at each scale.

$$L_x = \frac{\partial L}{\partial x} \quad L_y = \frac{\partial L}{\partial y}$$

- Continue smoothing incrementally with $dG(x,y,k\sigma)$ where $k=1.2$
- End, differential Harris scale space L_x and L_y is obtained.

- Harris autocorrelation matrices are calculated.

- Multiplication of the derivatives L_x^2, L_y^2 and $L_x L_y$ are calculated for each point of image.

- Multiplication of the derivatives are averaged by Gaussian filter having σ_I .

$$A11 = G(\sigma_I) * L_x^2$$

$$A12 = A21 = G(\sigma_I) * L_x L_y$$

$$A22 = G(\sigma_I) * L_y^2$$

Harris-Laplace Detector Algorithm (Continued)

- Harris Autocorrelation matrices are calculated.

$$\mu(x, \sigma_I, \sigma_D) = \sigma_D^2 G(\sigma_I) * \begin{bmatrix} L_x^2(x, \sigma_D) & L_x L_y(x, \sigma_D) \\ L_x L_y(x, \sigma_D) & L_y^2(x, \sigma_D) \end{bmatrix}$$

- End
- Cornerness measure is calculated.
 - Determinant of autocorrelation matrix components are calculated for each image coordinate.
 - Trace of autocorrelation matrix components are calculated for each image coordinate.
 - Cornerness measure is calculated for each image coordinate.

$$\text{cornerness} = \det(\mu(x, \sigma_I, \sigma_D)) - \alpha \text{trace}^2(\mu(x, \sigma_I, \sigma_D))$$

$$\alpha = 0.06$$

- End
- Finding local maxima of Cornerness measure
 - Local maxima of cornerness values are found in a region of 6 times σ_I around each point of the image.
 - Local maximum points are stored whose values are greater than some local maximum threshold.
 - End, Candidate Harris-Laplace feature points are obtained.
- Scale-normalized Laplacian of Gaussian values are calculated.
 - Scale-normalized LoG filter kernel is initialized for each scale level.

$$\sigma^2 \nabla^2 g(x, y, \sigma) = \left(\frac{x^2 + y^2 - 2\sigma^2}{\sigma^4} \right) e^{-\frac{x^2 + y^2}{2\sigma^2}}$$

Harris-Laplace Detector Algorithm (Continued)

- Scale-normalized LoG filtered images are obtained.

$$LoG(x, \sigma_I) = \sigma_I^2 (L_{xx}(x, \sigma_I) + L_{yy}(x, \sigma_I))$$

- End
- Finding local extrema of scale-normalized Laplacian of Gaussian
 - Scale-normalized LoG values belonging to candidate feature points are initialized.
 - The value of scale-normalized LoG response at the candidate feature point is checked whether or not it is a local maximum over the scales (i.e. $s-1, s+1$) or it is a local minimum over the scales
 - If scale-normalized LoG response is a local extremum, then the candidate feature point is accepted as the final Harris-Laplace feature point.
 - End, final Harris-Laplace feature points are obtained.

Multiscale Harris-Laplace detector implementation of Garcia (2008) is used for the extraction of Harris-Laplace feature points in this thesis [40].

3.4 Multiscale Hessian-Laplace Detector

Multiscale Hessian-Laplace detector consists of two detectors which are multiscale Hessian detector and Laplace detector combined together, similar to multiscale Harris-Laplace detector. Hessian detector is based on second order derivatives. Multiscale version of this detector is similar in nature to the multiscale version of Harris detector. The only difference is that Hessian matrix is used in order to extract feature points at a particular point in a particular scale. Laplace detector part is exactly the same with the one used in multiscale Harris-Laplace detector.

The diagram showing how Hessian-Laplace feature points are detected is presented in Figure 3-33 below.

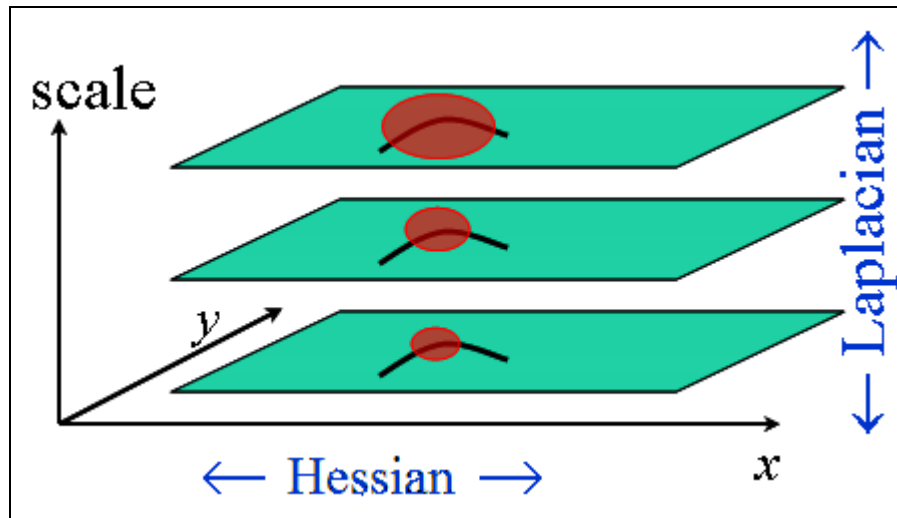


Figure 3-33 Diagram of Hessian-Laplace Detection [51].

The scale space representation of the input image is built, before multiscale Hessian detector is applied. Second order Gaussian differentials are needed for the multiscale Hessian detector. So, these second order Gaussian derivatives of the input image are constructed in scale space.

Firstly, first order Gaussian derivatives of the input image are obtained by using the method described in Harris-Laplace detector. Then, second order Gaussian derivatives L_{xx} , L_{yy} and L_{xy} are obtained by taking the convolution of the input image with the derivative filter kernel twice. Octaves and image pyramids are not used for this detector. Therefore, octave parameters and interpolation of input image are not applied in this differential Hessian scale space representation.

Second order Gaussian derivative images are shown in Figure 3-34, Figure 3-35, Figure 3-36, Figure 3-37, Figure 3-38, and Figure 3-39 below for the sample input image.

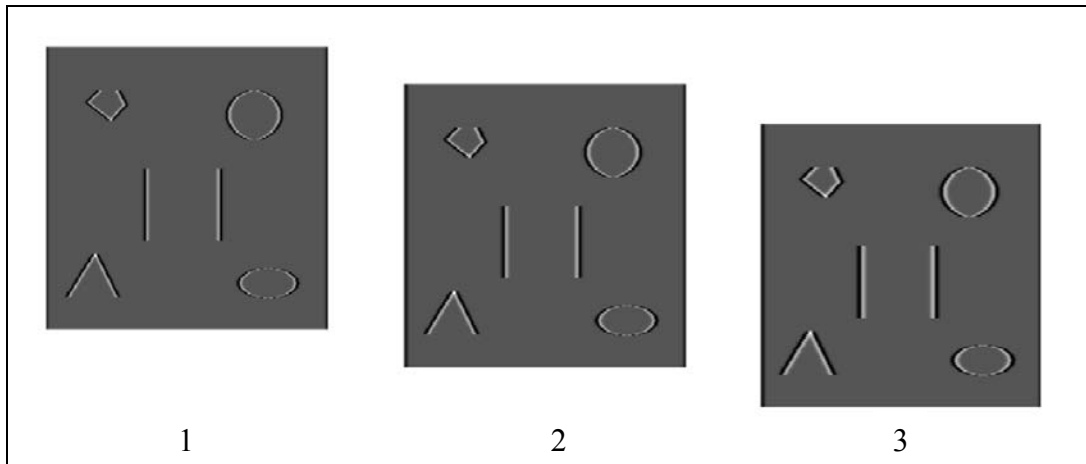


Figure 3-34 Second Order Gaussian Derivative Image-1, Lxx

(First Three Scale Levels, 1, 2, 3)

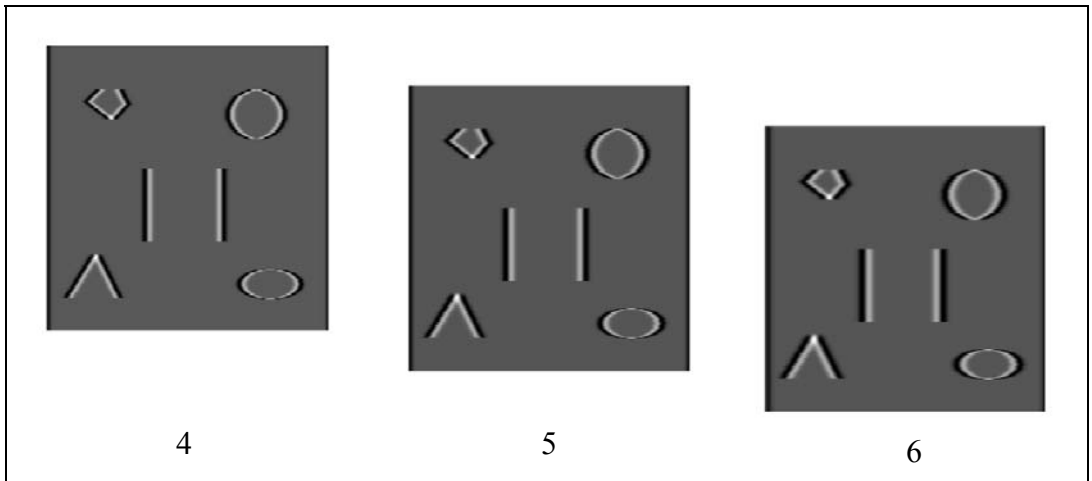


Figure 3-35 Second Order Gaussian Derivative Image-2, Lxx

(Second Three Scale Levels, 4, 5, 6)

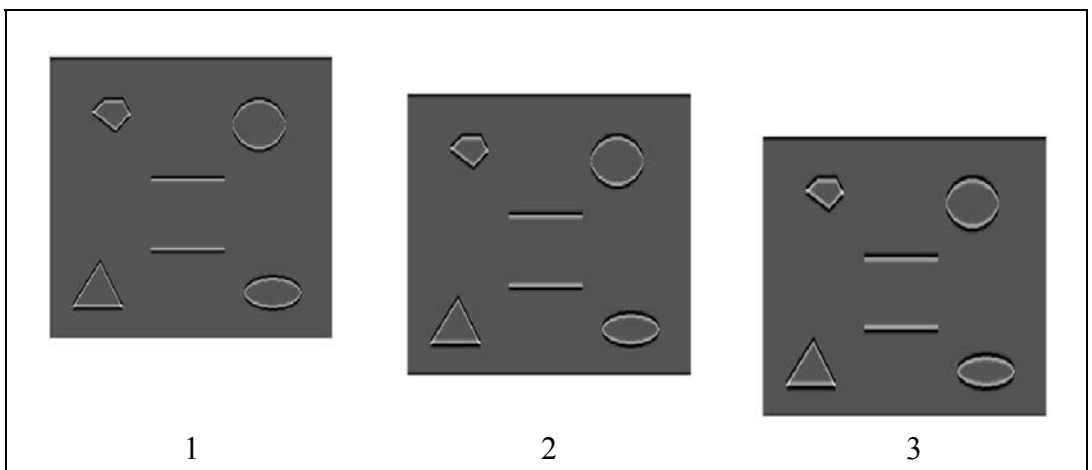


Figure 3-36 Second Order Gaussian Derivative Image-3, Lyy

(First Three Scale Levels, 1, 2, 3)

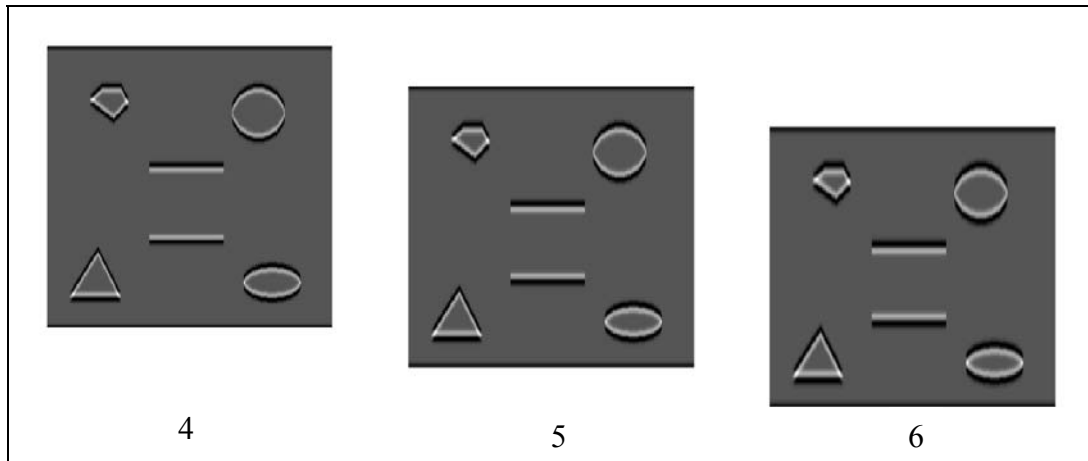


Figure 3-37 Second Order Gaussian Derivative Image-4, L_{yy}

(Second Three Scale Levels, 4, 5, 6)

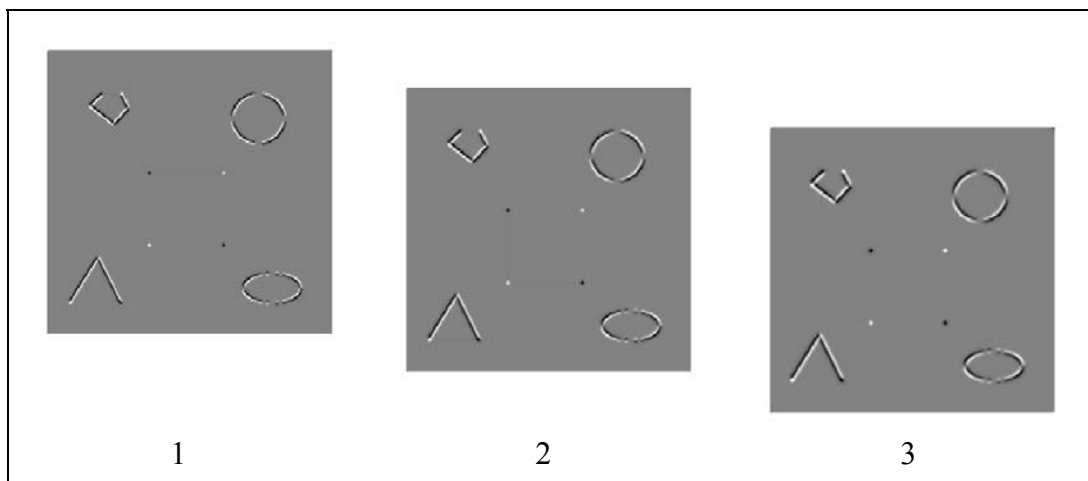


Figure 3-38 Second Order Gaussian Derivative Image-5, L_{xy}

(First Three Scale Levels, 1, 2, 3)

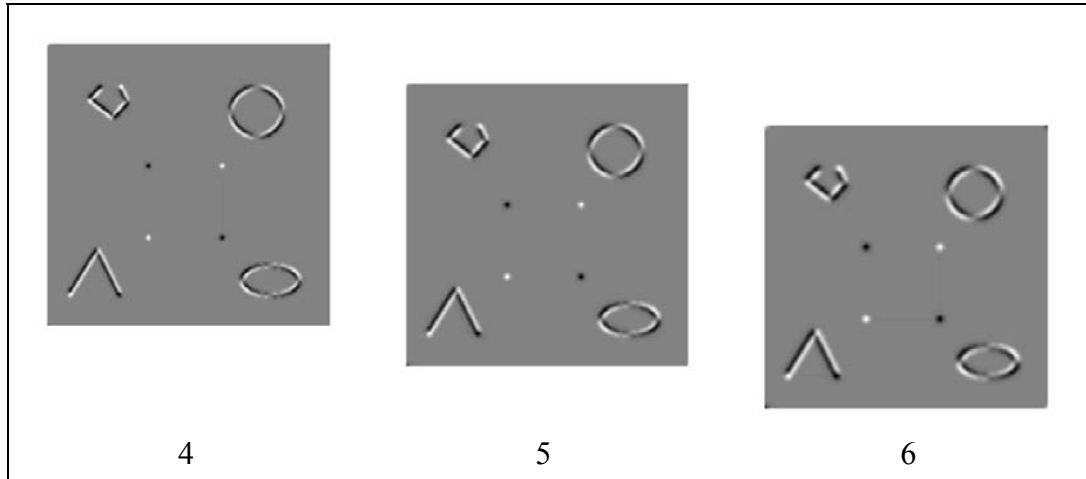


Figure 3-39 Second Order Gaussian Derivative Image-6, L_{xy}

(Second Three Scale Levels, 4, 5, 6)

The Hessian-Laplace detector is based on the Hessian matrix and second order derivatives, shown below.

$$H(x, y, \sigma) = \begin{bmatrix} L_{xx}(x, y, \sigma_D) & L_{xy}(x, y, \sigma_D) \\ L_{xy}(x, y, \sigma_D) & L_{yy}(x, y, \sigma_D) \end{bmatrix} \quad (3-33)$$

Hessian detector was developed by Beaudet in 1978 as it is previously stated in Section 3.1.1 [23]. Hessian detector is developed as a corner detector before Harris corner detector and it is utilized by most of the corner detectors. The measure that Beaudet used is the determinant of the Hessian matrix, defined as rotationally invariant measure of cornerness. The application of determinant of Hessian matrix as the local measure was not multiscale in Beaudet's original study.

Beaudet suggested his measure defined as a "rotationally invariant operator" DET shown below [44].

$$DET = I_{xx} I_{yy} - I_{xy}^2 \quad (3-34)$$

where $I(x,y)$ is the intensity value at pixels (x,y) of the input image. The feature points (corners) are detected at the place where the absolute value of DET takes its local extremum values [44]. Deriche and Giraudon suggested the use of Hessian detector in Beaudet's study at multiple scales [44].

Multiscale version of Hessian detector is invariant to scale changes. Multiscale version of this Hessian detector is used with Laplace detector, which is named as Hessian-Laplace detector. Candidate feature points are located in space at the local maxima of the determinant of multiscale Hessian matrices and final Hessian-Laplace feature points are located in scale dimension at the local maxima of the scale-normalized Laplacian of Gaussian. Hessian-Laplace detector finds blob like structures more accurately than SIFT and Harris-Laplace detectors in scale space. Laplace detector part of the Hessian-Laplace detector acts as a "matched filter" in which, it finds blobs better. Hence, shape of the scale-normalized Laplacian of Gaussian filter kernel is well suited to the blobs [24].

The scale resolution and integration scale parameters are input parameters for building Hessian scale space. Scale resolution (S) is the number of sublevels used in building the scale space. The local scale parameter σ_D is used in calculation of local derivatives of $L(x,y,\sigma)$. Local scale σ_D is taken as 0.7 times integration scale σ_I , as suggested in Mikolajczyk and Schmid [2]. An approach similar to simplified Harris-Laplace structure is used in finding the Hessian-Laplace feature points. Therefore, the input image in scale space is smoothed by a constant multiplicative factor k , namely scale factor where the smoothing kernel is equal to the derivative of $G(x,y,k\sigma)$ kernel and $k=1.2$. (Small scale factor of 1.2 is suggested for simplified Harris-Laplace in Mikolajczyk and Schmid [2].)

Hessian matrix is weighted by averaging with a Gaussian window function having integration scale σ_I . Hessian matrices components are calculated. The mathematical representation of this matrix is shown below.

$$\text{Weighted_Hessian} = G(\sigma_I) * \begin{bmatrix} L_{xx}(x, y, \sigma_D) & L_{xy}(x, y, \sigma_D) \\ L_{xy}(x, y, \sigma_D) & L_{yy}(x, y, \sigma_D) \end{bmatrix} \quad (3-35)$$

Candidate multiscale Hessian-Laplace feature points are obtained by finding the local maxima of the determinant of the Hessian matrix. Thus, the Hessian detector measure is the determinant of this matrix. The local maxima are searched in a circular region of 6 times σ_I around each point of the image. Local maxima are chosen on the basis of some threshold ($DET > \text{threshold}$) and they are stored as initial Hessian-Laplace feature points. The mathematical representation of determinant of weighted Hessian matrix is shown below.

$$DET = \sigma_I^2 (L_{xx}L_{yy} - L_{xy}^2) \quad (3-36)$$

The initial Hessian-Laplace feature points of the sample input image are shown in Figure 3-40 below. Note that both corners and blobs are detected initially.

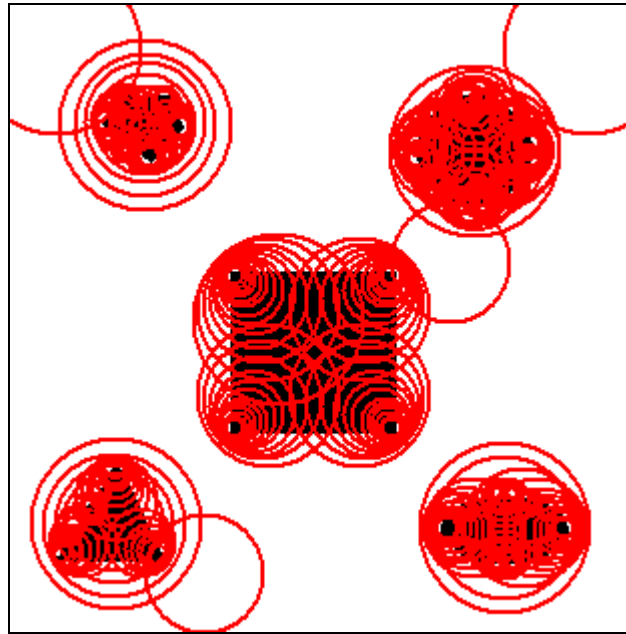


Figure 3-40 Initial Hessian-Laplace Feature Points

Laplace detector is utilized after this stage. Detailed information is given for Laplace detector in Section 3.3, Multiscale Harris-Laplace detector. The scale-normalized Laplacian of Gaussian filtered images are shown in Figure 3-30 and Figure 3-31 before, for the sample input image.

The scale-normalized Laplacian of Gaussian of the input image is obtained for each integration scale level. The LoG value of each candidate feature point is searched for its local extremum. The final feature points are obtained among the candidate points which give either local maximum or local minimum LoG values. As a result, final multiscale Hessian-Laplace feature points are extracted with their spatial coordinates x , y and scale parameter. The final Hessian-Laplace feature points are shown in Figure 3-41 below for the sample input image.

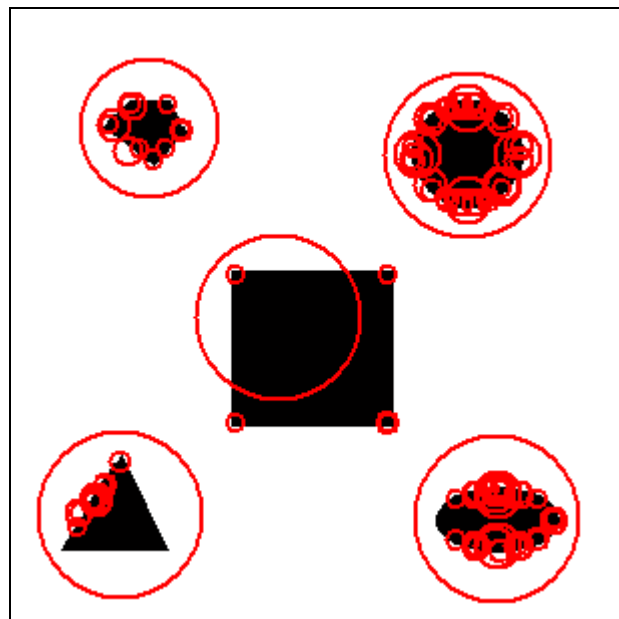


Figure 3-41 Final Hessian-Laplace Feature Points

The summary of the steps used in Hessian-Laplace feature detector are presented as a pseudocode in Table 3-3, Hessian-Laplace algorithm shown below.

Table 3-3 Hessian-Laplace Algorithm Steps

Hessian-Laplace Detector Algorithm

- Differential Hessian Scale Space of input image is constructed.
 - First order basic differential filter kernels are initialized in x and y direction.

$$h_x = [1 \quad 0 \quad -1] \quad h_y = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$$

- First order Gaussian derivative function is applied at these filter kernels.

$$\frac{\partial G}{\partial x} = -\frac{x}{\sigma^3 \sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}$$

- First order partial derivatives of $L(x,y,\sigma)$ in x and y directions are calculated at each scale.

$$L_x = \frac{\partial L}{\partial x} \quad L_y = \frac{\partial L}{\partial y}$$

- Second order Gaussian derivatives L_{xx} , L_{yy} and L_{xy} by taking the convolution of the input image with the derivative filter kernels twice.
- Continue smoothing incrementally with $dG(x,y,k\sigma)$ where $k=1.2$
- End, differential Hessian scale space L_{xx} , L_{yy} , and L_{xy} is obtained.

Hessian-Laplace Detector Algorithm (Continued)

- Weighted Hessian matrices are calculated.
 - Second order partial Gaussian derivatives L_{xx} , L_{yy} and L_{xy} are initialized.
 - Second order Gaussian derivatives are averaged by Gaussian filter having σ_I .

$$H11 = G(\sigma_I) * L_{xx}$$

$$H12 = H21 = G(\sigma_I) * L_{xy}$$

$$H22 = G(\sigma_I) * L_{yy}$$

- Hessian matrices are obtained for each image coordinate.

$$\text{Weighted_Hessian} = G(\sigma_I) * \begin{bmatrix} L_{xx}(x, y, \sigma_D) & L_{xy}(x, y, \sigma_D) \\ L_{xy}(x, y, \sigma_D) & L_{yy}(x, y, \sigma_D) \end{bmatrix}$$

- End
- Calculation of determinant of Hessian matrices
 - Components of weighted Hessian matrices are initialized for all image coordinates at all scale levels.
 - Determinant of Hessian matrices are calculated.

$$DET = \sigma_I^2 (L_{xx} L_{yy} - L_{xy}^2)$$

- End
- Finding local maxima of determinant measure
 - Local maxima of determinant values are found in a region of 6 times σ_I around each point of the image.
 - Local maximum points are stored whose values are greater than some local maximum threshold.
 - Candidate Hessian-Laplace feature points are obtained.
 - End

Hessian-Laplace Detector Algorithm (Continued)

- Scale-normalized Laplacian of Gaussian values are calculated.
 - Scale-normalized LoG filter kernel is initialized for each scale level.

$$\sigma^2 \nabla^2 g(x, y, \sigma) = \left(\frac{x^2 + y^2 - 2\sigma^2}{\sigma^4} \right) e^{-\frac{x^2 + y^2}{2\sigma^2}}$$

- Scale-normalized LoG filtered images are obtained.

$$LoG(x, \sigma_I) = \sigma_I^2 (L_{xx}(x, \sigma_I) + L_{yy}(x, \sigma_I))$$

- End
- Finding local extrema of scale-normalized Laplacian of Gaussian
 - Scale-normalized LoG values belonging to candidate feature points are initialized.
 - The value of scale-normalized LoG response at the candidate feature point is checked whether or not it is a local maximum over the scales (i.e. s-1, s+1) or it is a local minimum over the scales
 - If scale-normalized LoG response is a local extremum, then the candidate feature point is accepted as the final Hessian-Laplace feature point.
 - End, final Hessian-Laplace feature points are obtained.

Multiscale Harris-Laplace detector implementation of Garcia is modified by changing Harris detector part with Hessian matrix to make it multiscale Hessian detector for the simulations of this thesis [40].

3.5 SURF Detector

Speeded Up Robust Features (SURF) detector is a fast and efficient feature detector that has been recently developed by Bay et.al [3].

SURF feature detector uses a Hessian-based detector, which takes the determinant of Hessian matrix, as the local extremum measure for choosing interest points and scale. The detector of SURF is actually an approximation to the Hessian matrix, like Difference of Gaussian (DoG) is an approximation for scale-normalized Laplacian of Gaussian (LoG) detector. The convolution process used in the detection of feature points includes integral images and box-type Gaussian filter kernels, which makes the detection process very fast. Authors of the study also call it “Fast Hessian detector” for this reason [3].

Before obtaining SURF scale space and detection of feature points, integral image of the input image is obtained. Integral images, which are developed by Viola, and Jones [25], represent the sum of all pixels of the input image for its particular rectangular region. This region is defined by the coordinates x and y , and it is the rectangular area formed by this point and origin. The integral image at location (x,y) , is the sum of the pixel values above and to the left of (x,y) , inclusive, shown in Figure 3-42 below.

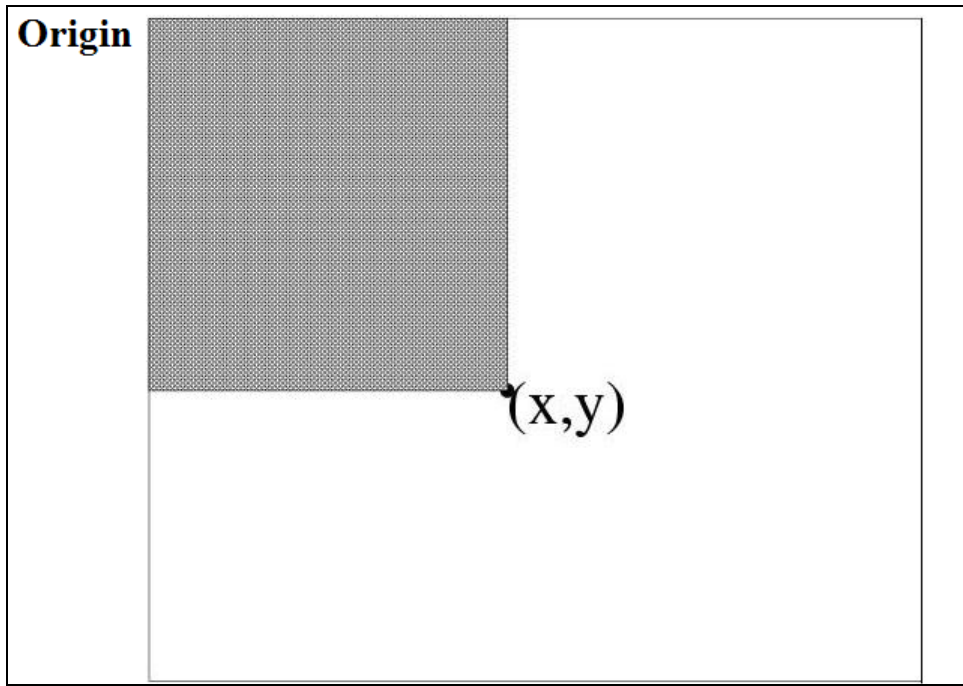


Figure 3-42 Rectangular Region for Integral Image

Thus, the corresponding integral image entry value is formulated as :

$$I_{\Sigma}(x, y) = \sum_{i=0}^{i \leq x} \sum_{j=0}^{j \leq y} I(i, j) \quad (3-37)$$

The integral image value $s(x,y)$ at pixel (x,y) is calculated in a single pass by using the following recursive formula.

$$s(x, y) = I(x, y) + s(x, y - 1) + s(x - 1, y) - s(x - 1, y - 1) \quad (3-38)$$

The values of integral image are the cumulative sums for all the pixels. Therefore, this $s(x,y)$ is an intermediate representation. While calculating the actual partial Gaussian second order derivatives of the input image, the integral image values corresponding to the inner rectangular regions of corresponding filter sizes should be calculated. In order to calculate the value of the integral image corresponding to

an arbitrary rectangular region inside the image, the following mathematical relation is used.

$$s(x, y, h, w) = s(x, y) + s(x + w, y + h) - s(x, y + h) - s(x + w, y) \quad (3-39)$$

where $s(x,y,h,w)$ is the sum of all pixels located in a rectangular region S to the left of pixel (x,y) with height h and width w . The calculation of integral image value (rectangle sum) for an arbitrary rectangular region S inside the image is shown in Figure 3-43 below.

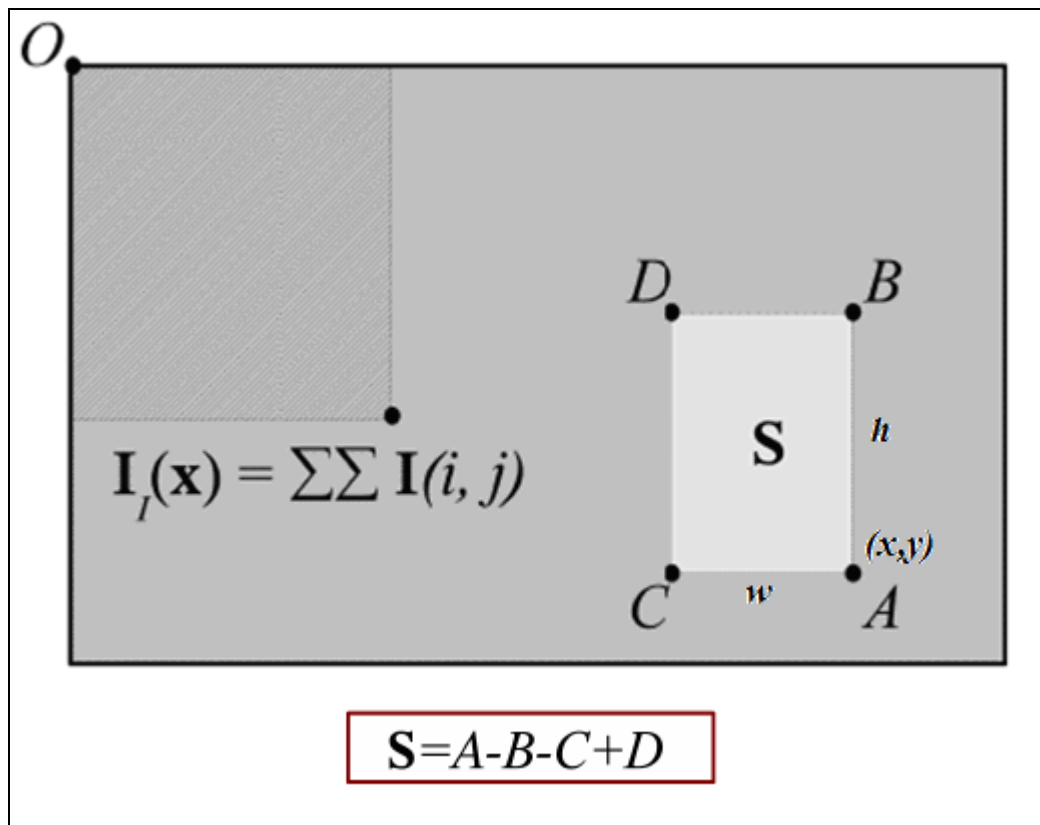


Figure 3-43 Calculation of an Arbitrary Rectangle Sum [50].

SURF feature detector is based on Hessian matrix. Therefore second order partial Gaussian differentials are constructed in scale space. Hessian matrix is described in Hessian-Laplace feature detector section.

Box filtering is used in order to obtain second order partial Gaussian derivatives of the input image. Box filtering is basically an average-of-surrounding-pixel kind of image filtering. Advantage of box filtering is that it can give approximate filtering results for a general use such that one can write general image filters like sharpen, emboss, smooth filters, etc.

Basic second order derivative filter kernels are used in image processing for taking differences between pixels. Three basic second order derivative filters in x, y and x-y directions are shown below.

$$h_{2x} = \begin{bmatrix} 1 & -2 & 1 \end{bmatrix} \quad (3-40)$$

$$h_{2y} = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} \quad (3-41)$$

$$h_{xy} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \quad (3-42)$$

Combining the use of both box filtering and basic second order derivative filters, the following box filters are used in order to obtain approximate second order partial Gaussian derivatives of input image.

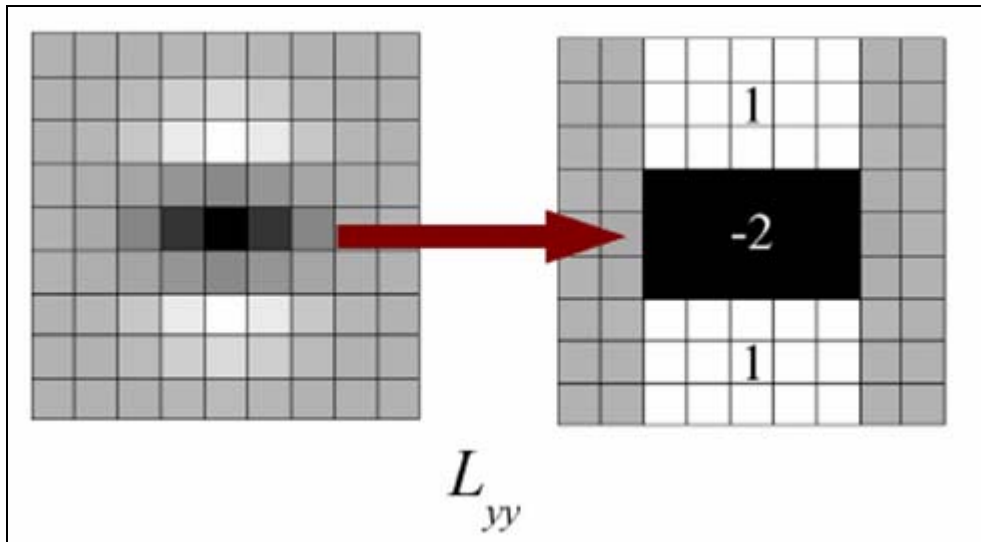


Figure 3-44 Box Filter for Second Order Partial Gaussian Derivative in y-direction [50].

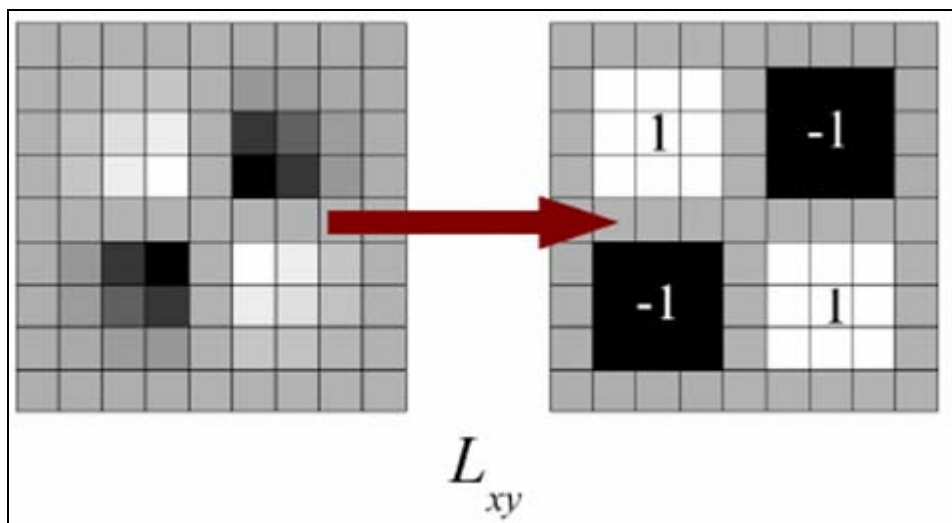


Figure 3-45 Box Filter for Second Order Partial Gaussian Derivative in xy-direction [50].

The filters on the left hand side of these images are the traditional discretized and cropped versions of derivative filter kernels, which are used in the construction of scale spaces of SIFT, Harris-Laplace and Hessian-Laplace detectors. Second order partial Gaussian derivatives of the input image are approximated by the use of box filters and integral images. Filters, which are shown on the right hand side of these figures, are the initial filter kernels of size 9x9. The filter kernel size of 9x9 boxes correspond to the actual second order Gaussian derivative filter kernels of standard deviation of 1.2, $\sigma=1.2$. The second order partial Gaussian derivative filter in x direction is actually the transpose of the box filter in y direction, shown in Figure 3-44.

Scale space of SURF feature detector is implemented by using image pyramids. Scale space is constructed by up-scaling the filter kernel sizes, rather than subsampling the input image like in SIFT detector. The initial filter kernel size is 9x9 and the image is filtered by filter kernels of 15x15, 21x21, 27x27 and so on for the next scale levels of the first octave. The filter size increase is doubled for the next octaves. It becomes 12 for the next octave, beginning from 39x39 filter kernel. The scale space structure of the SURF detector is shown in Figure 3-46 below.

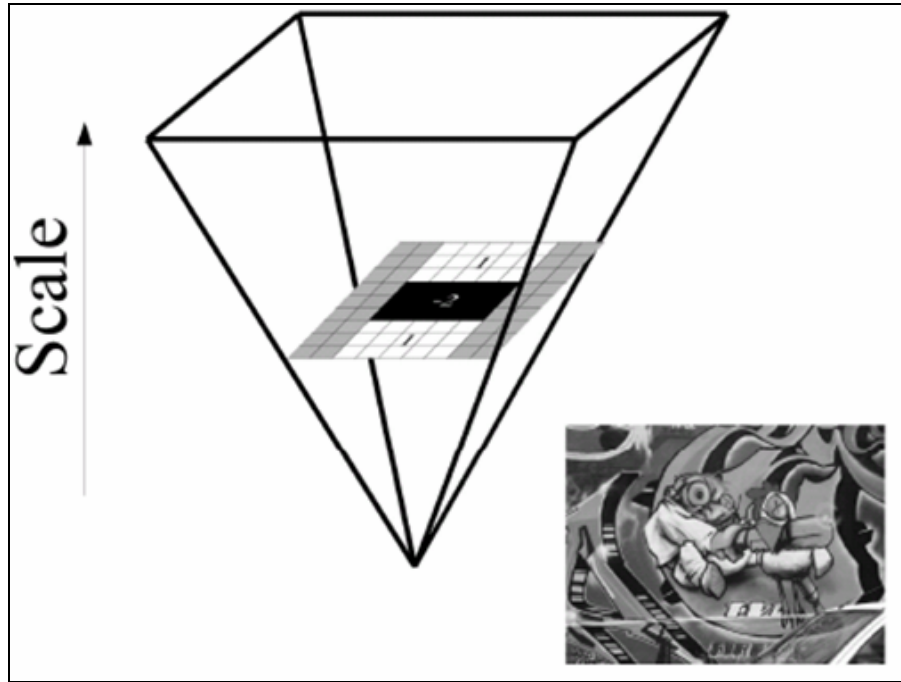


Figure 3-46 Scale Space Structure of SURF Feature Detector [50].

The local measure used in the selection of SURF feature points is the determinant of the Hessian matrix. The second order Gaussian differentials obtained are the approximations to L_{xx} , L_{yy} , and L_{xy} . The approximated Hessian matrix obtained is shown below.

$$H(x, y, \sigma) = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix} \quad (3-43)$$

The local maxima of the determinant of this approximated Hessian matrix is used in order to select the space coordinates x , y and the scale of the feature point. Frobenius norm of the second order Gaussian partial differentials are applied in order to balance the weights of the derivatives in the value of determinant. Filter responses are normalized for a constant Frobenius norm of derivatives.

Frobenius norm is very similar to Euclidean norm. Frobenius norm of a matrix is defined mathematically as it is shown below.

$$|A|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2} \quad (3-44)$$

Weights of the derivatives are balanced by Frobenius norm of the second order Gaussian partial derivatives, as it is shown below.

$$\frac{|L_{xy}(1.2)|_F |D_{xx}(9)|_F}{|L_{xx}(1.2)|_F |D_{xy}(9)|_F} \cong 0.9 \quad (3-45)$$

The expression for determinant of Hessian matrix after Frobenius norm is applied is shown below.

$$\det(H(x, y, \sigma)) = D_{xx}D_{yy} - (0.9D_{xy})^2 \quad (3-46)$$

Initial feature points are obtained by finding the local maxima of the determinant measure above. However, the initial feature points do not represent the true location of the local maximum points corresponding to the local measure used. Therefore, the feature points need to be localized in space and scale coordinates to obtain the true location of the local maxima. The same method of SIFT detector subpixel localization is used to localize the feature points [48]. “Non-maximum suppression” is applied for subpixel localization in 3x3x3 neighborhood for this purpose.

As a result of the localization stage, final SURF feature points are extracted with their spatial coordinates x, y and scale parameter. These final SURF feature points are shown in Figure 3-47 below.

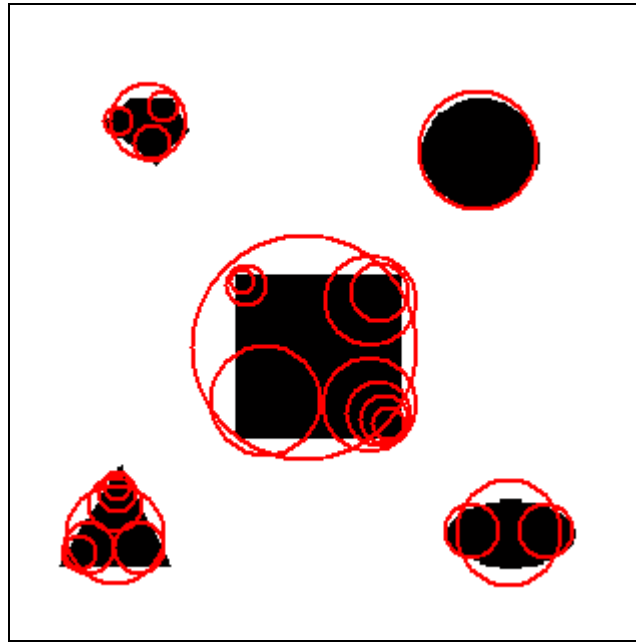


Figure 3-47 Final SURF Feature Points

The summary of the steps used in SURF feature detector are presented as a pseudocode in Table 3-4, SURF algorithm shown below.

Table 3-4 SURF Algorithm Steps

SURF Detector Algorithm

- Finding integral image of the input image
 - Integral image value $s(x,y)$ at pixel (x,y) is calculated in a single pass.

$$s(x, y) = I(x, y) + s(x, y - 1) + s(x - 1, y) - s(x - 1, y - 1)$$
 - The integral image values are stored as a look-up table for later use in construction of differential scale space.
 - End
- Differential SURF Scale Space of input image is constructed.
 - Integral image look-up table is initialized.
 - Box filter of size 9x9 is initialized based on the following second order differential filter kernels.

$$h_{2x} = [1 \quad -2 \quad 1] \quad h_{2y} = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$$

$$h_{xy} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

- Integral image values corresponding to the inner rectangular regions of size 9x9 is calculated.

$$s(x, y, h, w) = s(x, y) + s(x + w, y + h) - s(x, y + h) - s(x + w, y)$$
 height (h) and width (w) of the rectangular regions are 9.
- Second order partial Gaussian derivatives of input image D_{xx} , D_{yy} , and D_{xy} are calculated by using box filters.
- Box filters of sizes 15x15, 21x21 and 27x27 with an increase of 6 units are initialized for next scale levels of the first octave.

SURF Detector Algorithm (Continued)

- Second order partial Gaussian derivatives of input image D_{xx} , D_{yy} , and D_{xy} are calculated for next scales.
- Box filter size is doubled.
- Box filters of sizes 39x39, 51x51, 63x63, and so on with an increase of 12 units are initialized for next scale levels of the second octave.
- Continue smoothing until number of octaves corresponding to the size of box filters is less than the size of the input image.
- End, differential SURF scale space D_{xx} , D_{yy} , and D_{xy} is obtained.
- Calculation of determinant of Hessian matrices
 - Components of approximate Hessian matrices are initialized for all image coordinates at all scale levels.
 - Weights of the derivatives are balanced by Frobenius norm of the second order Gaussian partial derivatives.

$$\frac{|L_{xy}(1.2)|_F |D_{xx}(9)|_F}{|L_{xx}(1.2)|_F |D_{xy}(9)|_F} \cong 0.9$$

- Determinant of Hessian matrices are calculated.

$$\det(H(x, y, \sigma)) = D_{xx}D_{yy} - (0.9D_{xy})^2$$

- End

SURF Detector Algorithm (Continued)

- Finding local maxima of determinant measure
 - Local maxima of determinant values are found in a region around each point of the image.
 - Local maximum points are stored whose values are greater than some local maximum threshold.
 - Candidate SURF feature points are obtained.
 - End
- Subpixel localization of initial SURF feature points
 - Initial feature points are initialized.
 - Non-maximum suppression is applied in a 3x3x3 neighborhood of each initial feature point.
 - Initial feature points are localized in space and scale.
 - The value of the maximum of the determinant is interpolated in space and scale.
 - End, final SURF feature points are obtained.

SURF detector implementation of Bay et.al is used for the extraction of SURF feature points in this thesis [41].

3.6 Orientation of Interest Points

Orientation of interest points is the histogram of gradient directions calculated in the neighborhood of the corresponding interest point. It is used in the calculation of descriptor vectors of each interest point to obtain rotation invariance. Orientation of feature points are calculated after the extraction of them using feature detectors.

Orientation of SIFT, Harris-Laplace and Hessian-Laplace feature points is calculated on the basis of histogram of gradient directions, which is used in SIFT descriptors. However, orientation of SURF features is different from SIFT format orientations, since SURF descriptors have different form. The reason is that SIFT descriptors are used with SIFT, Harris-Laplace and Hessian-Laplace features, whereas SURF descriptor is used with SURF features in this study. SURF feature orientations are calculated on the basis of histogram of the “Haar-wavelet” responses in x- and y-directions in the neighborhood of the corresponding interest point [3].

After the refinement or localization stages of initial SIFT, Harris-Laplace and Hessian-Laplace feature points, a set of good features are obtained for each detector. The effects of rotation are eliminated by finding the dominant orientation component of each feature point at its corresponding scale. Gaussian scale space of the input image is used for this purpose. Each feature point has a scale component at this stage and the correct image of the Gaussian scale space at the right scale or at the closest scale of the feature point is analyzed for finding the orientation.

Correct image of the Gaussian scale space belonging to correct scale is chosen and orientation calculations are performed with this image. The chosen image $L(x,y,\sigma)$ is indicated with the equation given below.

$$L(x, y) = G(x, y, \sigma) * I(x, y) \quad (3-47)$$

where I is the input image and $G(x,y,\sigma)$ is the Gaussian filter kernel.

Gradient magnitude $|G(x, y)|$ and orientation $\angle G(x, y)$ of each feature point is calculated at spatial coordinates x, y and at the closest scale σ of that feature point by using relations of pixel differences shown below:

$$|G(x, y)| = \sqrt{(L(x + 1, y) - L(x - 1, y))^2 + (L(x, y + 1) - L(x, y - 1))^2} \quad (3-48)$$

$$\angle G(x, y) = \tan^{-1}((L(x, y + 1) - L(x, y - 1)) / (L(x + 1, y) - L(x - 1, y))) \quad (3-49)$$

Orientation histogram is formed from the gradient orientations of sample points $L(x,y)$ within a region around the feature point. This region is a Gaussian weighted circular window centered on the feature point with 1.5 times that of the scale of the feature point, that is $\sigma_w = 1.5\sigma$.

The orientation histogram has 36 bins covering 360 degree range of orientations. Each sample added to the histogram is weighted by its gradient magnitude and by a Gaussian weighted circular window with 1.5 times corresponding scale. After this step, histogram is smoothed by a moving average filter. The typical diagram of orientation histogram is shown in Figure 3-48 below.

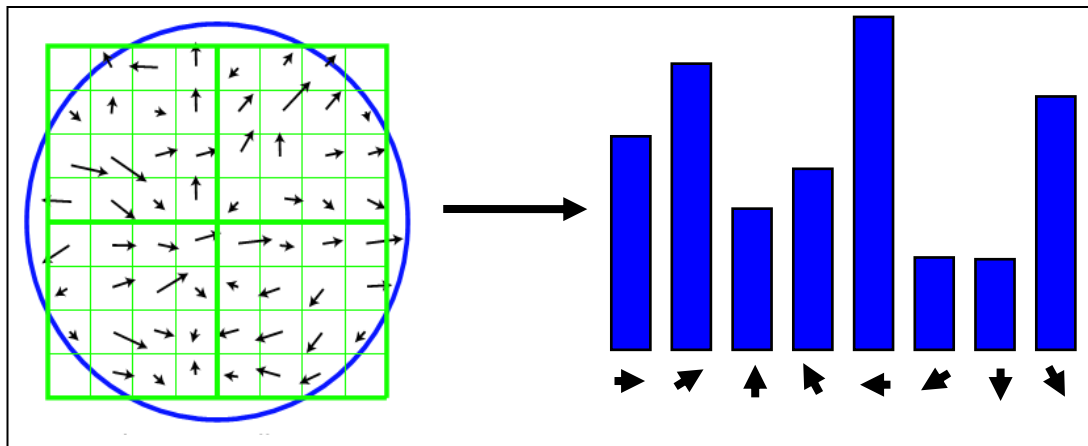


Figure 3-48 Typical Orientation Histogram Diagram [49].

Strongest peaks in the orientation histogram are found. The highest peak in the histogram is detected and any other orientation peak within 80 percent of the highest peak are also obtained [1]. Therefore, feature points are obtained with multiple orientations at the same location. Thus, for each location and scale,

multiple SIFT features can be generated. Finally, 3 closest histogram values to each peak are used to interpolate (fit to a parabola) a better accurate peak [1].

As a result of this stage, SIFT, Harris-Laplace, and Hessian-Laplace features have four components of x , y spatial coordinates, σ of the corresponding scale, and dominant orientation of the feature point. The SIFT, Harris-Laplace, and Hessian-Laplace feature points of the sample input image are shown with their dominant orientations in Figure 3-49, Figure 3-50, and Figure 3-51 below, respectively.

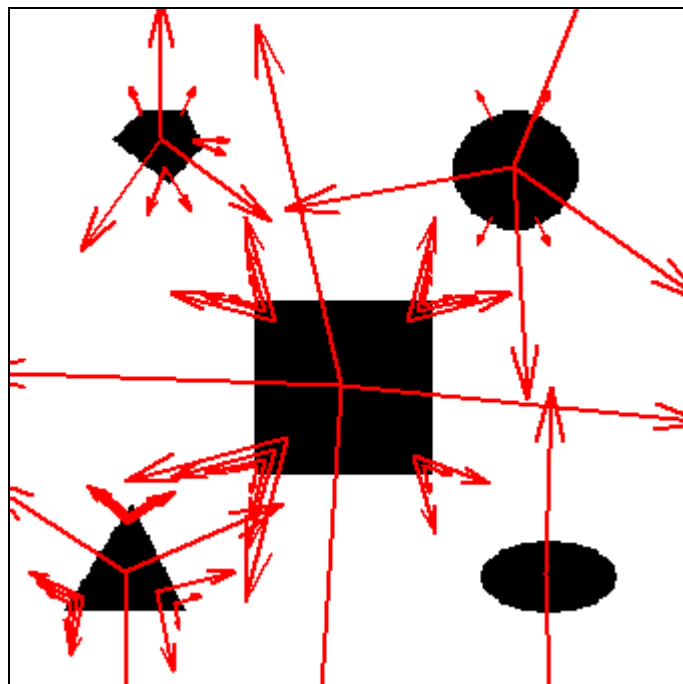


Figure 3-49 SIFT Features With Orientations

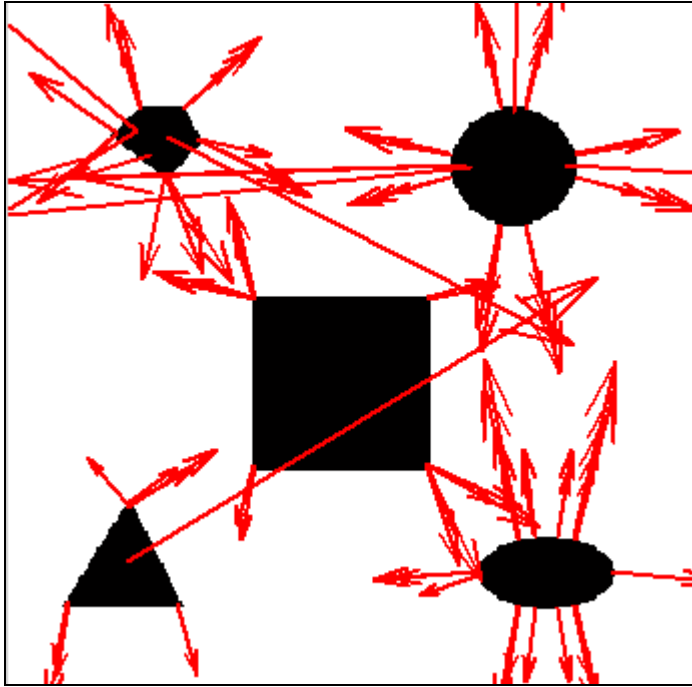


Figure 3-50 Harris-Laplace Feature Points with Orientations

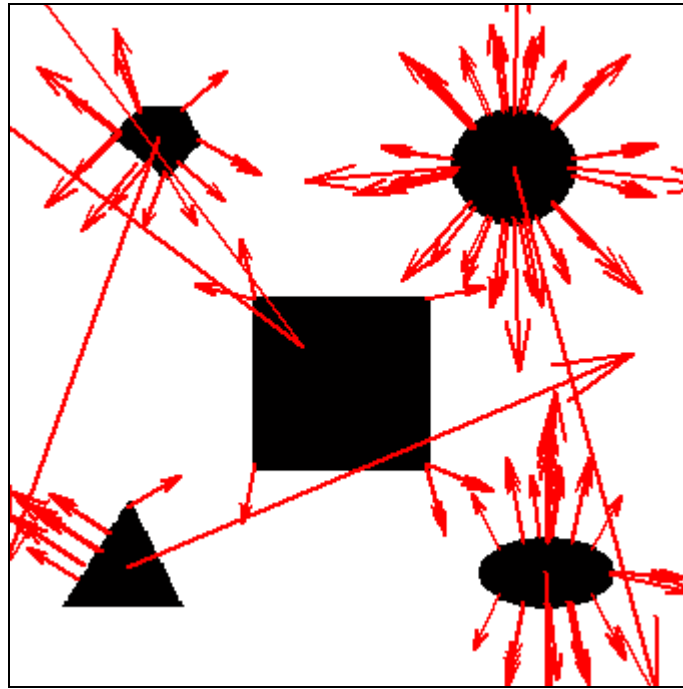


Figure 3-51 Hessian-Laplace Feature Points with Orientations

SURF based orientation of feature points is calculated on the basis of the “Haar wavelet” responses around the interest point in x- and y-directions [3]. The calculation of Haar wavelet responses is performed at the corresponding scale of the interest point.

Haar wavelet is the first known wavelet with the simplest form. Haar used these functions to give an example of a countable orthonormal system for the space of square-integrable functions on the real line. The disadvantage of the Haar wavelet is that it is not continuous and therefore not differentiable [45].

Haar wavelet’s mother wavelet function $\psi(t)$ is shown below.

$$\psi(t) = \begin{cases} 1 & 0 \leq t < 1/2, \\ -1 & 1/2 \leq t < 1, \\ 0 & \text{otherwise} \end{cases} \quad (3-50)$$

Scaling function $\phi(t)$ of Haar wavelet is also shown below.

$$\phi(t) = \begin{cases} 1 & 0 \leq t < 1, \\ 0 & \text{otherwise} \end{cases} \quad (3-51)$$

Haar wavelet types used for SURF orientation calculations are shown in Figure 3-52 below. Side lengths of the wavelets are four times the scale of the feature point (4s).



Figure 3-52 Haar Wavelet Types [50].

(Left : x response, Right : y response)

Haar wavelet responses of each feature point are calculated in x- and y-directions in a circular neighborhood of radius, six times the scale of the feature point (6s) around that interest point. Then, these responses are weighted with a Gaussian of standard deviation 2.5s centered at the feature point. The responses calculated are separate for horizontal and vertical directions. Dominant orientations are obtained by calculating the sum of all responses within a sliding orientation interval of an angle covering $\pi/3$. Horizontal and vertical responses are summed and the largest norm vector in the corresponding window gives its orientation angle as the orientation of the interest point [3].

3.7 Local Region Descriptors

Local region descriptors or feature descriptors are vectors that describe and represent information about a local area of an image [42]. This area is actually a circular one around the interest points determined by feature detectors and the diameter of which depends on the scale of the feature point. Feature descriptors are used effectively in image matching and retrieval by finding correspondences between the points in reference and test images [42].

Mikolajczyk, and Schmid, showed that SIFT type descriptor gave best performance results when it is used with Harris and Hessian type detectors with both multiscale Laplace and affine invariant version [24]. Moreels, and Perona, showed similar results that support the successful performance of SIFT descriptor when it is used with several feature detectors including Harris-Laplace and Hessian-Laplace detectors [18]. The successful performance of SIFT descriptor is also reported in studies of Zhang et.al [19], Brown, and Lowe [15].

The feature points consist of spatial coordinates x and y , corresponding scale and orientation up to this stage. The SIFT descriptor vectors for each feature point found by SIFT, Harris-Laplace and Hessian-Laplace detectors are calculated at this step. SURF type descriptors are also obtained for each feature point found by SURF detector. These descriptor vectors are important in matching algorithm between two distinctive feature points of two reference and training images.

The SIFT descriptor is the statistics of the feature points based on the orientations around their neighborhoods. They are calculated for a region around a keypoint, like in orientation, as opposed to directly for the keypoint. SIFT descriptor is a three dimensional histogram of the position and orientation of the gradient.

The input for SIFT descriptor calculation is the Gaussian scale space of the input image. SIFT descriptor is obtained for each octave separately as the descriptor function operates on single octave of Gaussian scale space for SIFT feature points.

SIFT descriptors are obtained for the Gaussian scale space of Harris-Laplace and Hessian-Laplace detectors including only one multiscale structure, not several octaves.

Each keypoint is centered on the region under constraint. Then, this region is divided into 4x4 bins in each direction x and y, resulting in a total of 16 subregions. Gradient statistics is collected in each of these subregions as in the calculation of dominant orientations. Magnitudes and orientations are calculated for points in region around keypoint using its nearest scale. The radius of this circular region used in the gradient statistics calculation is shown below.

$$radius = 4mag_factor.\sigma / 2 \quad (3-52)$$

where σ is the closest scale of the keypoint and “mag_factor” is the magnification factor parameter.

Fractional displacements from the center are calculated. (i.e. Each center of the layout is x, y spatial coordinate of the feature points.) The displacements of each bin are calculated, normalized with respect to the feature point orientation and extension. This is performed in order to ensure orientation invariance. In other words, the gradient orientations and coordinates of descriptor are rotated relative to orientation of keypoint. The diagram showing these operations is presented in Figure 3-53 below.

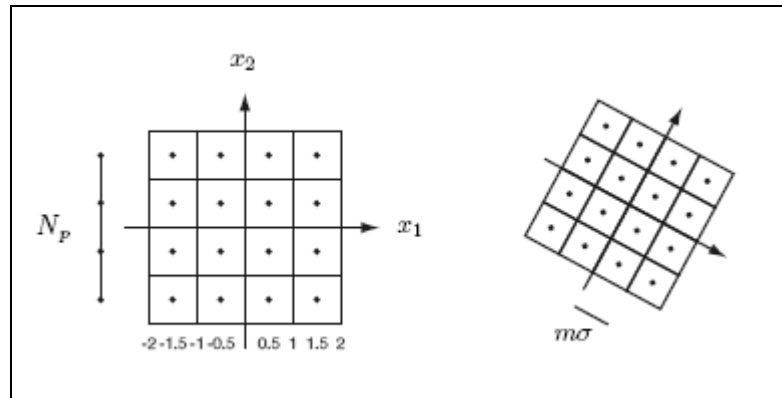


Figure 3-53 Fractional Displacements and Orientation Invariance [47].

Magnitude of each point is weighted with σ of one half the width of the descriptor window ($NBP/2$). There are two reasons for this operation, which are firstly it stops sudden changes in descriptor due to small changes in the position of the window and lastly, it gives less weight to gradients far from keypoint [1].

The Gaussian weighted histogram is distributed into 8 adjacent bins by summing each of the bins proportionally to its distance from the respective center. At the end, the histogram is normalized to unit length. This last operation is performed in order to ensure invariance to illumination change [1]. The generic SIFT descriptor obtained at the end of these operations is shown in Figure 3-54 below.

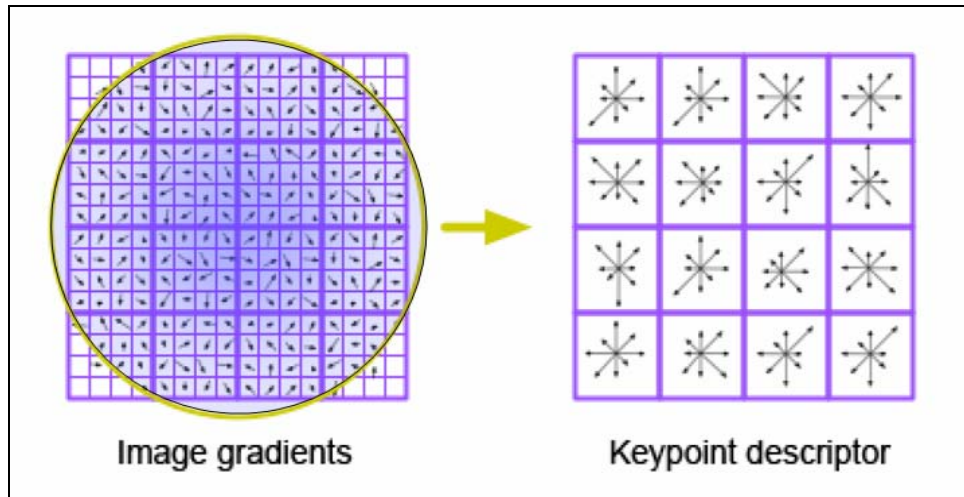


Figure 3-54 4x4x8 Typical SIFT Descriptor [49].

As a result, the descriptors are calculated for each feature point within its 4 units (4 bins) of neighborhoods in each spatial direction x and y, resulting in 4x4 bins of region. The orientation histogram is obtained in each of 16 unit regions (4x4 bins) with 8 angular bins or intervals. The SIFT descriptor vector contains 128 components containing the orientation histogram information in this way.

SURF descriptor is based on histogram of Haar wavelet responses in horizontal and vertical directions around the local area of interest points found by SURF feature detector [3].

SURF descriptors of SURF feature points are obtained in a square region around the feature point centered on it. This region is subdivided into 4x4 square subregions with 5x5 regularly spaced sample points inside. The region is oriented towards the orientation angle of the feature point. The Haar wavelet responses are calculated in x and y directions for each subregion. These responses are then weighted by a Gaussian kernel (of standard deviation $\sigma=3.3s$, where s is the scale of feature point) centered at the feature point in order to increase the robustness towards geometric deformations and localization errors. The wavelet responses d_x and d_y are summed

up separately for each subregion. The sum of absolute value of these responses is also extracted for each subregion in order to bring in information about the polarity of the intensity changes. Therefore SURF descriptor vectors of size $4 \times 4 \times 4$, 64 are obtained for each corresponding feature point. SURF descriptor values are then normalized so that invariance to contrast is obtained [3].

The diagram showing how SURF descriptor is obtained is given in Figure 3-55 below.

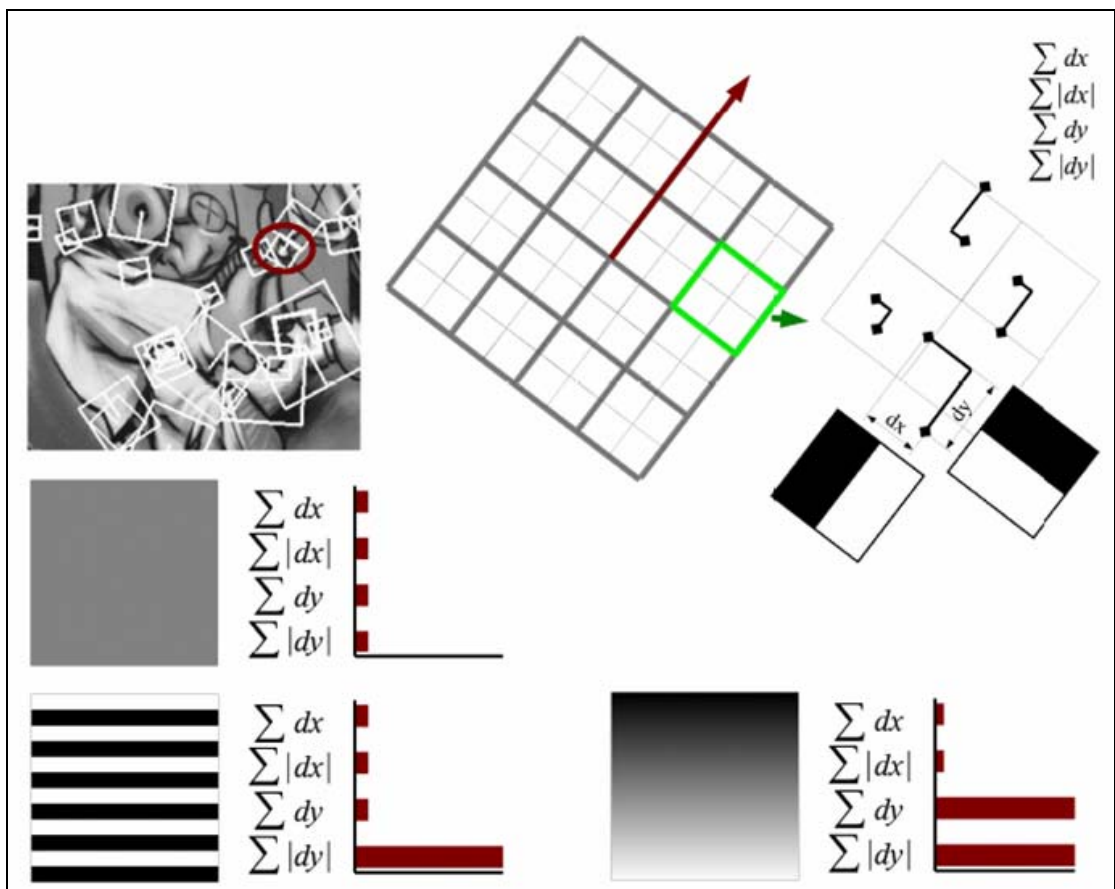


Figure 3-55 Diagram of SURF Descriptor [50].

Bay et.al, 2006 [3] evaluated their SURF descriptor performance in their study [3]. They performed experiments similar to Mikolajczyk, and Schmid's study in [24]. They showed that SURF descriptor used with SURF feature points performed the best, when it is compared with other descriptors, including SIFT descriptor [3].

Matching algorithm and calculation method for matching performance is discussed in the next section.

3.8 Matching Accuracy of Feature Points

Object recognition requires accurate matching of feature points between two images in applications. Matching algorithm is important for this purpose. This algorithm is explained first in this section. Then, definition of matching accuracy is given.

Matching algorithm of Lowe, [1] is used in this study. It is based on the nearest neighbor feature point between the one in the reference image and in the training (test) image. This nearest neighbor feature point is defined as the point with minimum Euclidean distance between its descriptor vector in the training image and the compared feature point descriptor vector in the reference image. The closest first neighbor feature point does not give accurate match always as explained in Lowe, [1]. Lowe used second closest neighbor feature point and decided that the best criteria for the best match to be that the first closest neighbor feature point and the second closest neighbor feature point should be away from each other. As a result, the first closest neighbor feature point can be stated as the best unique match to the feature point analyzed in the reference image.

The feature points are extracted by SIFT, multiscale Harris-Laplace, multiscale Hessian-Laplace, and SURF feature detectors for each reference and training images in visual and infrared-bands. Then, descriptors are found for each of the feature point in both types of reference and training images. After this calculation step, matched feature points are found on the basis of nearest neighborhood criteria of Lowe between the reference and the training image features.

The matching algorithm implementation of Vedaldi is used for this purpose [39].

The matching performance of the feature detectors under the scale and rotation changes performed in the experiments is evaluated by using the matching accuracy criterion, which is obtained from Precision and Recall rates of the matching.

Mikolajczyk, and Schmid, evaluated the matching performance of different type descriptors along with different feature detector combinations by using Recall and “1-precision” rate criteria. Recall rate depends on the number of correspondences and number of correct matches. Mikolajczyk, and Schmid obtained number of correspondences by finding homographies between the reference image and the transformed image. They took into account a certain overlap error between the local regions defined by the feature points and found the number of correspondences based on this overlap error [24].

Recall rates are defined by the equation below :

$$Recall = \frac{number_of_correspondences}{number_of_correct_matches} \quad (3-53)$$

Number of correspondences between the reference image and the transformed or training image is calculated by comparing the local regions formed by feature points and by counting the number of corresponding points with eye inspection. This is based on the very simple idea and logic explained in the study of Mikolajczyk, and Schmid, [24]. Finding number of corresponding areas by using homographies and overlap error is a relatively complicated task.

Number of correct matches is also calculated by counting the total number of correct matches with eye inspection in the analysis of this study. The feature points outside the region we are interested, are neglected and not used in the matching accuracy calculations.

“1-precision” rate is defined as :

$$1 - \text{Precision} = \frac{\text{number_of_false_matches}}{\text{number_of_correct_matches} + \text{number_of_false_matches}} \quad (3-54)$$

The precision rate is calculated by the relation (3-55) below :

$$\text{Precision} = \frac{\text{number_of_correct_matches}}{\text{number_of_total_matches}} \quad (3-55)$$

Number of total matches is calculated by the simulation code, which includes correct, false matches and matches outside the region of interest. Matches outside the region of interest are neglected and subtracted from the total number of matches found by the code.

Matching accuracy is calculated after recall and precision rates are obtained for each reference and transformed or training image pair in the analysis of this study. Matching accuracy is calculated as follows :

$$\text{Matching_Accuracy} = \left(\frac{\text{Precision_Rate} + \text{Recall_Rate}}{2} \right) \times 100 \quad (3-56)$$

The ideas behind interest point extraction, mathematical details of feature detectors, description of local region defined by interest points and matching of these local regions between reference and transformed image pairs are explained up to this point in this chapter. Experimental setups, simulations performed and results of the experiments are discussed in the next chapter.

CHAPTER 4

SIMULATIONS AND RESULTS

Comparison of images on the basis of correspondences and finding matches is a relatively difficult task, especially for infrared-band images. As a consequence of this fact, design of controlled experiments for this study becomes important. Having these facts in mind, experimental setups are designed thoroughly and carefully, considering different aspects of the problem.

Four types of experimental studies are performed to validate the facts and ideas described in Chapter 3. Firstly, matching of MWIR and LWIR-band images of a blackbody is evaluated for different temperatures for each feature detector. Secondly, matching performance for images of real objects, face and a vehicle, in LWIR and visual-bands are evaluated by using each feature detector. Thirdly, matching performance of each feature detector is evaluated for images in all of the bands with different rotations of the mask. Finally, matching performance of each detector is evaluated for images in each band with increasing distances.

A brief explanation on the way how infrared images are obtained and represented in simulations is discussed in the next section. Then, each of the four experiments is explained along with their experimental setups, visual and graphical results.

4.1 Infrared Band Images

Infrared-band images are used in this study in order to evaluate the effectiveness of feature point detectors and matching performance of descriptors in infrared-bands. Moreover, the performance of the feature detectors each with its descriptor used is evaluated for visual-band.

Images obtained in two bands of infrared spectrum are used in the analysis of experiments. These bands are listed below.

1. Mid-Wave infrared-band (1-5 μm in wavelength)
2. Long-Wave infrared-band (8-12 μm in wavelength)

Two types of infrared cameras are used during the controlled experiments, which are MWIR camera and LWIR camera. The infrared cameras are high resolution and high speed cameras. Focal Plane Array (FPA) type detectors are used in each infrared camera. MWIR camera operates in 1-5 μm infrared-band, whereas LWIR camera operates in 8-12 μm infrared-band. These spectral bands are each of the camera detector's active detection ranges. The infrared cameras used in the experiments are also separate components of the IRLook (An advanced mobile, infrared signature measurement, data reduction, and analysis system), a visualization system by ASELSAN. The details of this system are explained in the study of Çukur et.al [43].

Infrared cameras are controlled and set up by their separate computers. These cameras have their own control and data acquisition software, installed in their computers. Both of the infrared cameras operate on the basis of Automatic Gain Control (AGC). They have analog video outputs as well as 14-bit digital video output. The video frames can be stored on the computers with their special data acquisition software.

The outputs of MWIR and LWIR cameras are 256x256 pixel infrared video frames. Each pixel has its associated 14-bit intensity value. The 14-bit digital levels should be between 3000 and 13000 units in order for the image data to be meaningful. Integration time parameter is adjusted in this study to obtain images within this range accordingly.

Spectral band pass filters can be installed in each of the infrared cameras. Several band pass filters are used with the infrared cameras, in order to obtain meaningful and comparable infrared radiation data in the images.

MWIR and LWIR cameras have special high quality lenses, in order to focus the target with high accuracy and fill the camera field of view. Some of these lenses belonging to each infrared camera are used in the experiments in order to fill the field of view of blackbody and other real image targets.

4.2 LWIR versus MWIR Infrared Images Experiment

LWIR versus MWIR infrared images experiment is performed in order to understand the effect of temperature on the matching accuracy in each of the corresponding infrared-band. This experiment is important in order to understand which infrared-band is most accurate in matching performance of objects with changing temperature provided that the operational and physical parameters are controlled and fixed. Visual-band images are not used in this experiment.

4.2.1 Experimental Setup

Experimental studies for MWIR versus LWIR infrared images experiment are performed by taking images in two infrared-bands, namely mid-wave infrared-band (MWIR) and long-wave infrared-band (LWIR).

Two infrared cameras, which are described in Section 4.1, are used in this experiment. 100mm focal length lenses of MWIR and LWIR cameras are used. The

images are taken from 4m distance fixed. Both of the infrared cameras are located on a turntable standing above the tripod, so that the focal plane of each of the cameras is closer to each other for the experiment.

Infrared photos of a blackbody device are taken. This blackbody device is 18cm by 18cm in size and it provides uniform heat distribution over its surface at a constant (set) temperature with good accuracy and stability, when it is operated. These blackbody devices are used as reference sources for infrared cameras. Information about blackbody concept and blackbody devices is discussed in Section 2.1. Blackbody temperature is set at appropriate temperatures for each infrared camera, as discussed in the next paragraphs. The photos are taken after blackbody is heated and its temperature is stabilized.

A typical mask built from white color cartoon is used during these experiments. Several dimension rectangular, hexagon and triangular shapes are formed by cutting the surface of the mask. These shapes are formed in order to obtain enough number of distinctive feature points in the analysis of the images. The mask is placed in front of the blackbody device before taking photos. It is rotated only 5 degrees to the left at 4m distance of the measurement. 5 degrees of rotation of the mask is performed in order to obtain a transformed or test image for the comparison of matching performance between reference images where mask is not rotated. The schema of the mask is shown in Figure 4-1 below.

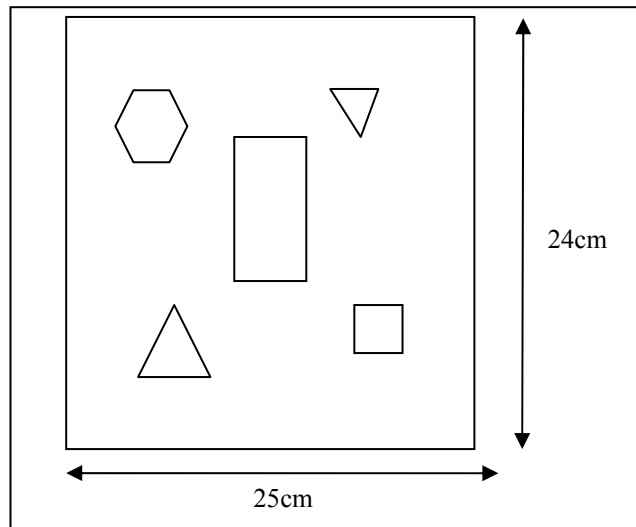


Figure 4-1 White Cartoon Mask of Experiment 1

The heating surface of the blackbody is black in color. The blackbody device (without mask) used in this experiment is shown in Figure 4-2 below. The device with mask is shown in Figure 4-3 below. These visual-band images are not used in the analysis of this experiment, presented for visualizing the experimental setup only.



Figure 4-2 Blackbody Device without Mask (Experiment 1)

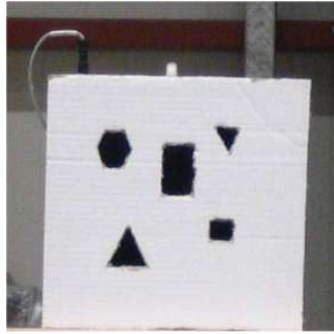


Figure 4-3 Blackbody Device with Mask (Experiment 1)

The typical infrared image obtained is presented in Figure 4-4 below. Note that the hot regions in the image are in white color, whereas colder regions relative to the hotter areas are in dark color. The images used in the analysis of this experiment are presented later in this section.



Figure 4-4 Typical Infrared Image of Blackbody Device with Mask (Experiment 1)

Spectral filters are used in each infrared camera in order to obtain meaningful and comparable digital levels of infrared images. The spectral band pass filters used for each infrared camera are listed below.

1. MWIR Camera, 3.5-5.0 μm bandpass filter
2. LWIR Camera, 8.6-9.3 μm bandpass filter

Two temperature values, low and high temperatures with respect to ambient temperature, are used in this experiment for understanding the effect of temperature on matching performance of feature detectors. The low temperature value is set at 32°C, whereas the high temperature is adjusted at 112°C. These values are used to adjust blackbody temperature for the LWIR infrared images.

The external factor that needs to be fixed is the response of the detectors of each infrared camera. MWIR and LWIR cameras have different type detectors. Their responsivity specification also changes. In order to compare images taken by LWIR camera detector and the images taken by MWIR camera detector, integration times of each camera should be adjusted to make the coming energy to the camera detectors equal.

Equalization of detector responses or coming energy from blackbody device is performed by following the procedure listed below. The masks are not used during this equalization.

1. Blackbody device is set at temperature value of LWIR-band for the test temperature. (i.e. 32°C)
2. The blackbody device is waited so that the temperature value is stabilized.
3. LWIR-band camera integration time is set at 300 μs (fixed).
4. A measurement from LWIR camera is taken.
5. Mean digital value is read from the Data Acquisition software of LWIR camera and it is noted.
6. Measurement from MWIR camera is observed real time. Mean digital value is read from the Data Acquisition software of MWIR camera. The

integration time of MWIR camera is adjusted until mean digital value is made equal with the value obtained in Step 5.

7. The integration time of MWIR camera that makes mean digital values equal is noted.

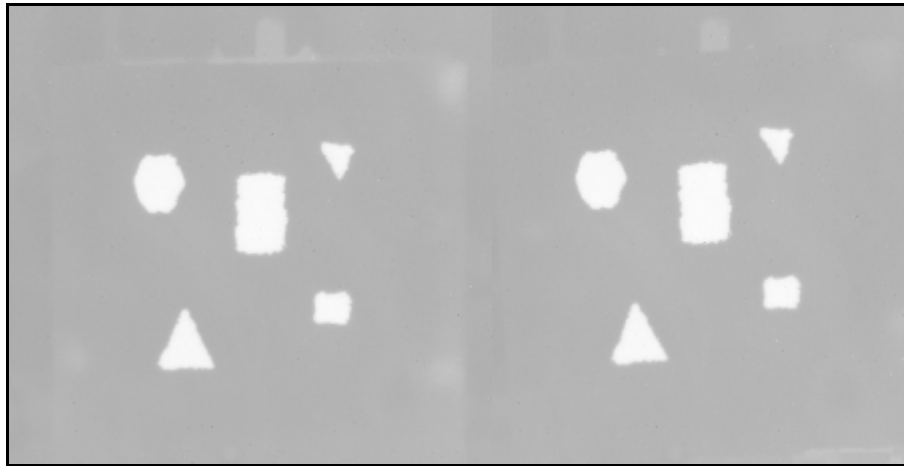
The resultant integration values found for MWIR camera is summarized in Table 4-1 below.

Table 4-1 Integration Times for MWIR Camera

No	Test Set	Temperature for LWIR and MWIR Band	Integration Time of LWIR Camera (μs)	Integration Time of MWIR Camera (μs)
1.	Low	32°C	300	805
2.	High	112°C	300	273

The detector responses of LWIR and MWIR cameras are equalized by adjusting the integration times given in Table 4-1. Blackbody temperature is adjusted at 32°C, LWIR camera integration time is 300 μs and MWIR integration time is set at 805 μs . Infrared images are captured with this low temperature configuration. Then, high temperature configuration with blackbody at 112°C, LWIR integration time at 300 μs and MWIR integration time at 273 μs is applied.

Infrared images are captured with mask not rotated and also for rotated 5 degrees to the left for each configuration. The obtained infrared images, that are analyzed, are presented in Figure 4-5, Figure 4-6, Figure 4-7, and Figure 4-8, for LWIR-band 32°C, MWIR-band 32°C, LWIR-band 112°C, and MWIR-band 112°C, respectively.

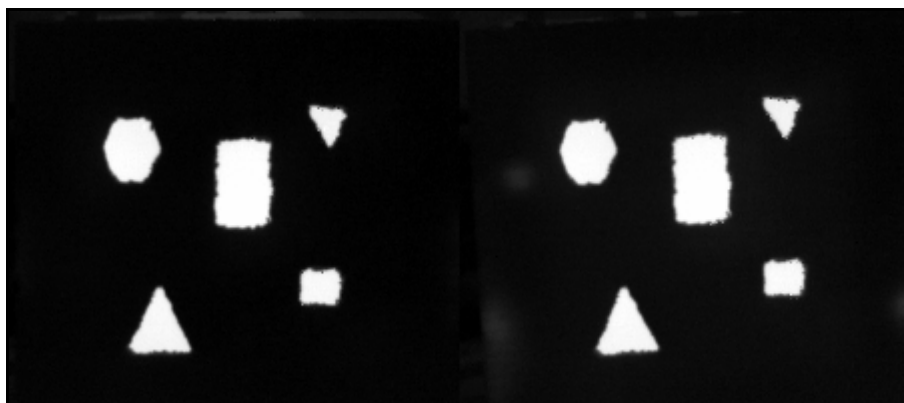


(a)

(b)

Figure 4-5 LWIR Band Infrared Images, 32°C Blackbody, 4m.

a) Mask not rotated, (b) Mask rotated 5 degrees to left



(a)

(b)

Figure 4-6 MWIR Band Infrared Images, 32°C Blackbody, 4m.

(a) Mask not rotated, (b) Mask rotated 5 degrees to left



(a)

(b)

Figure 4-7 LWIR Band Infrared Images, 112°C Blackbody, 4m.

(a) Mask not rotated, (b) Mask rotated 5 degrees to left



(a)

(b)

Figure 4-8 MWIR Band Infrared Images, 112°C Blackbody, 4m.

(a) Mask not rotated, (b) Mask rotated 5 degrees to left

The reference images are the ones with mask not rotated on the left part of each image pair. The transformed images are the images captured with mask in front of the blackbody rotated 5 degrees to left. The results of this experiment are presented in the next section.

4.2.2 Results of LWIR versus MWIR Experiment

Infrared images taken in LWIR and MWIR-bands are inputs to the feature detectors. The reference images, with mask not rotated and the transformed or training images, with mask rotated 5 degrees to left are read by MATLAB and feature points are extracted by using SIFT, Harris-Laplace, Hessian-Laplace and SURF detectors.

Feature points of LWIR infrared image pair belonging to 112°C, extracted by SIFT, Harris-Laplace, Hessian-Laplace and SURF detectors are presented in Figure 4-9, Figure 4-10, Figure 4-11, and Figure 4-12 below, respectively. In these images, red circles denote the local regions determined by each feature point. The center of each circle is the position of the feature point. The larger area of the circle means that the feature point is detected at a higher scale.

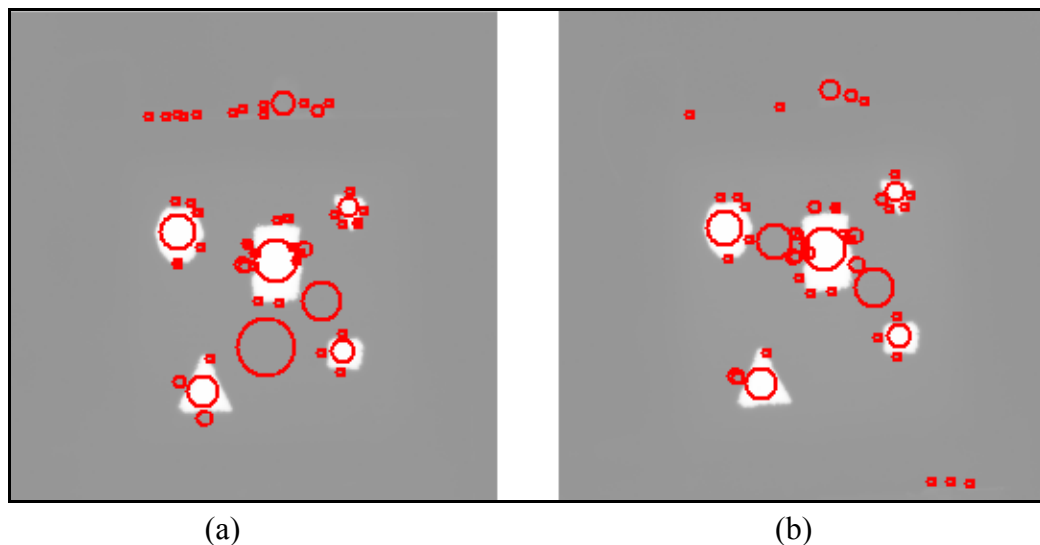
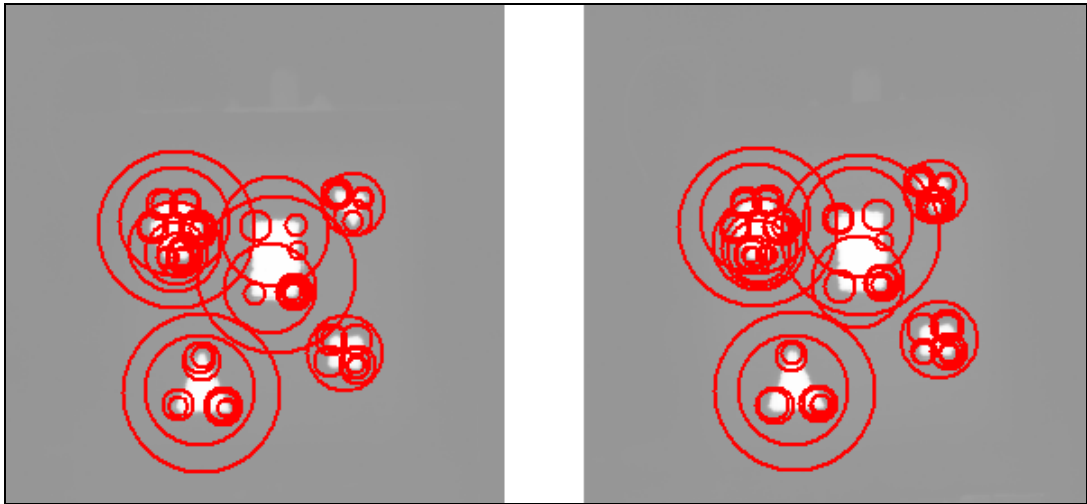


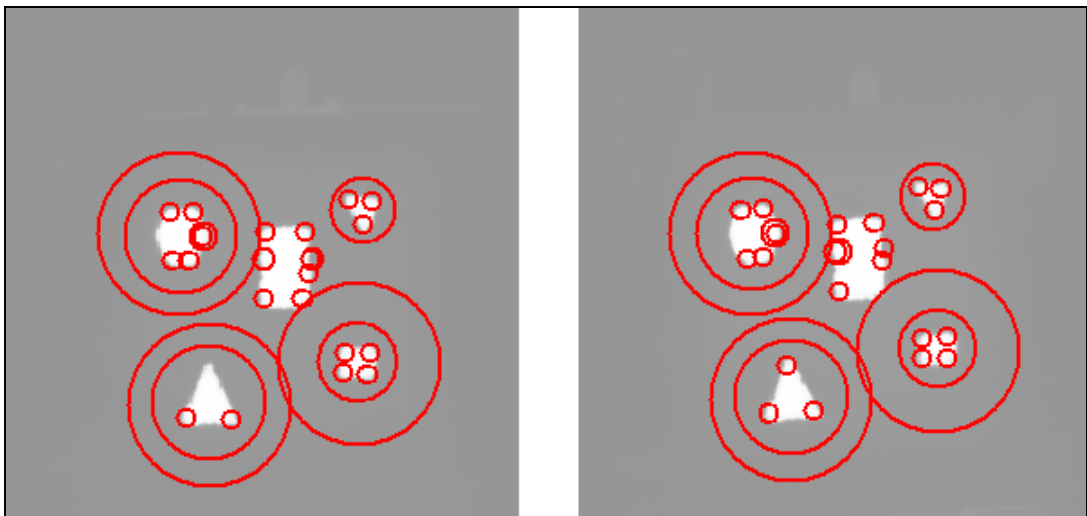
Figure 4-9 LWIR 112°C Infrared Image Feature Points Extracted by SIFT Detector.

(a) Mask not rotated, (b) Mask rotated 5 degrees to left



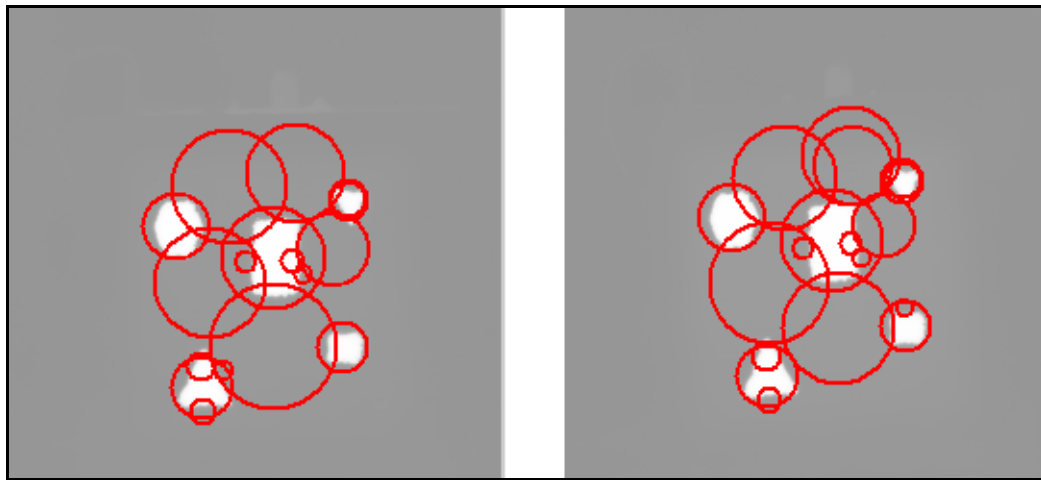
(a) (b)
 Figure 4-10 LWIR 112°C Infrared Image Feature Points Extracted by Harris-Laplace Detector.

(a) Mask not rotated, (b) Mask rotated 5 degrees to left



(a) (b)
 Figure 4-11 LWIR 112°C Infrared Image Feature Points Extracted by Hessian-Laplace Detector.

(a) Mask not rotated, (b) Mask rotated 5 degrees to left



(a)

(b)

Figure 4-12 LWIR 112°C Infrared Image Feature Points Extracted by SURF Detector.

(a) Mask not rotated, (b) Mask rotated 5 degrees to left

The extracted features of the other temperatures and MWIR-band images have parallel extracted feature figures to that of LWIR 112°C images shown above. It can be seen that SIFT, Hessian-Laplace and SURF feature points are detected at blob-like regions, whereas Harris-Laplace feature points are detected near or at corners. Also, feature points extracted are rotation invariant, since almost the entire feature regions found in the reference images on the left are observed in the transformed image as well.

The matched feature points for LWIR 112°C image pairs is shown in Figure 4-13, Figure 4-14, Figure 4-15, and Figure 4-16 below for SIFT, Harris-Laplace, Hessian-Laplace and SURF detectors, respectively. Only one of the typical matched pair is presented in each of these figures.

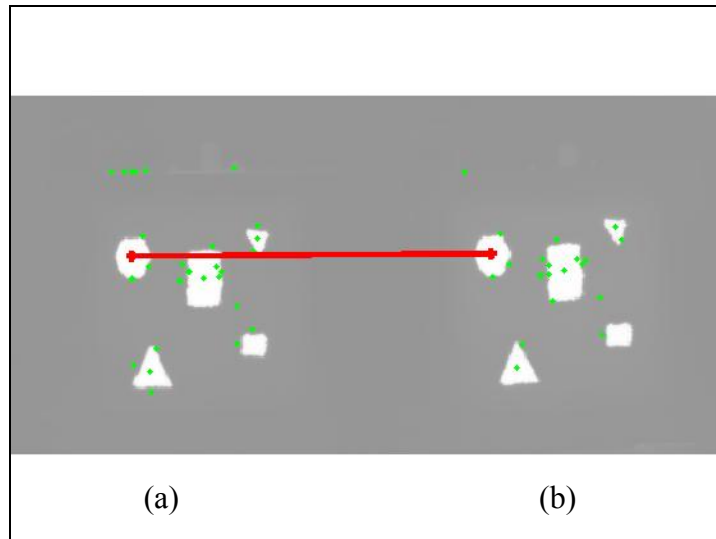


Figure 4-13 Matched Feature Points of LWIR 112°C Infrared Images, SIFT detector.

(a) Mask not rotated, (b) Mask rotated 5 degrees to left

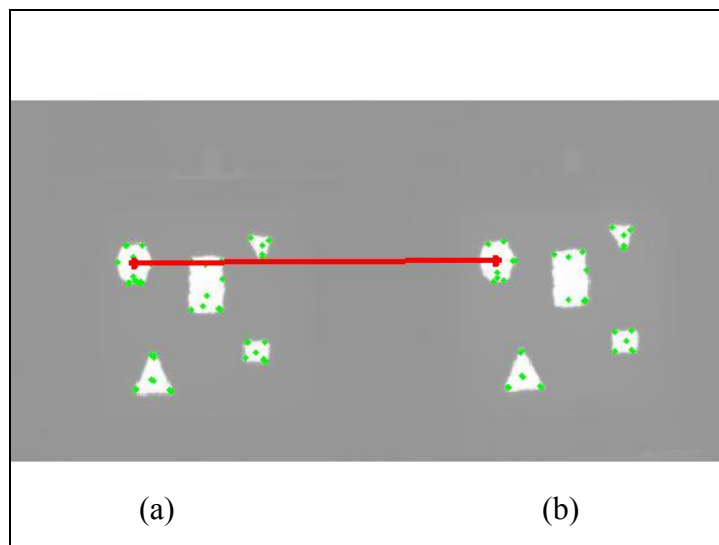


Figure 4-14 Matched Feature Points of LWIR 112°C Infrared Images, Harris-Laplace detector.

(a) Mask not rotated, (b) Mask rotated 5 degrees to left

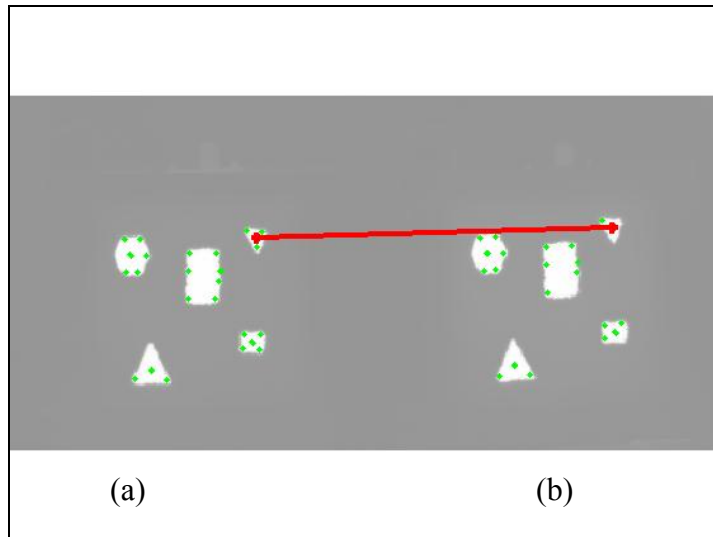


Figure 4-15 Matched Feature Points of LWIR 112°C Infrared Images, Hessian-Laplace detector.

(a) Mask not rotated, (b) Mask rotated 5 degrees to left

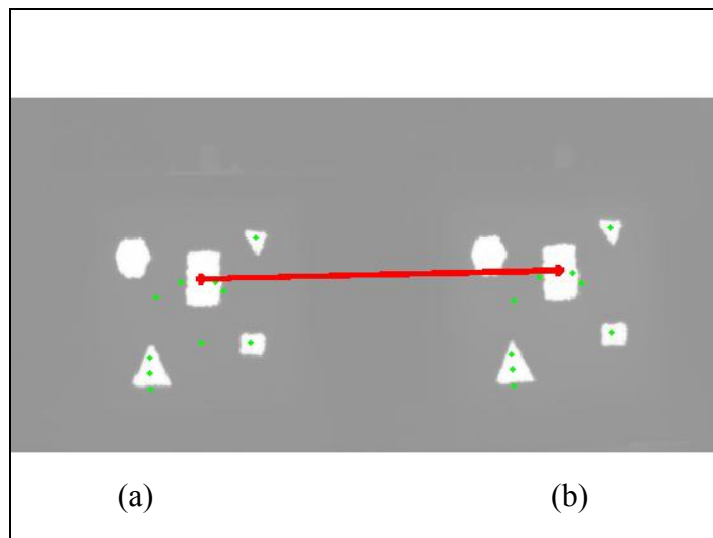


Figure 4-16 Matched Feature Points of LWIR 112°C Infrared Images, SURF detector.

(a) Mask not rotated, (b) Mask rotated 5 degrees to left

Number of correspondences and number of correct matches are counted with eye inspection. Then, precision and recall rates are calculated and matching accuracy is found. The graphical results showing the change of matching accuracy with respect to temperature values are presented in Figure 4-17 and Figure 4-18 below for LWIR and MWIR-bands, respectively.

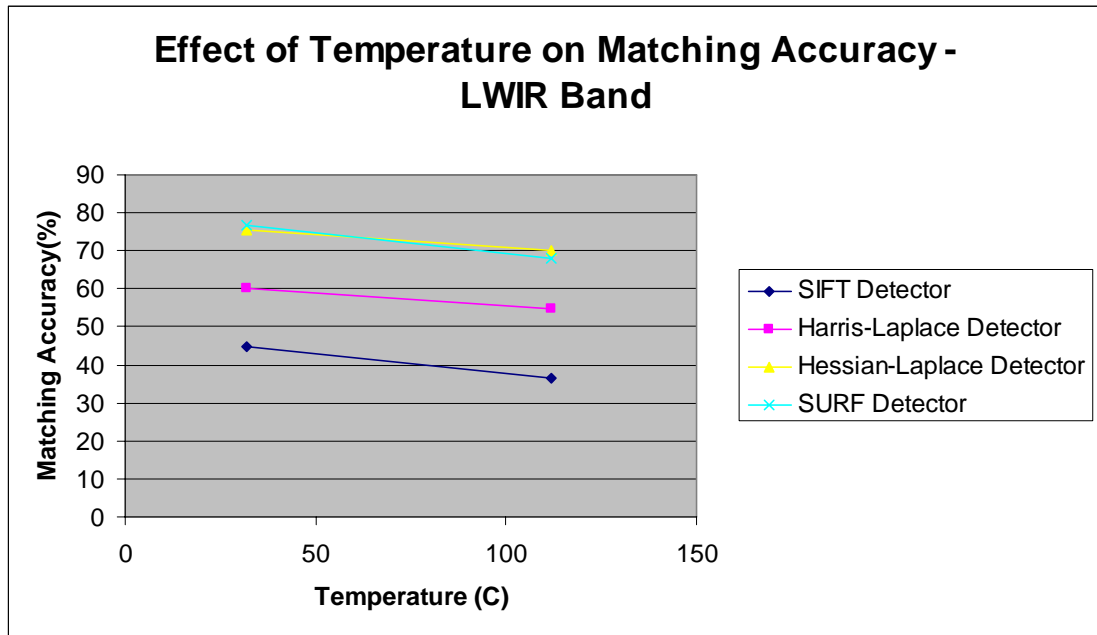


Figure 4-17 Matching Accuracy with Changing Temperature-LWIR Band

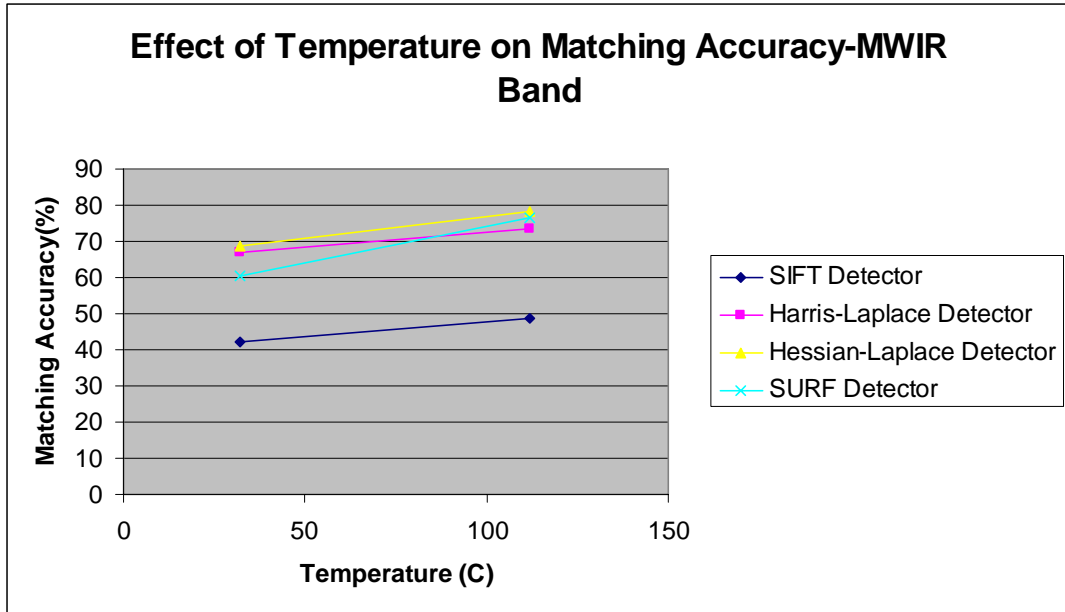


Figure 4-18 Matching Accuracy with Changing Temperature-MWIR Band

The matching accuracy values increase with increasing temperature in MWIR-band. Matching accuracy of targets is better in MWIR-band for all detectors as the temperature of the targets increase. LWIR-band images have opposite results obtained for the MWIR-band. Matching accuracy values decrease with increasing temperature for LWIR-band images.

The discussion of these results is given in the Conclusions chapter of this study.

4.3 LWIR versus Visual Band Images Experiment

LWIR versus visual-band images experiment is performed with real targets. The aim of this experiment is to examine which feature detector is more effective in finding correspondences of face and vehicle targets in different bands. MWIR-band is not used in this experiment, since MWIR-band is more effective in identifying matches of hot targets, as it is observed in LWIR versus MWIR infrared images

experiment. Face and vehicle are low temperature targets and it is important to understand which feature detector is successful in these bands.

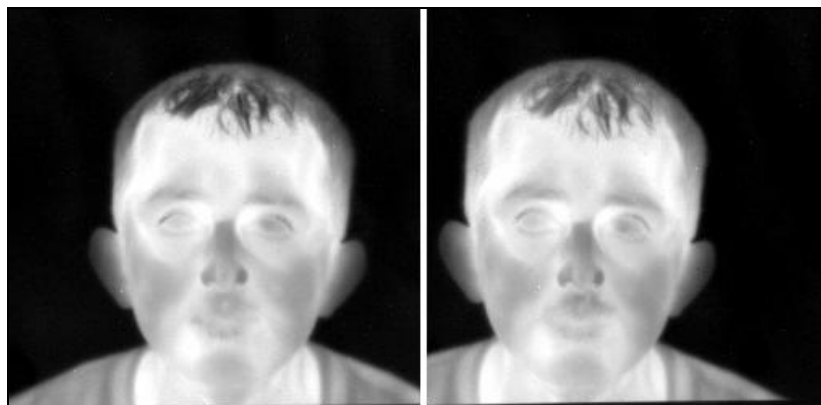
4.3.1 Experimental Setup

Experimental studies for LWIR versus visual-band images experiment are performed by taking images in two bands, namely long-wave infrared-band (LWIR) and visual-band.

LWIR infrared camera and an off-the-shelf Canon Powershot A75 visual digital camera are used during this experiment. Face and vehicle images are captured in both bands. 100mm focal length lens of LWIR camera is used for face images, whereas 50mm focal length lens of LWIR is used for vehicle images.

The face images are taken from 4m for changing rotation and at 4m, 8m and 12m for changing distance without any rotation of face. Vehicle images are taken at 43.5m for changing rotation, on the other hand. Both of the infrared and visual-band cameras are located on a turntable standing above the tripod, so that the focal plane of each of the cameras is closer to each other for the experiment.

8.6-9.3 μm spectral bandpass filter is used in LWIR infrared camera for face images. Face images with rotation change are captured in LWIR and visual-bands. Face images without rotation is taken as reference image at 4m distance. Then, a slided face image to the left for LWIR-band and to the right for visual-band (i.e. transformed image without rotation) and another face image turned to the left for LWIR-band and to the right for visual-band with approximately 5 degrees rotation are taken at 4m distance again. The image pairs that are analyzed are shown in Figure 4-19, Figure 4-20, Figure 4-21, and Figure 4-22 below, for LWIR (first two images) and visual-bands (second two images), respectively.

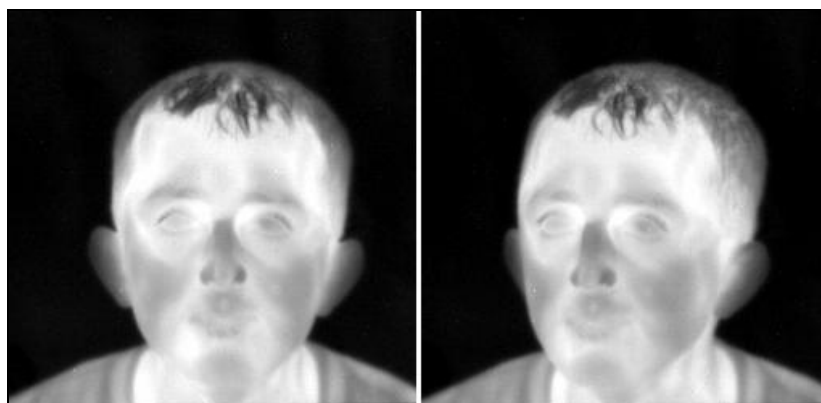


(a)

(b)

Figure 4-19 LWIR Band Infrared Images of Face, 4m, Slided

(a) Reference image, (b) Face slided to left, transformed image



(a)

(b)

Figure 4-20 LWIR Band Infrared Images of Face, 4m, Turned to left

(a) Reference image, (b) Face turned approximately 5 degrees to left, transformed image



(a) (b)
Figure 4-21 Visual Band Images of Face, 4m, Slided

(a) Reference image, (b) Face slided to right, transformed image



(a) (b)
Figure 4-22 Visual Band Images of Face, 4m, Turned to right

(a) Reference image, (b) Face turned approximately 5 degrees to right, transformed image

Face images with changing distances are captured in LWIR and visual-bands. The reference images are the face images taken at 4m distance for each band in this case. Transformed images are the images at 8m and 12m for each band. The image pairs that are analyzed, are shown in Figure 4-23 and Figure 4-24 for LWIR-band, in Figure 4-25, and Figure 4-26 for visual-band below, respectively.

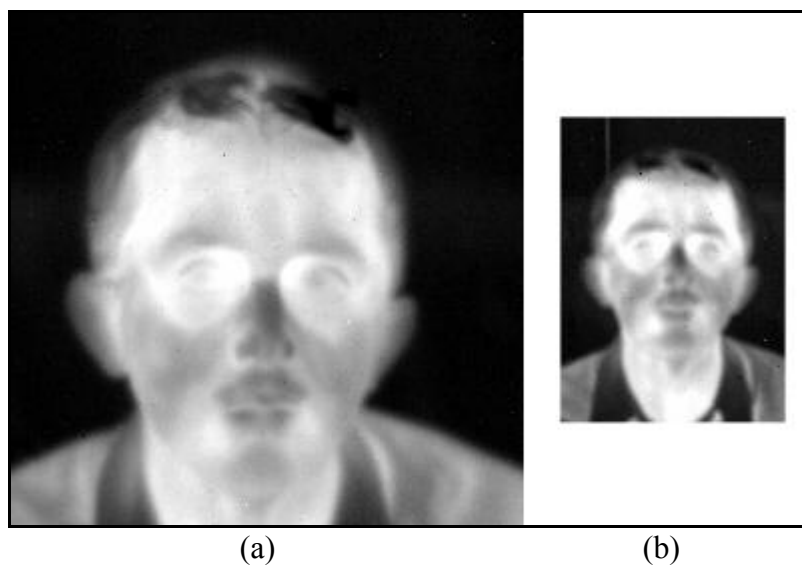


Figure 4-23 LWIR Band Infrared Images of Face for 4m and 8m.

(a) Reference image at 4m, (b) Transformed image at 8m

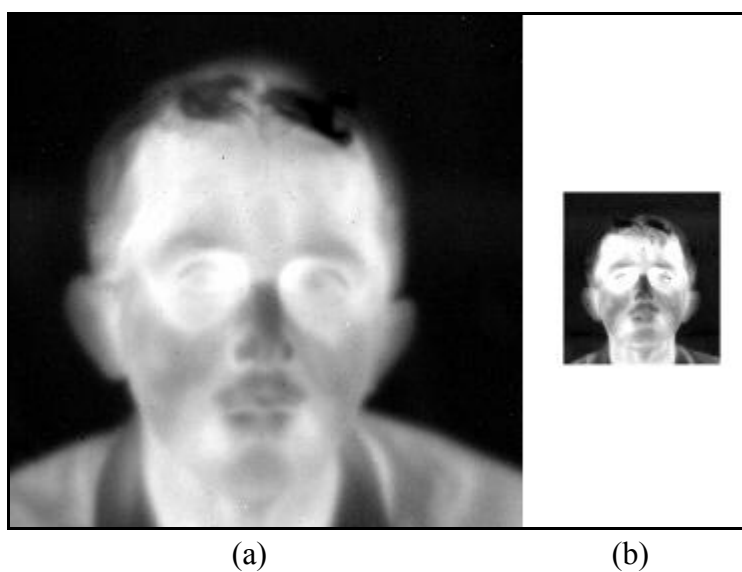


Figure 4-24 LWIR Band Infrared Images of Face for 4m and 12m.

(a) Reference image at 4m, (b) Transformed image at 12m



(a) (b)
Figure 4-25 Visual Band Images of Face for 4m and 8m.

(a) Reference image at 4m, (b) Transformed image at 8m



(a) (b)
Figure 4-26 Visual Band Images of Face for 4m and 12m.

(a) Reference image at 4m, (b) Transformed image at 12m

Vehicle images with rotation change are captured in LWIR and visual-bands at 43.5m distance fixed. LWIR images of the vehicle are captured in 8.6-9.3 μm spectral band. Vehicle images without rotation is taken as reference image, whereas transformed images are taken with vehicle approximately 5 degrees and 45 degrees turned to right in both bands. The image pairs that are analyzed are shown in Figure 4-27, Figure 4-28, Figure 4-29, and Figure 4-30 below, for LWIR (first two pair of images) and visual-bands (second two pair of images), respectively.



Figure 4-27 LWIR Band Infrared Images of Vehicle, 0 degree and 5 degree.

(a) Reference image, not rotated, (b) Transformed image rotated approximately 5 degrees right



Figure 4-28 LWIR Band Infrared Images of Vehicle, 0 degree and 45 degree.

(a) Reference image, not rotated, (b) Transformed image rotated approximately 45 degrees right



(a)

(b)

Figure 4-29 Visual Band Images of Vehicle, 0 degree and 5 degree.

(a) Reference image, not rotated, (b) Transformed image rotated approximately 5 degrees right



(a)

(b)

Figure 4-30 Visual Band Images of Vehicle, 0 degree and 45 degree.

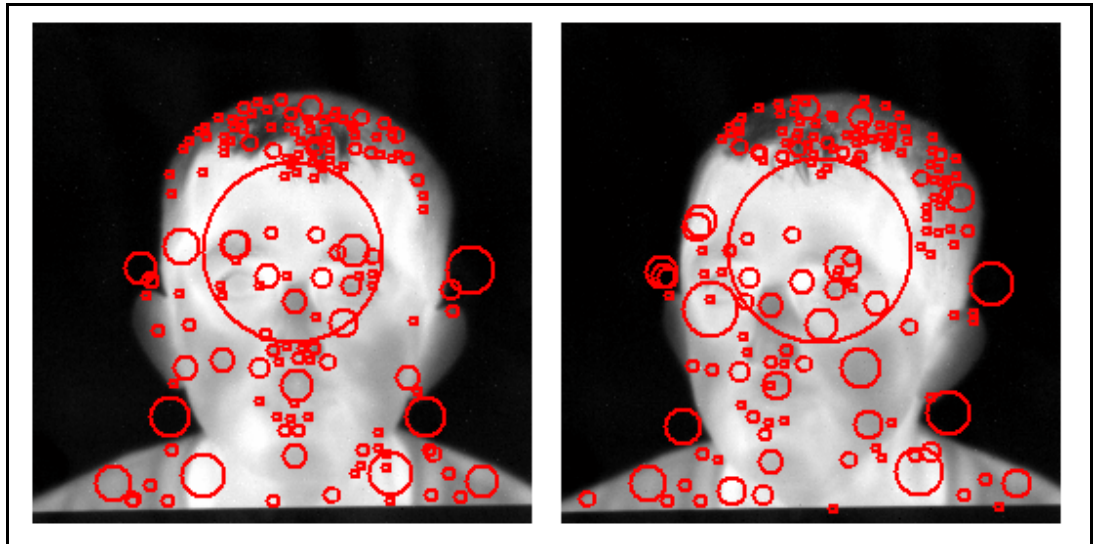
(a) Reference image, not rotated, (b) Transformed image rotated approximately 45 degrees right

The results of this experiment are presented in the next section.

4.3.2 Results of LWIR versus Visual Band Images Experiment

Feature points belonging to LWIR and visual-band face image pairs are extracted. The feature points of LWIR face images are presented in Figure 4-31, Figure 4-32, Figure 4-33, and Figure 4-34 below, extracted by SIFT, Harris-Laplace, Hessian-Laplace and SURF detectors, respectively. In these images, red circles indicate the local regions defined by each feature point. The radius of each circle is the

indication of the scale of that feature point. If the circle is larger, the feature point is extracted at a larger scale.

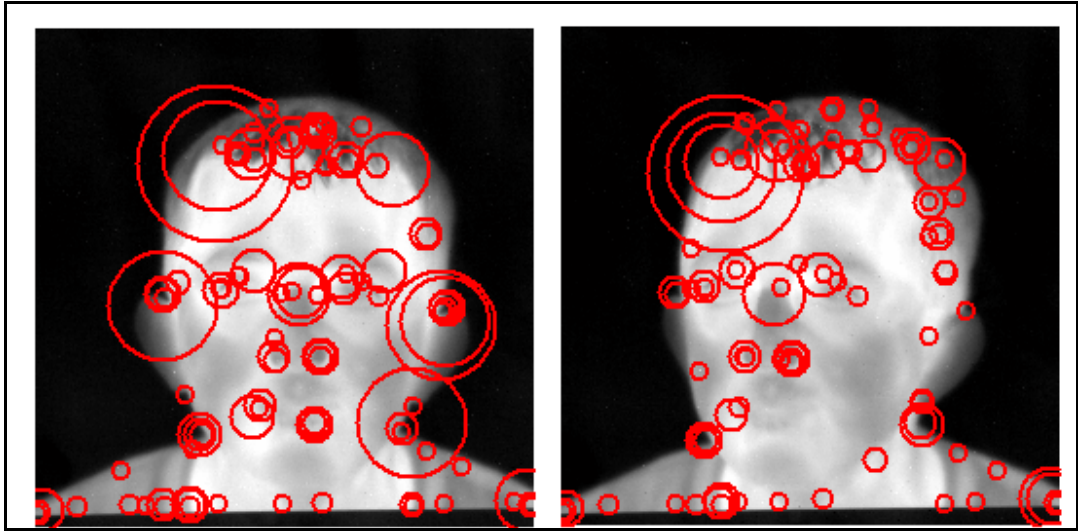


(a)

(b)

Figure 4-31 LWIR Band Face Image Feature Points Extracted by SIFT Detector.

(a) Reference image, (b) Face turned approximately 5 degrees to left, transformed image

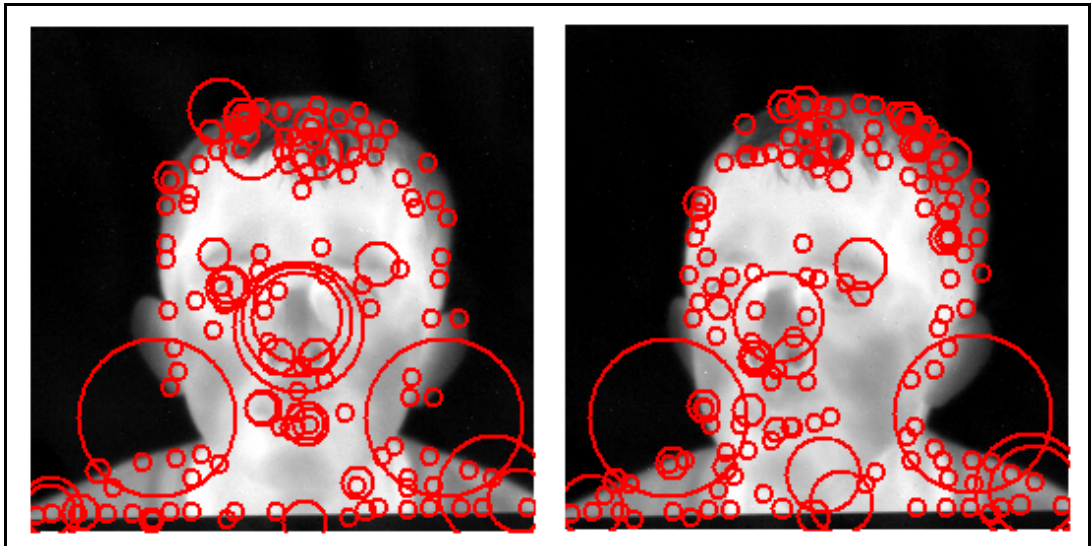


(a)

(b)

Figure 4-32 LWIR Band Face Image Feature Points Extracted by Harris-Laplace Detector.

(a) Reference image, (b) Face turned approximately 5 degrees to left, transformed image

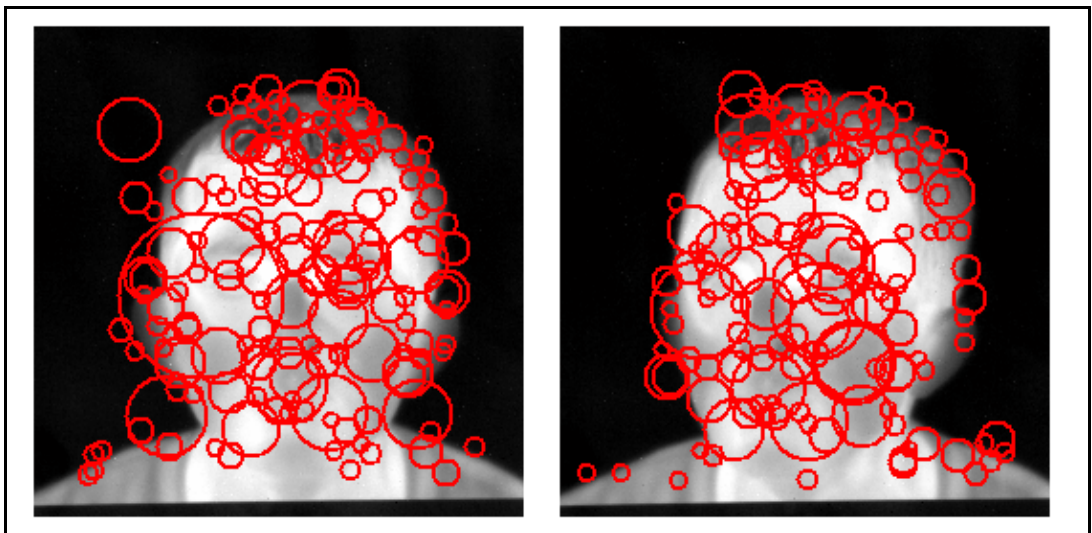


(a)

(b)

Figure 4-33 LWIR Band Face Image Feature Points Extracted by Hessian-Laplace Detector.

(a) Reference image, (b) Face turned approximately 5 degrees to left, transformed image



(a)

(b)

Figure 4-34 LWIR Band Face Image Feature Points Extracted by SURF Detector.

(a) Reference image, (b) Face turned approximately 5 degrees to left, transformed image

The extracted features of slided face images in LWIR-band have similar feature figures to those shown above, so they are not presented. It can be observed that, SIFT, Hessian-Laplace and SURF feature points are detected at round, curved and blob-like regions, whereas Harris-Laplace feature points are detected near regions having similar shape to corners.

Feature points of visual-band face images are presented in Figure 4-35, Figure 4-36, Figure 4-37, and Figure 4-38 below, for each detector.

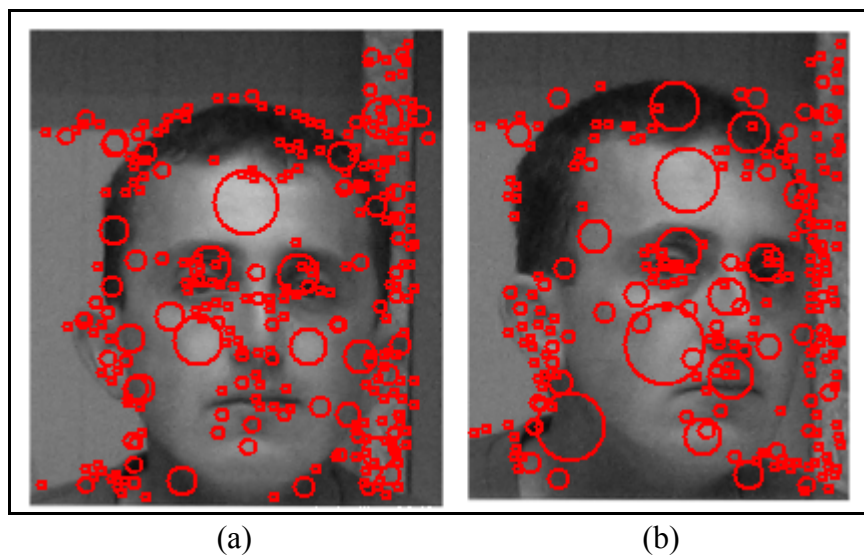
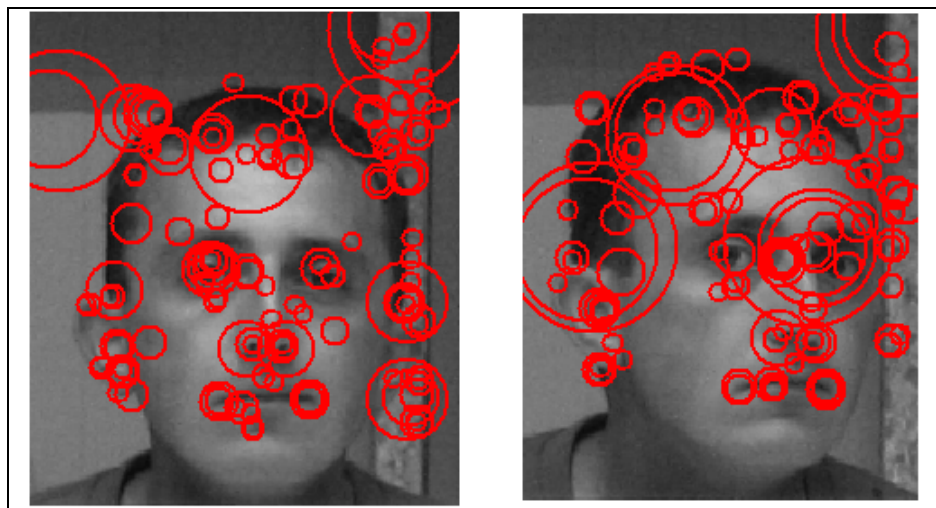


Figure 4-35 Visual Band Face Image Feature Points Extracted by SIFT Detector.

(a) Reference image, (b) Face turned approximately 5 degrees to right, transformed image

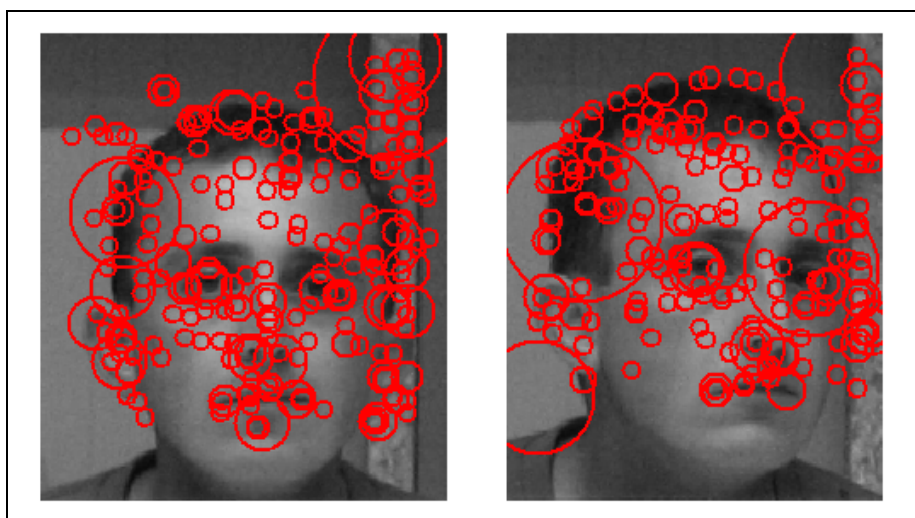


(a)

(b)

Figure 4-36 Visual Band Face Image Feature Points Extracted by Harris-Laplace Detector.

(a) Reference image, (b) Face turned approximately 5 degrees to right, transformed image

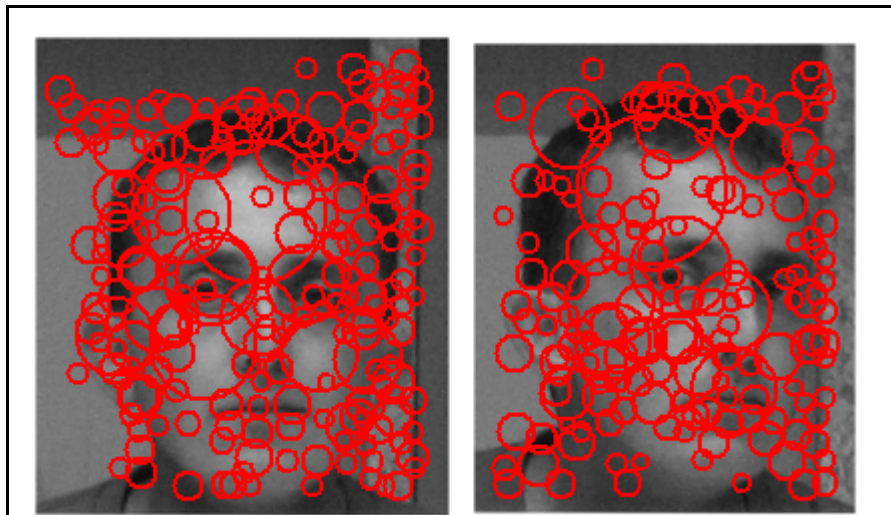


(a)

(b)

Figure 4-37 Visual Band Face Image Feature Points Extracted by Hessian-Laplace Detector.

(a) Reference image, (b) Face turned approximately 5 degrees to right, transformed image



(a)

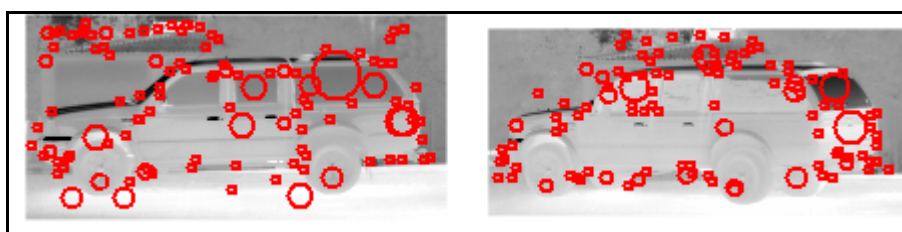
(b)

Figure 4-38 Visual Band Face Image Feature Points Extracted by SURF Detector.

(a) Reference image, (b) Face turned approximately 5 degrees to right, transformed image

The extracted features of slided face images in visual-band have similar feature figures to those shown above, so they are not presented.

The extracted feature results of the vehicle LWIR images are presented in Figure 4-39, Figure 4-40, Figure 4-41, and Figure 4-42 below, for each detector, respectively.



(a)

(b)

Figure 4-39 LWIR Band Vehicle Image Feature Points Extracted by SIFT Detector.

(a) Reference image, not rotated, (b) Transformed image rotated approximately 5 degrees right

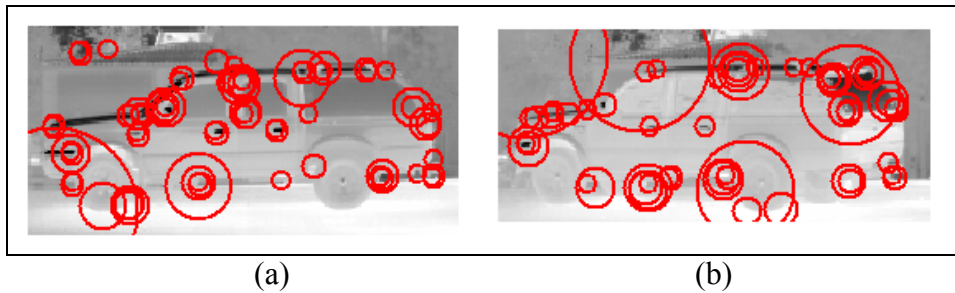


Figure 4-40 LWIR Band Vehicle Image Feature Points Extracted by Harris-Laplace Detector.

(a) Reference image, not rotated, (b) Transformed image rotated approximately 5 degrees right

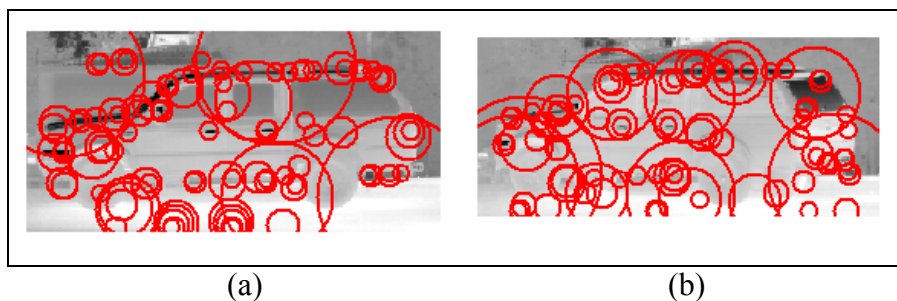


Figure 4-41 LWIR Band Vehicle Image Feature Points Extracted by Hessian-Laplace Detector.

(a) Reference image, not rotated, (b) Transformed image rotated approximately 5 degrees right

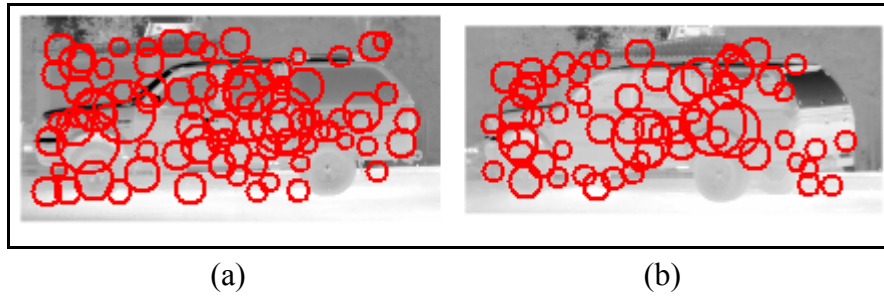


Figure 4-42 LWIR Band Vehicle Image Feature Points Extracted by SURF Detector.

(a) Reference image, not rotated, (b) Transformed image rotated approximately 5 degrees right

The extracted features of vehicle images turned approximately to right 45 degrees in LWIR-band have similar feature figures to those shown above, so they are not presented.

Matched features of LWIR-band face images are presented in Figure 4-43, Figure 4-44, Figure 4-45, and Figure 4-46 below for each feature detectors.

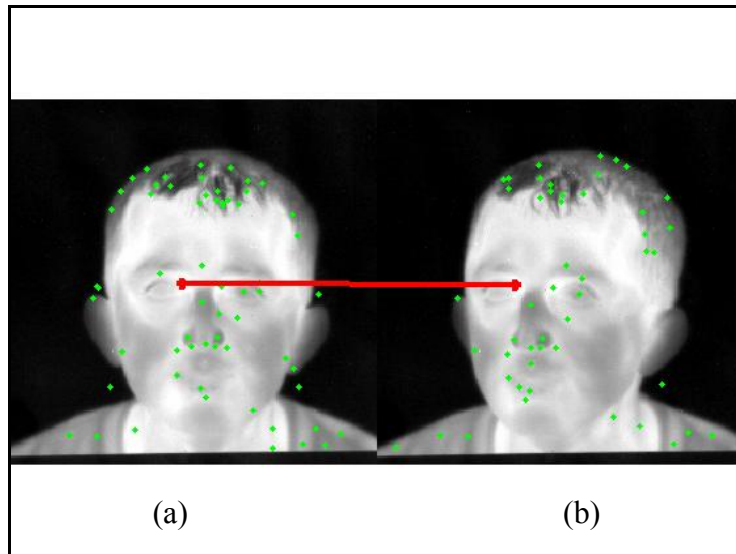


Figure 4-43 Matched Feature Points of LWIR Band Face Images, SIFT Detector.

(a) Reference image, (b) Face turned approximately 5 degrees to left, transformed image

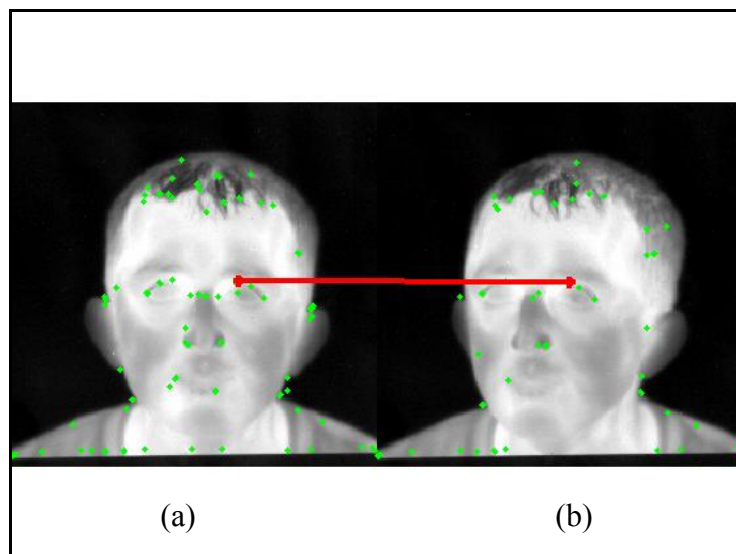


Figure 4-44 Matched Feature Points of LWIR Band Face Images, Harris-Laplace Detector.

(a) Reference image, (b) Face turned approximately 5 degrees to left, transformed image

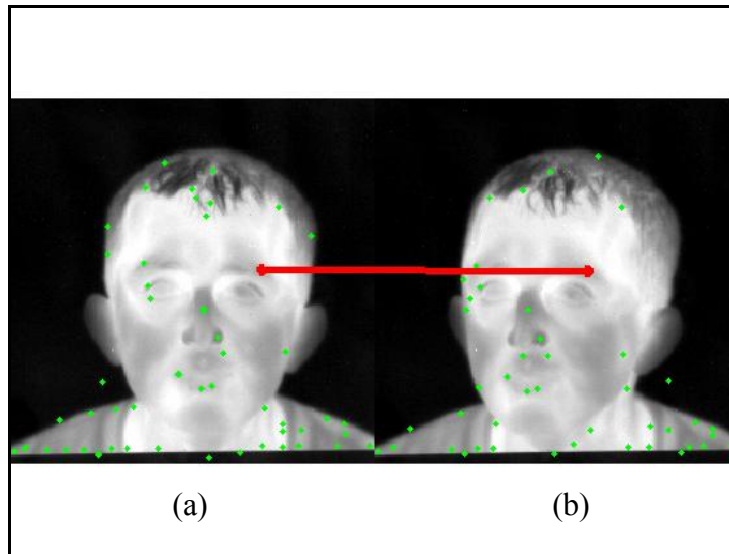


Figure 4-45 Matched Feature Points of LWIR Band Face Images, Hessian-Laplace Detector.

(a) Reference image, (b) Face turned approximately 5 degrees to left, transformed image

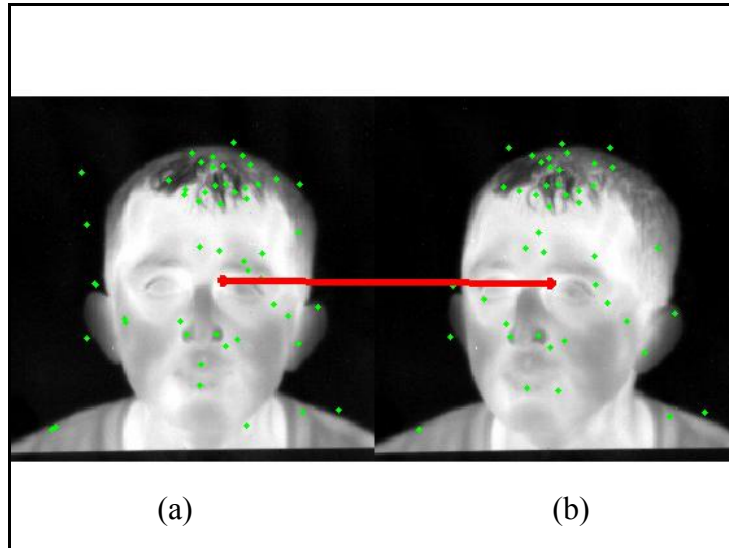


Figure 4-46 Matched Feature Points of LWIR Band Face Images, SURF Detector.

(a) Reference image, (b) Face turned approximately 5 degrees to left, transformed image

Some of the features of LWIR-band face images are not matched correctly. False matches of LWIR-band face images are presented in Figure 4-47, Figure 4-48, Figure 4-49, and Figure 4-50 below for each feature detectors. False matches are included in the analysis of this experiment like analysis of other experiments in matching accuracy calculations as it is explained in Section 3.8.

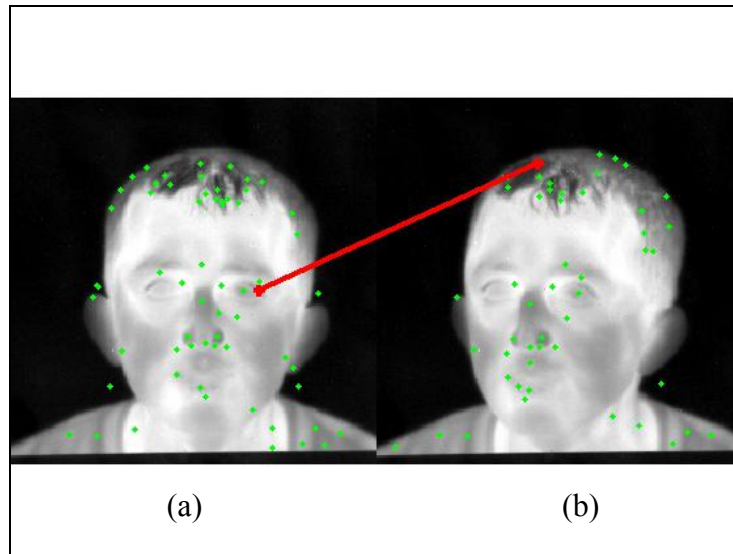


Figure 4-47 False Matches of LWIR Band Face Images, SIFT Detector.

(a) Reference image, (b) Face turned approximately 5 degrees to left, transformed image

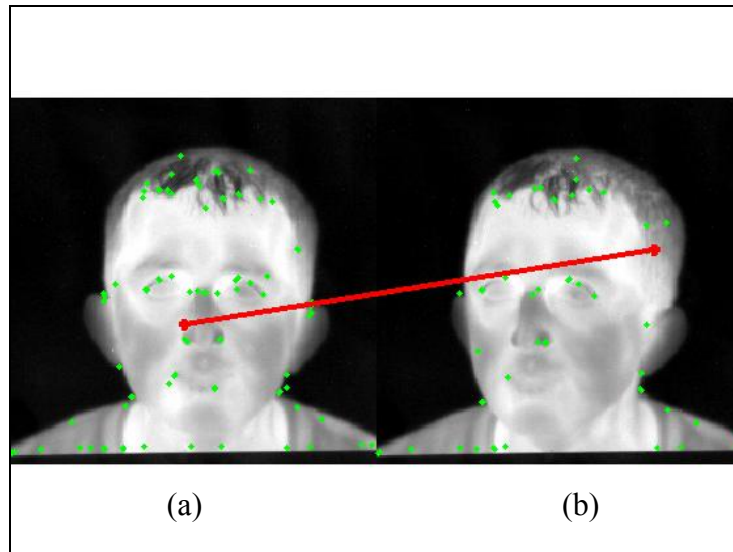


Figure 4-48 False Matches of LWIR Band Face Images, Harris-Laplace Detector.

(a) Reference image, (b) Face turned approximately 5 degrees to left, transformed image

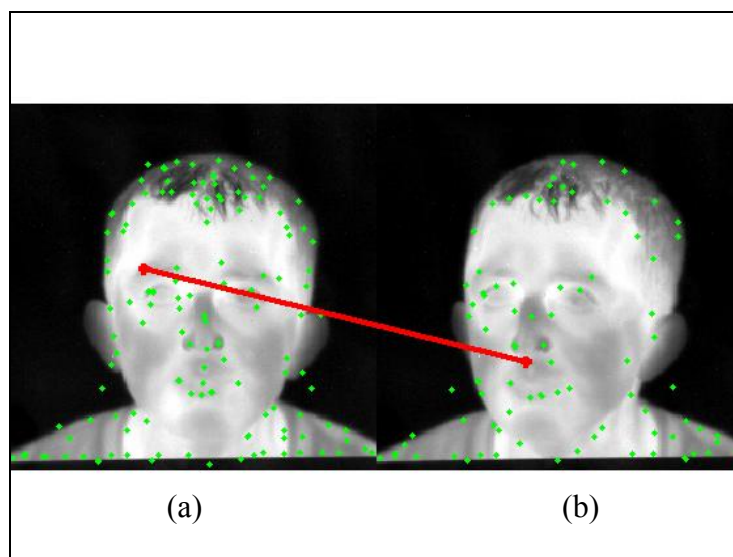


Figure 4-49 False Matches of LWIR Band Face Images, Hessian-Laplace Detector.

(a) Reference image, (b) Face turned approximately 5 degrees to left, transformed image

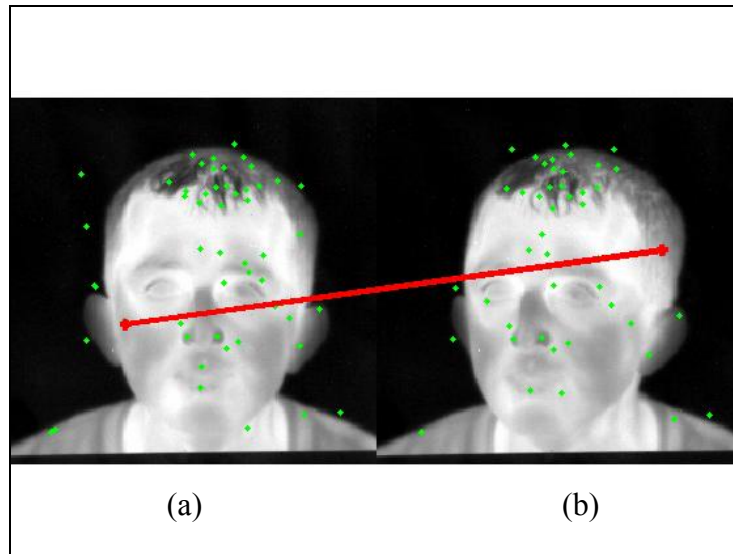


Figure 4-50 False Matches of LWIR Band Face Images, SURF Detector.

(a) Reference image, (b) Face turned approximately 5 degrees to left, transformed image

The matched feature points for LWIR-band face images comparing the reference image, face not rotated, and the slided ones are not presented. Moreover, the matched points for visual-band face images comparing the reference image, face not rotated, and the transformed images (slided and approximately 5 degrees rotated) are also not presented. Because, the matched points are similar in all of them.

The matched points outside the face region are not included in matching accuracy calculations. The graphical results showing the change of matching accuracy with respect to rotation angle are presented in Figure 4-51 and Figure 4-52, below for LWIR and visual-bands, respectively.

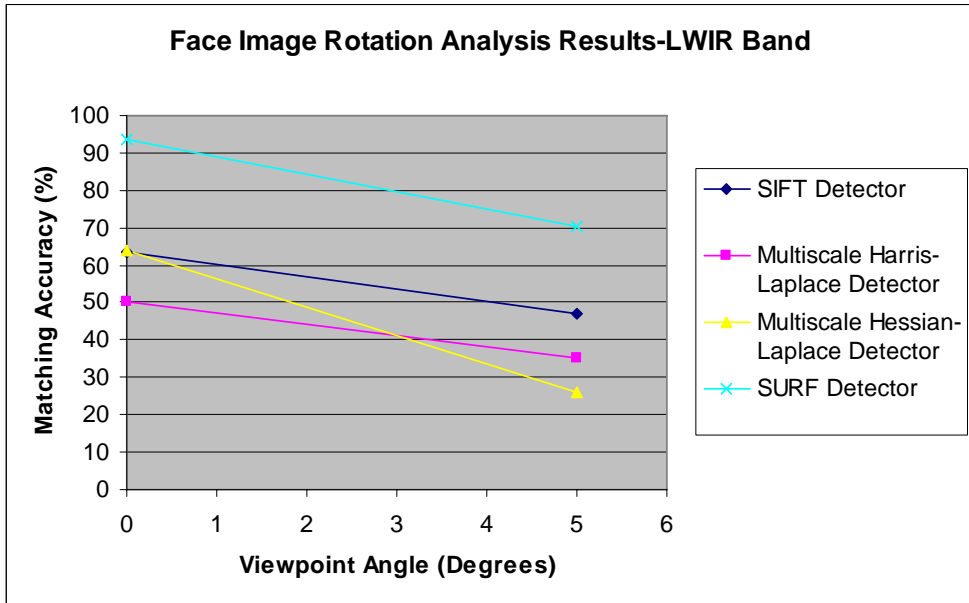


Figure 4-51 Matching Accuracy of Face Images with Changing Rotation Angle-
LWIR Band

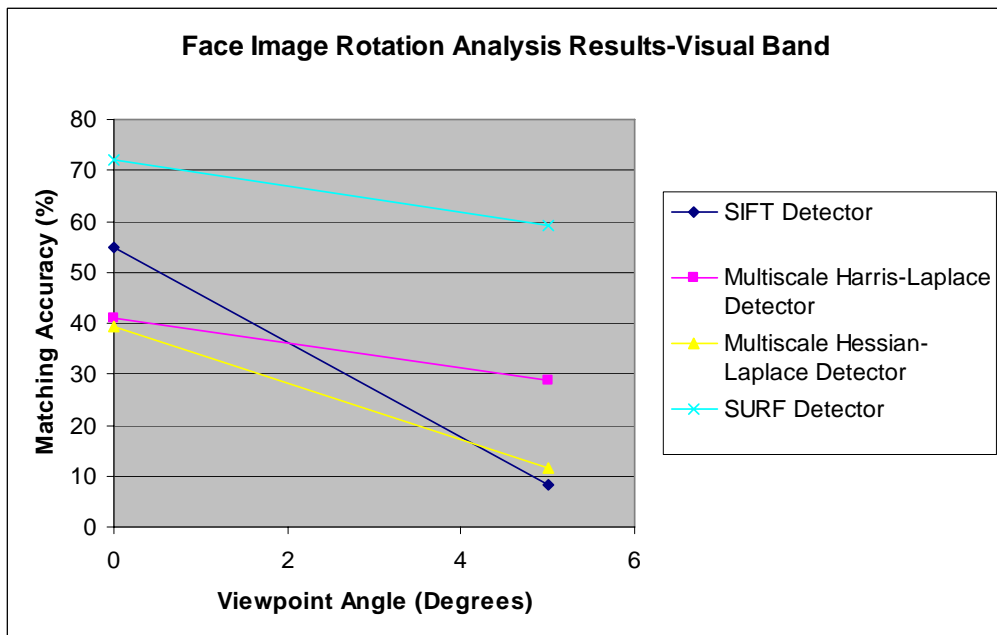


Figure 4-52 Matching Accuracy of Face Images with Changing Rotation Angle-
Visual Band

Matching accuracy between reference face images and transformed face images decrease as the rotation angle increases in both LWIR and visual-bands, for all detectors utilized. The zero rotation angles correspond to the slided face images. Moreover, matching accuracy performance of SURF detector for face images in terms of rotation angle is better in LWIR and visual-bands, when compared with other feature detector-descriptor performances.

Matching accuracy values of multiscale Hessian-Laplace detector decrease much faster than other detectors in LWIR-band. Because, Hessian detector is sensitive to noise than other detectors and there exists noise in LWIR-band images.

Matched features for LWIR and visual-band face images for changing distance case are not presented. Because, these matching points are similar to those of face images for changing rotation case. Graphical results are given instead.

The graphical results showing the change of matching accuracy of face images with respect to changing distances are presented in Figure 4-53 and Figure 4-54, below for LWIR and visual-bands respectively.

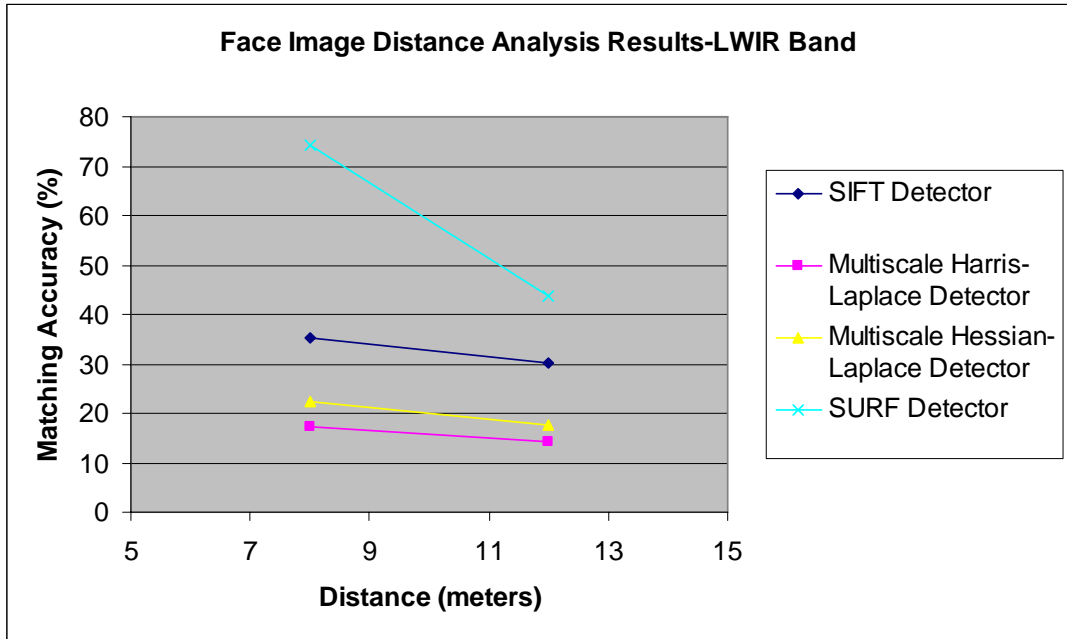


Figure 4-53 Matching Accuracy of Face Images with Changing Distance-LWIR Band

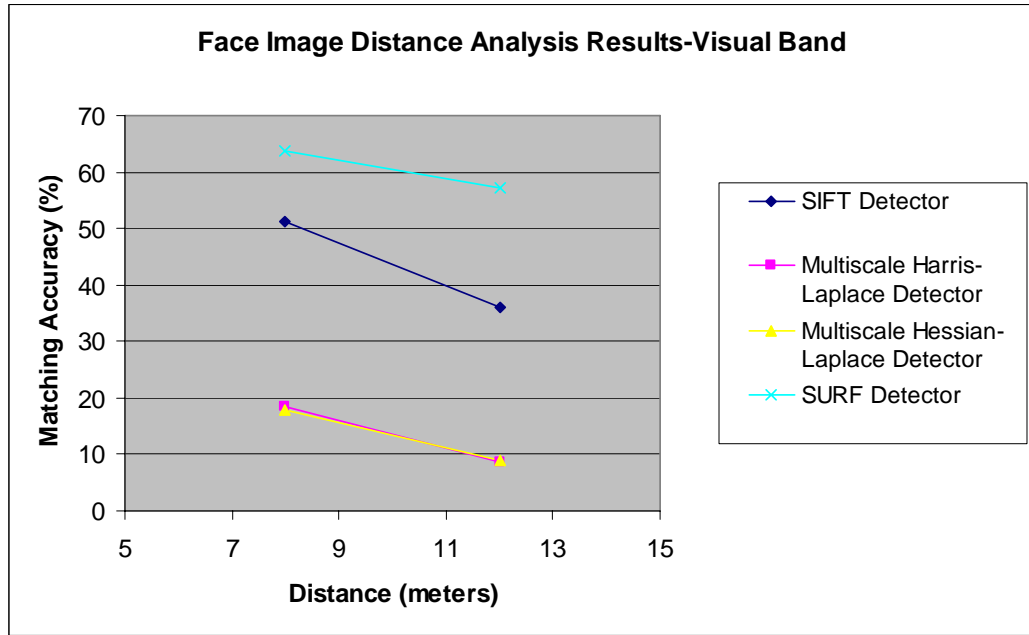


Figure 4-54 Matching Accuracy of Face Images with Changing Distance-Visual Band

Matching accuracy of face images decrease as the distance of measurement increases from 8m to 12m. This result is true for both LWIR and visual-bands, for all feature detectors utilized. Besides, matching accuracy performance of SURF detector for face images is better in terms of distance for both LWIR-band and visual-bands.

Matched features of vehicle images are shown in Figure 4-55, Figure 4-56, Figure 4-57, and Figure 4-58 below, for each feature detector.

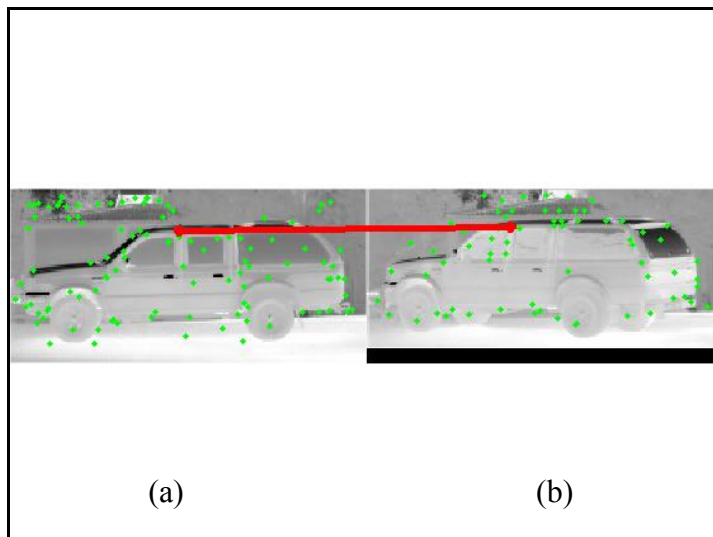


Figure 4-55 Matched Features of LWIR Band Vehicle Images-SIFT Detector.

(a) Reference image, not rotated, (b) Transformed image rotated approximately 5 degrees right

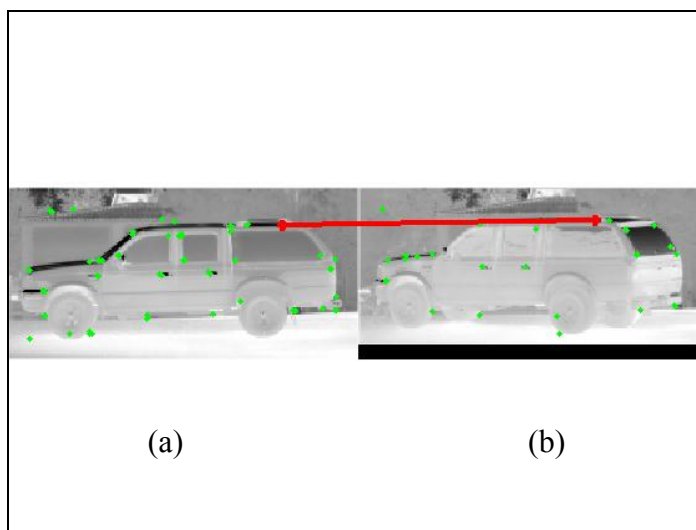


Figure 4-56 Matched Features of LWIR Band Vehicle Images-Harris Laplace Detector.

(a) Reference image, not rotated, (b) Transformed image rotated approximately 5 degrees right

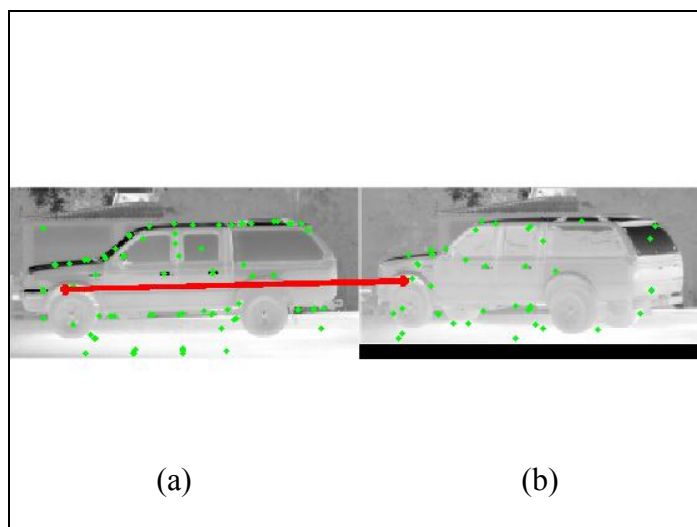


Figure 4-57 Matched Features of LWIR Band Vehicle Images-Hessian Laplace Detector.

(a) Reference image, not rotated, (b) Transformed image rotated approximately 5 degrees right

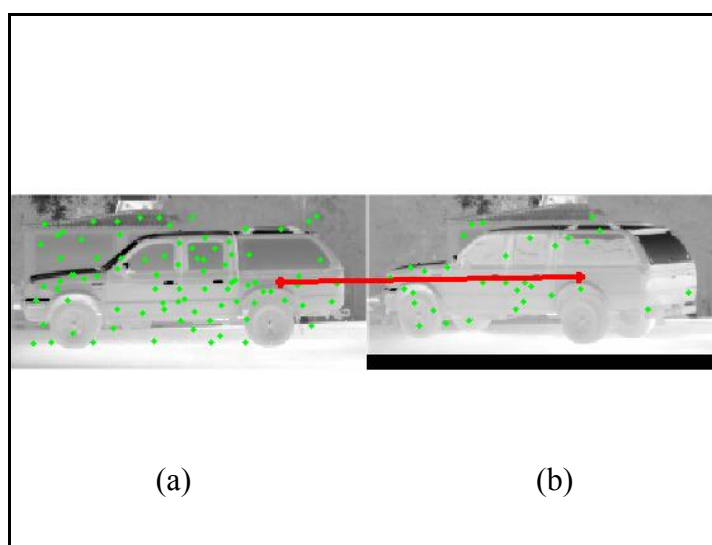


Figure 4-58 Matched Features of LWIR Band Vehicle Images-SURF Detector.

(a) Reference image, not rotated, (b) Transformed image rotated approximately 5 degrees right

Matched features of LWIR-band and visual-band vehicle images comparing the reference images, and the transformed images with vehicle approximately 45 degrees rotated, are not presented. Moreover, matched points for visual-band vehicle images comparing the reference images and the transformed images, approximately 5 degrees rotated, are not given. Because, all of the matched features are similar.

Matched features outside the vehicle region of interest are not considered for matching accuracy calculations. The graphical results showing the change of matching accuracy with respect to rotation angle are presented in Figure 4-59 and Figure 4-60 below, for LWIR and visual-bands respectively.

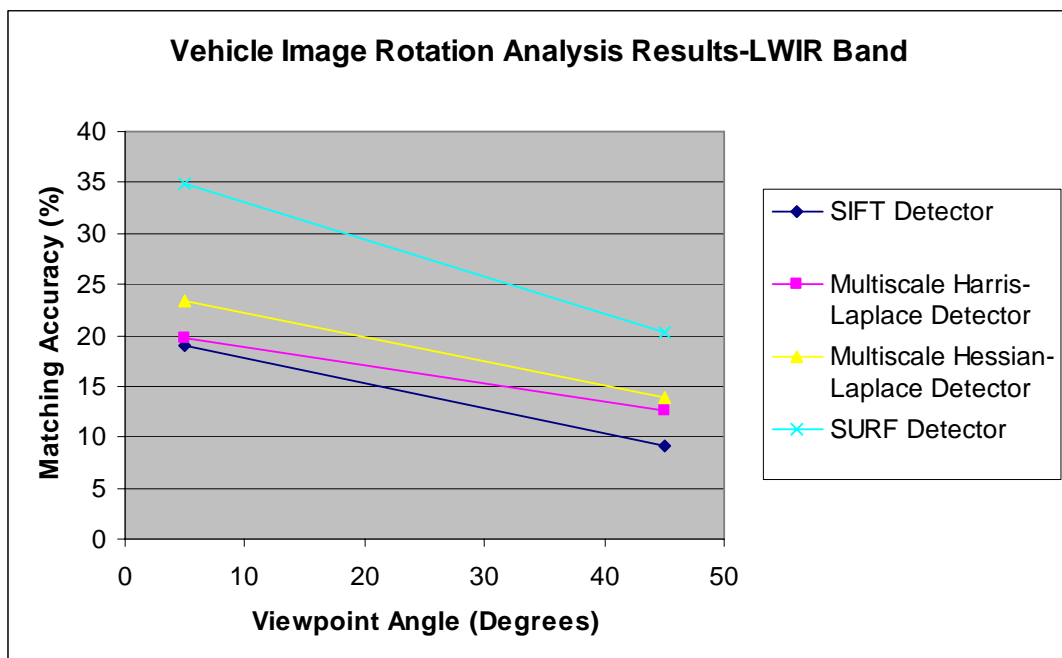


Figure 4-59 Matching Accuracy of Vehicle Images with Changing Rotation Angle-
LWIR Band

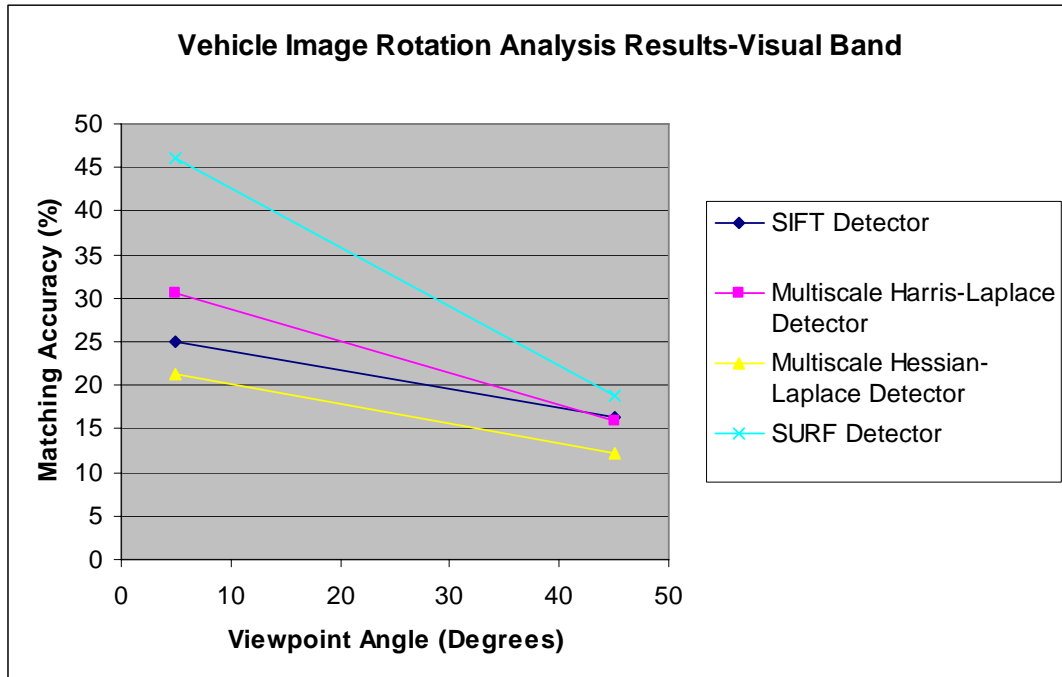


Figure 4-60 Matching Accuracy of Vehicle Images with Changing Rotation Angle-Visual Band

Matching accuracy of vehicle images decrease when rotation angle increases from 5 degrees to 45 degrees, for both LWIR and visual-bands, for all detectors utilized. Moreover, SURF detector performance in terms of matching accuracy for vehicle images is better in LWIR and visual-bands, as rotation angle changes, when compared with other feature detectors.

The discussion of these results is given in Conclusions chapter of this study. The experiments that are performed for the evaluation of matching performance of each feature detector in all of the bands (LWIR, MWIR, and Visual) with changing rotation of the mask in front of a blackbody and this experiment's results are discussed in the next section.

4.4 Matching Performance for Different Rotation Angles

Matching performance with various rotations experiment is performed in order to understand the effect of rotation angle of the mask on the matching accuracy in each of the corresponding band with all feature detectors. Mask is used in front of a blackbody in order to create sufficient number of features, discussed in the next section in detail.

4.4.1 Experimental Setup

Experimental studies for this experiment are performed by capturing images in all of the bands, namely mid-wave infrared-band (MWIR), long-wave infrared-band (LWIR), and visual-band.

Two infrared cameras (MWIR camera and LWIR camera) and an off-the-shelf Cannon Powershot A75 visual digital camera are used in these controlled experiments. Infrared cameras are already described in Section 4.1. 100mm focal length lenses of MWIR and LWIR cameras are used. The effect of rotation angle of the mask on matching accuracy is investigated at each different distance of measurement, so images are captured from 8m, 15m, 22.5m, 30m, 35m, and 43m distances. All of the three cameras are located on a turntable standing above the tripod, so that the focal plane of each of the three cameras is closer to each other for the experiment.

Infrared images of a blackbody device are captured. This blackbody device is different from the one used in previous experiments. It is 50cm by 50cm in size and it provides uniform heat distribution over its surface at a constant (set) temperature with good accuracy and stability during its operation. Such blackbody devices are used as reference sources for calibrating infrared cameras. Temperature of the blackbody is fixed and set at 150°C for all of the measurements. The photos are captured after blackbody is heated and its temperature is stabilized.

Mask built from white color cartoon is used during these experiments. Several shapes such as rectangle, pentagon, hexagon and square, are formed by cutting the surface of the mask. These shapes are formed in order to obtain enough number of distinctive feature points in the analysis of the images. Mask is rotated 5 degrees and 45 degrees to the right at each distance of measurement. 5 degrees and 45 degrees of rotation of the mask is performed in order to obtain a transformed or test image for the comparison of matching performance between reference images where mask is not rotated. The schema of the mask is shown in Figure 4-61 below.

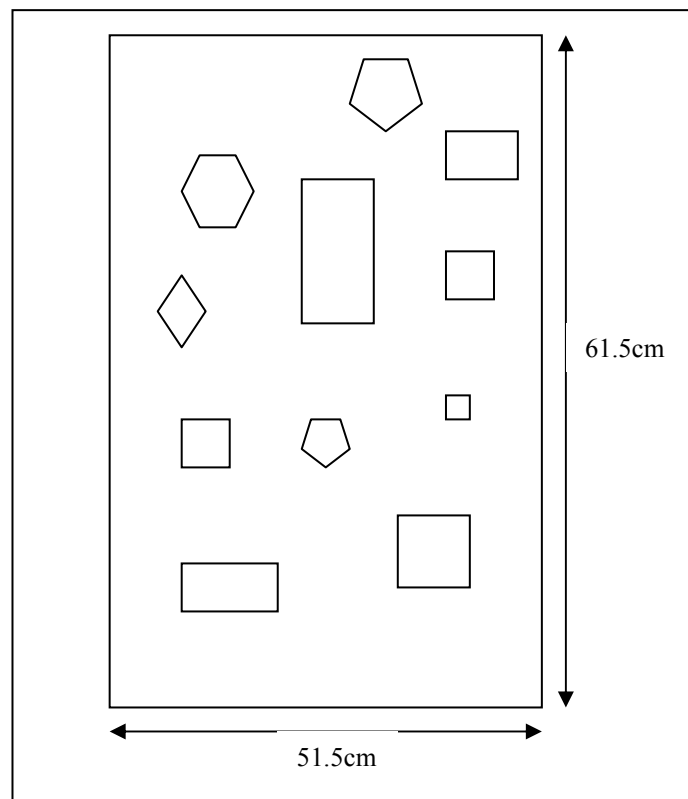


Figure 4-61 White Cartoon Mask of Experiment 3

The heating surface of the blackbody is black in color. The blackbody device (without mask) used in this experiment is shown in Figure 4-62 below. The device

with mask is shown in Figure 4-63 below. These visual-band images are not used in the analysis of this experiment. They are presented for visualizing the experimental setup only.

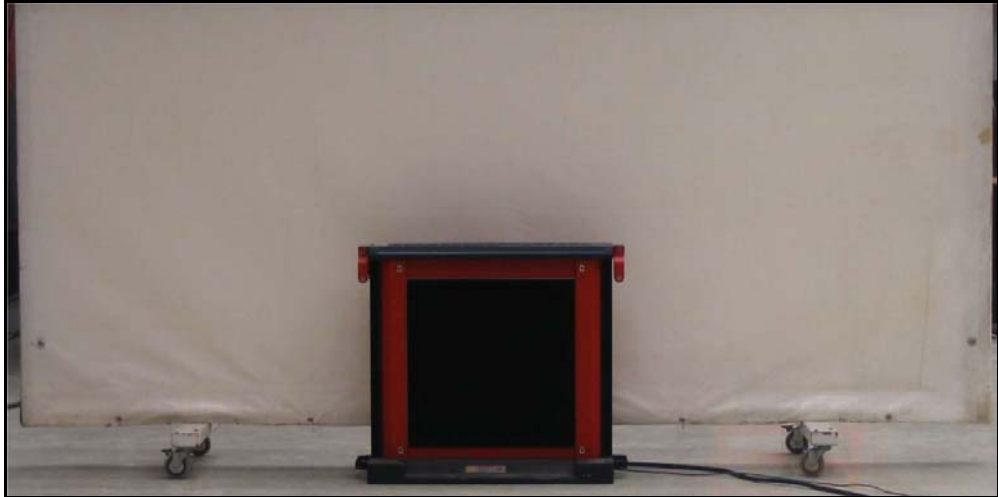


Figure 4-62 Blackbody Device without Mask (Experiment 3)

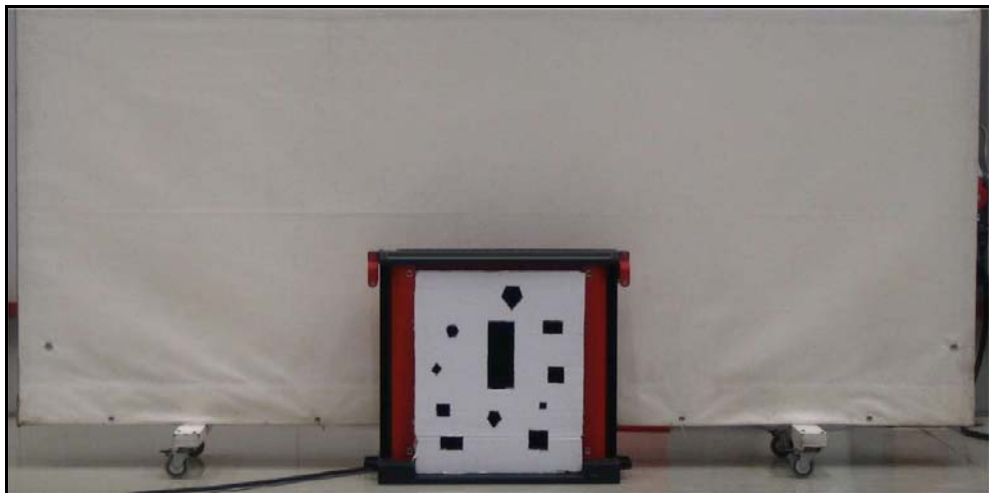


Figure 4-63 Blackbody Device with Mask (Experiment 3)

The typical infrared images obtained are presented in Figure 4-64 below. Note that the hot regions in the images are in white color, whereas colder regions relative to the hotter areas are in dark color. The images used in the analysis of this experiment are presented later in this section.

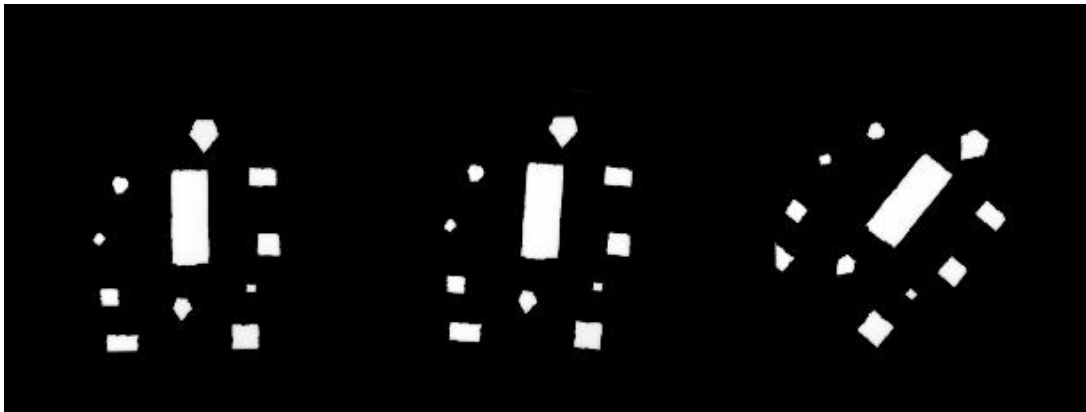


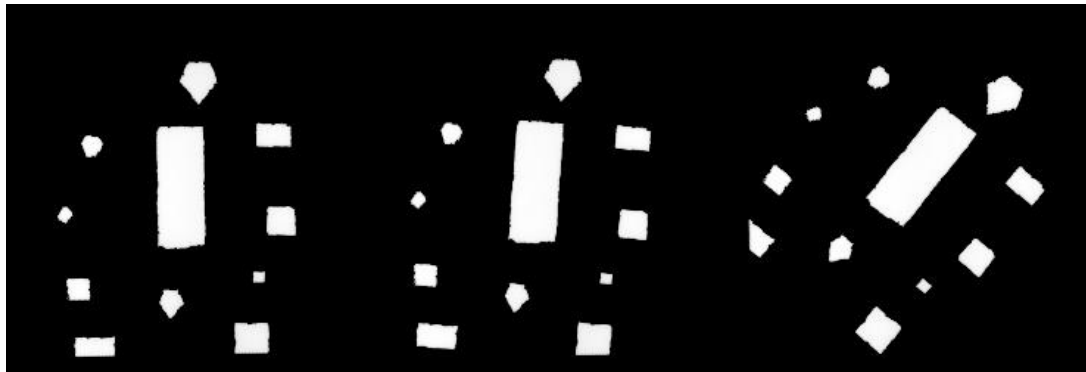
Figure 4-64 Typical Infrared Images of Blackbody Device with Mask
(Experiment 3)

The spectral band pass filters used for each infrared camera are listed below.

1. MWIR Camera, 2.8-3.5 μm bandpass filter
2. LWIR Camera, 8.6-9.3 μm bandpass filter

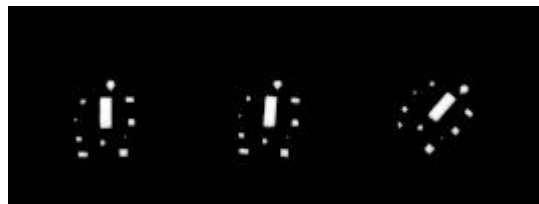
After stabilization of the blackbody device, LWIR camera integration time is set at 162 μs and MWIR camera integration time is set at 550 μs in order to have images with comparable digital levels. The equivalence of integration times of LWIR and MWIR cameras is not important in this experiment, since the matching performance between LWIR and MWIR camera images is not investigated.

Infrared images captured with mask rotations of 5 degrees and 45 degrees in LWIR, MWIR, and visual-bands are presented below. The images obtained at 8m and 43m distances are shown for comparison.



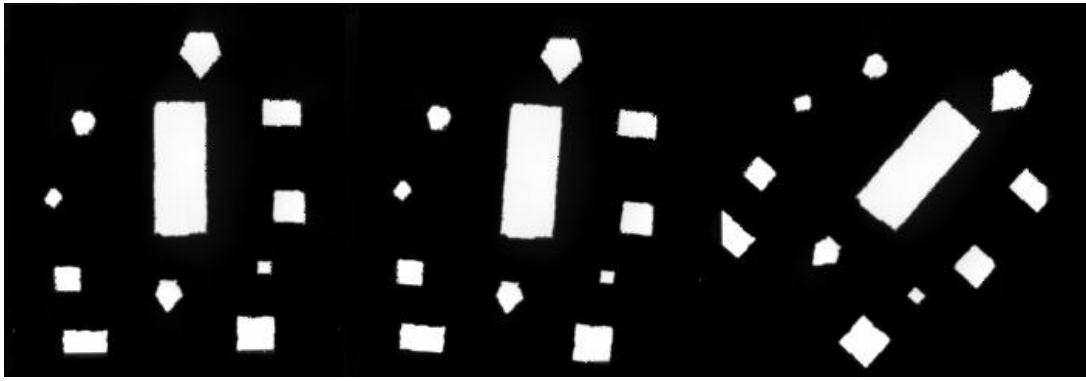
(a) (b) (c)
Figure 4-65 LWIR Band Infrared Images with 150°C Blackbody at 8m.

(a) Reference image, mask not rotated, (b) Mask rotated 5 degrees to right, (c) Mask rotated 45 degrees to right



(a) (b) (c)
Figure 4-66 LWIR Band Infrared Images with 150°C Blackbody at 43m.

(a) Reference image, mask not rotated, (b) Mask rotated 5 degrees to right, (c) Mask rotated 45 degrees to right



(a)

(b)

(c)

Figure 4-67 MWIR Band Infrared Images with 150°C Blackbody at 8m.

(a) Reference image, mask not rotated, (b) Mask rotated 5 degrees to right, (c) Mask rotated 45 degrees to right



(a)

(b)

(c)

Figure 4-68 MWIR Band Infrared Images with 150°C Blackbody at 43m.

(a) Reference image, mask not rotated, (b) Mask rotated 5 degrees to right, (c) Mask rotated 45 degrees to right

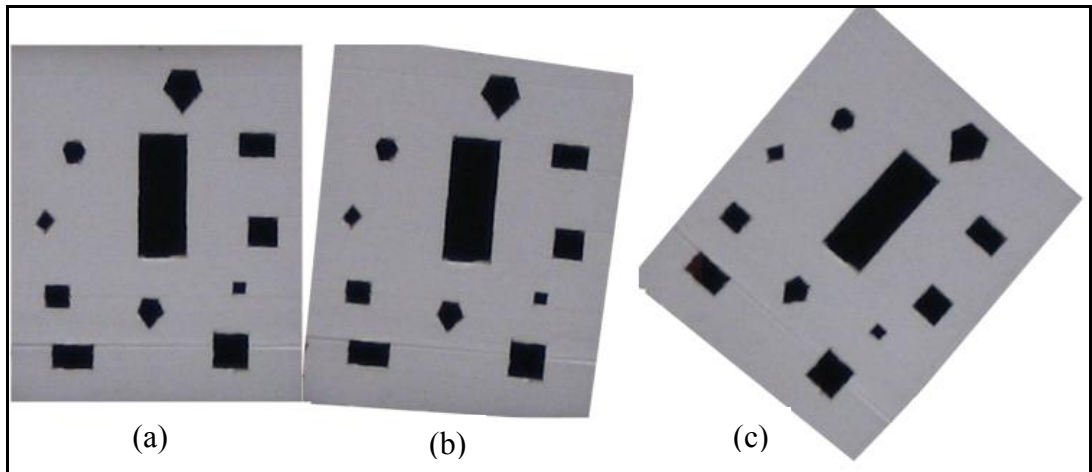


Figure 4-69 Visual Band Infrared Images with 150°C Blackbody at 8m.

(a) Reference image, mask not rotated, (b) Mask rotated 5 degrees to right, (c) Mask rotated 45 degrees to right

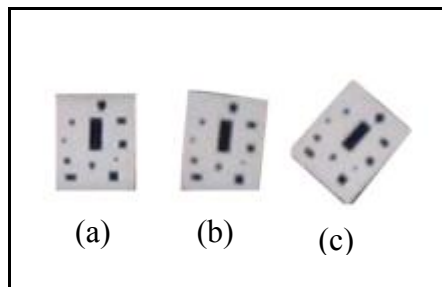


Figure 4-70 Visual Band Infrared Images with 150°C Blackbody at 43m.

(a) Reference image, mask not rotated, (b) Mask rotated 5 degrees to right, (c) Mask rotated 45 degrees to right

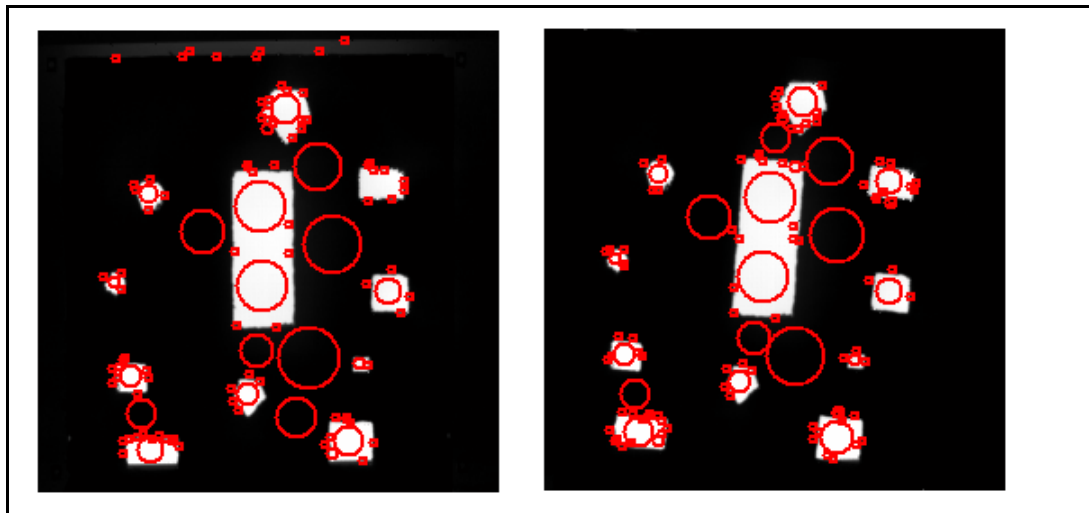
Each reference image for corresponding band and distance is compared with the transformed images, either of the one which is the image with mask 5 degrees rotated to the right or the other which is the image with mask 45 degrees rotated to

the right. The analyzed images at 15m, 22.5m, 30m, and 35m distances are not presented, due to space limitations.

The results of this experiment are presented in the next section.

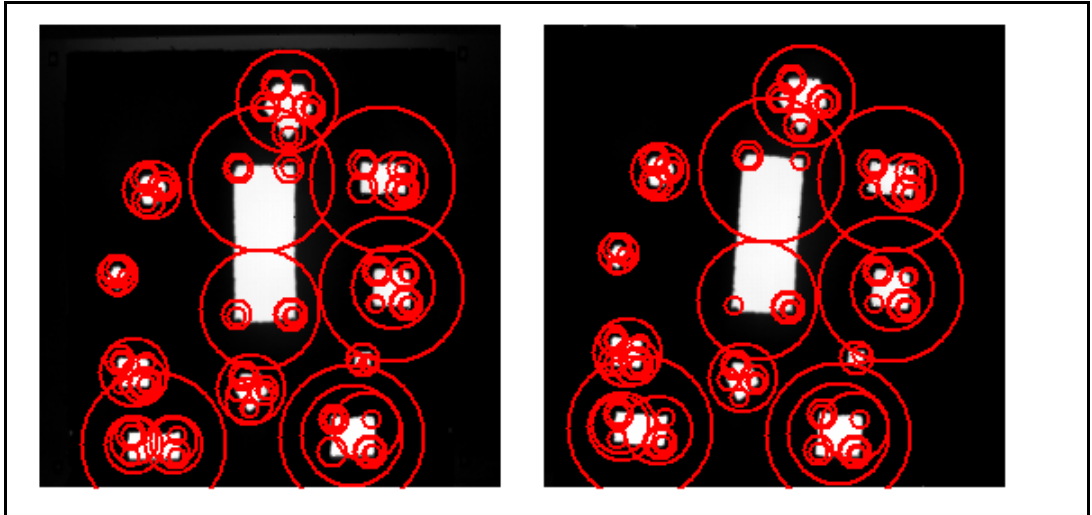
4.4.2 Results of Matching Performance with Different Rotations Experiment

Feature points of MWIR infrared image pairs belonging to 150°C blackbody with mask at 8m distance are presented in Figure 4-71, Figure 4-72, Figure 4-73, and Figure 4-74 below, extracted by SIFT, Harris-Laplace, Hessian-Laplace and SURF detectors, respectively.



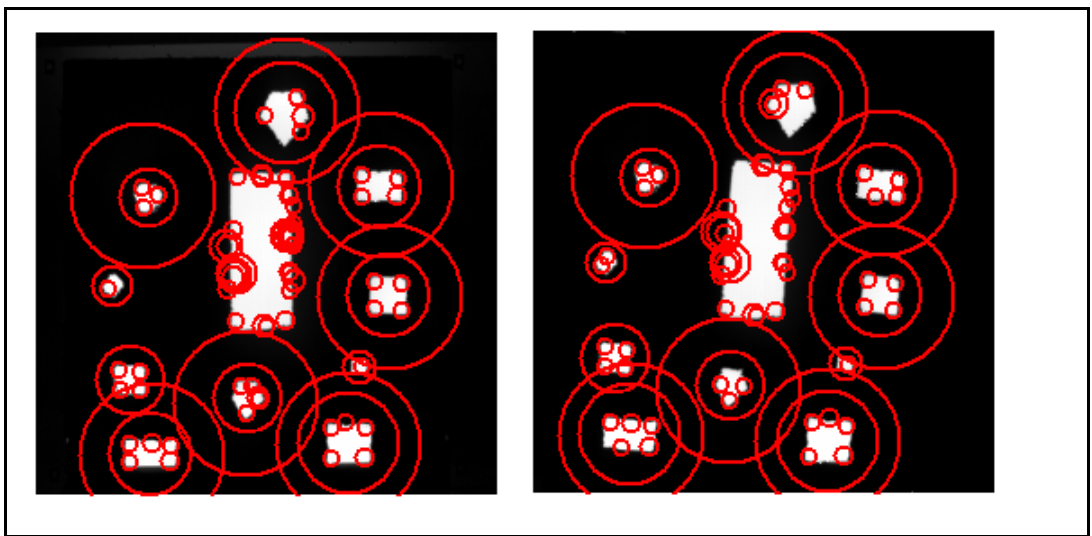
(a) (b)
Figure 4-71 MWIR Band Image Feature Points Extracted by SIFT Detector, 150°C Blackbody, 8m.

(a) Reference image, mask not rotated, (b) Mask rotated 5 degrees to right



(a) (b)
 Figure 4-72 MWIR Band Image Feature Points Extracted by Harris-Laplace
 Detector, 150°C Blackbody, 8m.

(a) Reference image, mask not rotated, (b) Mask rotated 5 degrees to right



(a) (b)
 Figure 4-73 MWIR Band Image Feature Points Extracted by Hessian-Laplace
 Detector, 150°C Blackbody, 8m.

(a) Reference image, mask not rotated, (b) Mask rotated 5 degrees to right

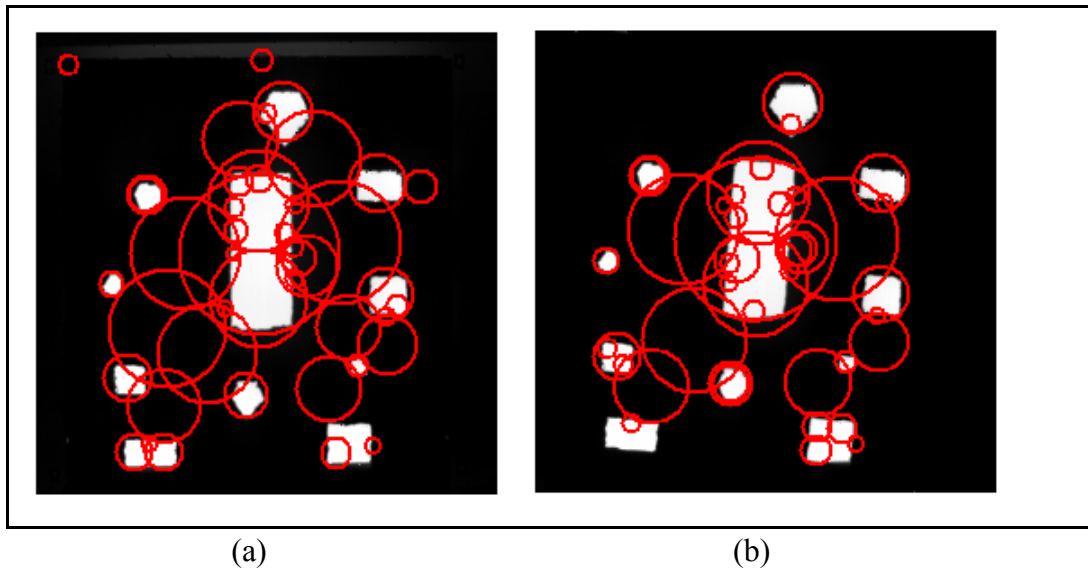


Figure 4-74 MWIR Band Image Feature Points Extracted by SURF Detector, 150°C Blackbody, 8m.

(a) Reference image, mask not rotated, (b) Mask rotated 5 degrees to right

The extracted features of LWIR-band images are similar to the figures of MWIR 150°C images shown above, so they are not presented. SIFT, Hessian-Laplace and SURF detectors detect blob-like regions accurately, as it can be observed from the figures. Harris-Laplace is strong for corner detection at different scales.

Feature points of visual-band image pairs belonging to 150°C blackbody with mask at 8m distance, are presented in Figure 4-75, Figure 4-76, Figure 4-77, and Figure 4-78 below, extracted by SIFT, Harris-Laplace, Hessian-Laplace and SURF detectors, respectively.

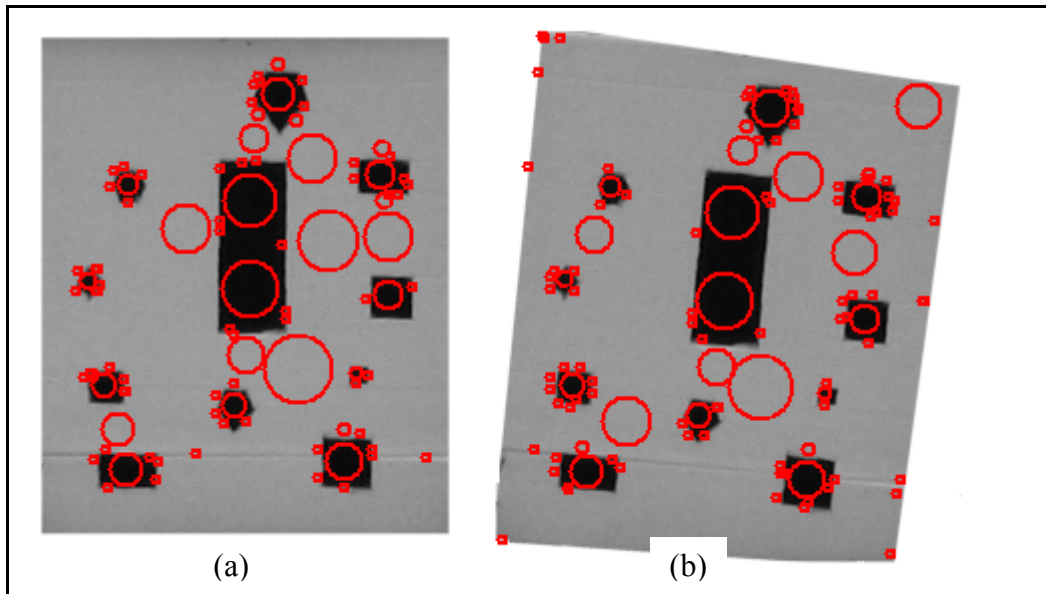


Figure 4-75 Visual Band Image Feature Points Extracted by SIFT Detector, 150°C
Blackbody, 8m.

(a) Reference image, mask not rotated, (b) Mask rotated 5 degrees to right

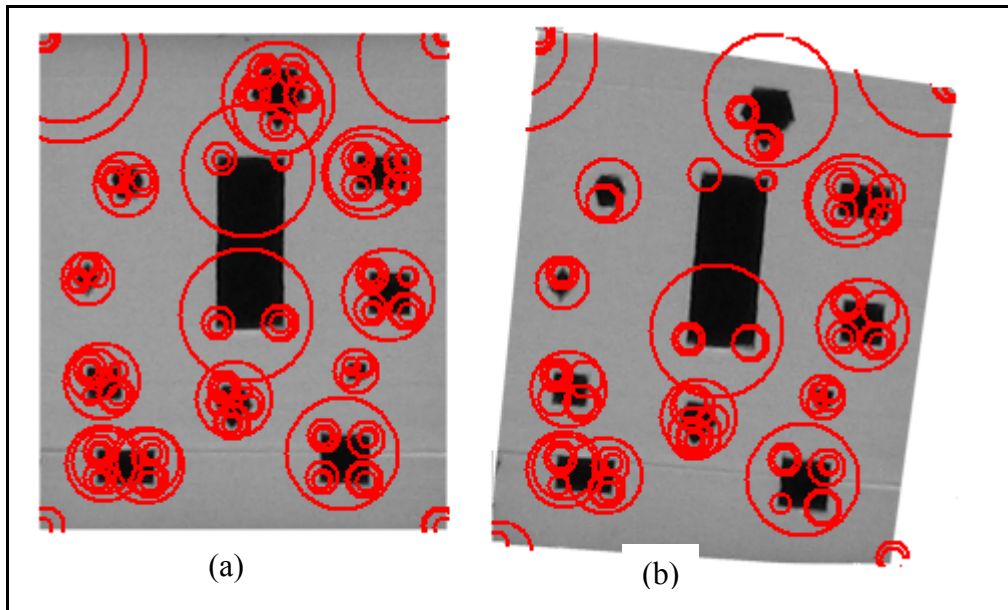


Figure 4-76 Visual Band Image Feature Points Extracted by Harris-Laplace Detector, 150°C Blackbody, 8m.

(a) Reference image, mask not rotated, (b) Mask rotated 5 degrees to right

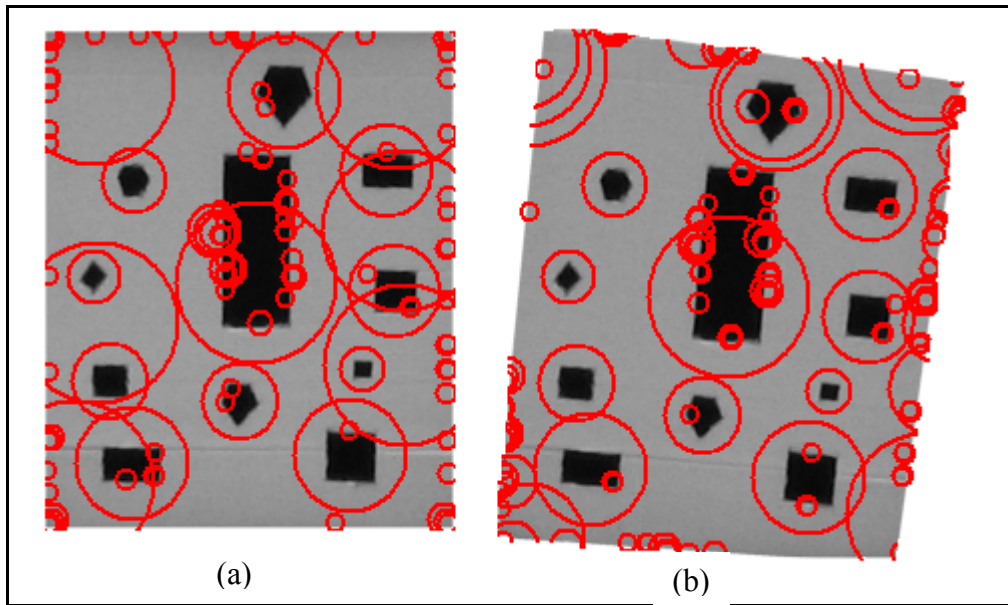


Figure 4-77 Visual Band Image Feature Points Extracted by Hessian-Laplace Detector, 150°C Blackbody, 8m.

(a) Reference image, mask not rotated, (b) Mask rotated 5 degrees to right

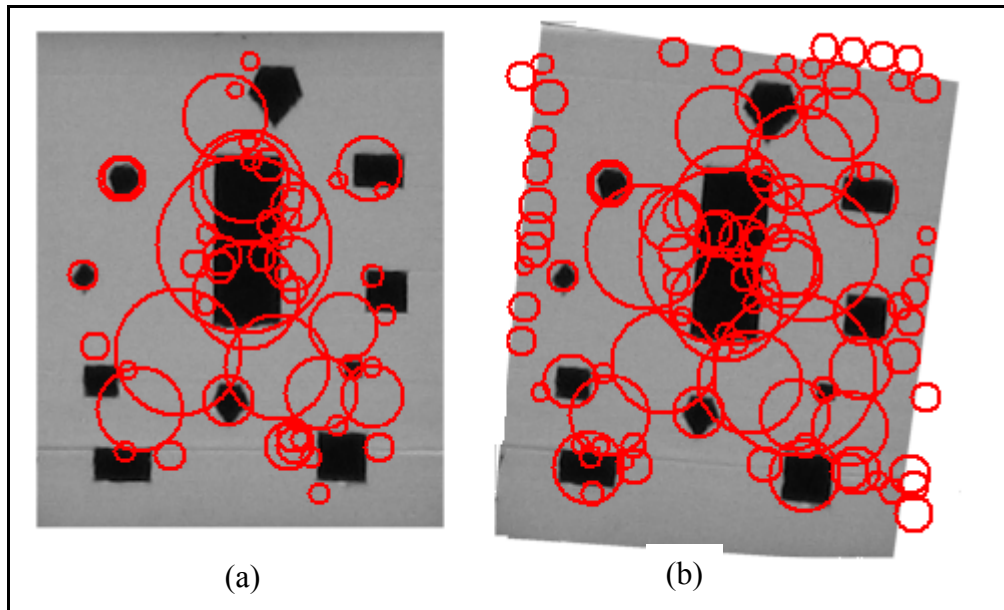


Figure 4-78 Visual Band Image Feature Points Extracted by SURF Detector, 150°C Blackbody, 8m.

(a) Reference image, mask not rotated, (b) Mask rotated 5 degrees to right

Black and white color contrast is effective in visual-band feature detection. It can be observed that SIFT, Hessian-Laplace and SURF feature detectors are effective in finding round, curved structures and blobs. Harris-Laplace feature detector is good at finding corners.

Feature regions of the 45 degree rotated transformed images are not presented for all the bands due to the space limitations.

After feature point extraction of reference and transformed images, descriptor vectors are calculated and matches between two images are obtained. Matched feature points for MWIR 150°C blackbody image pairs is shown in Figure 4-79, Figure 4-80, Figure 4-81, and Figure 4-82 below, for each feature detector.

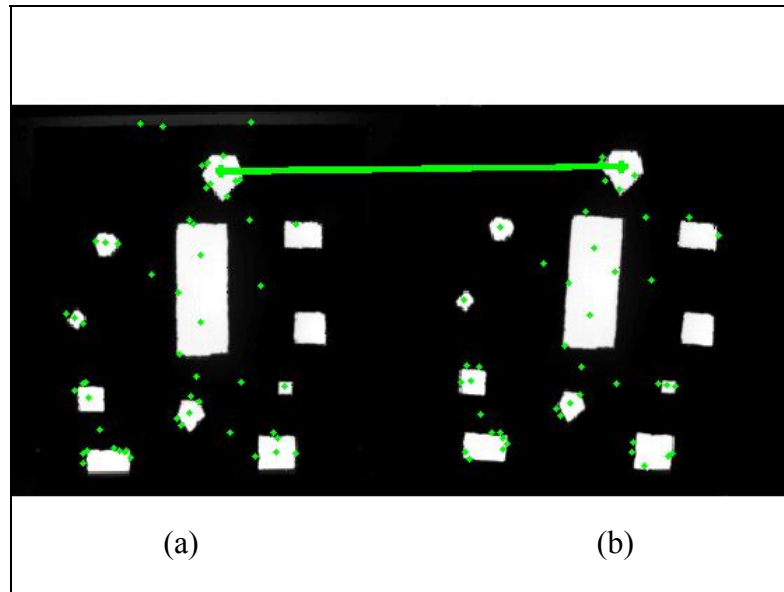


Figure 4-79 Matched Feature Points of MWIR Band Images, SIFT detector, 150°C
Blackbody, 8m.

(a) Reference image, mask not rotated, (b) Mask rotated 5 degrees to right

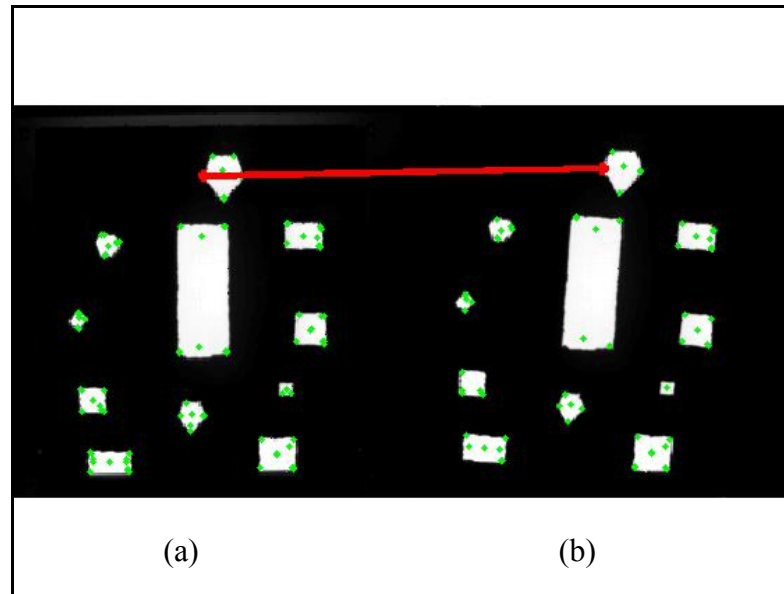


Figure 4-80 Matched Feature Points of MWIR Band Images, Harris-Laplace Detector, 150°C Blackbody, 8m.

(a) Reference image, mask not rotated, (b) Mask rotated 5 degrees to right

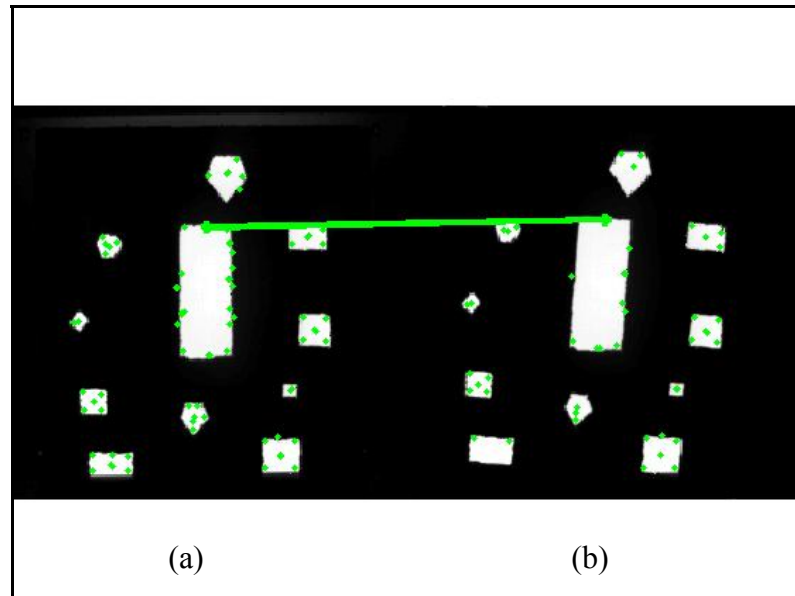


Figure 4-81 Matched Feature Points of MWIR Band Images, Hessian-Laplace Detector, 150°C Blackbody, 8m.

(a) Reference image, mask not rotated, (b) Mask rotated 5 degrees to right

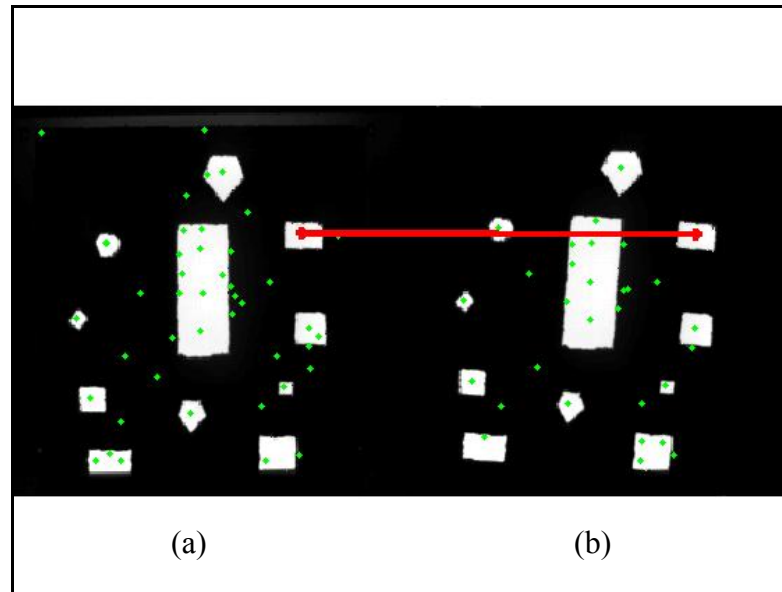


Figure 4-82 Matched Feature Points of MWIR Band Images, SURF Detector, 150°C Blackbody, 8m.

(a) Reference image, mask not rotated, (b) Mask rotated 5 degrees to right

Matched feature points, outside the region of interest, where rectangular, pentagon, hexagon and square shapes exist, are not used in the calculation of matching accuracy. The effect of rotations of the mask on the matching accuracy is presented in Figure 4-83, Figure 4-84, and Figure 4-85 below, for SIFT detector in MWIR, LWIR and visual-bands, respectively.

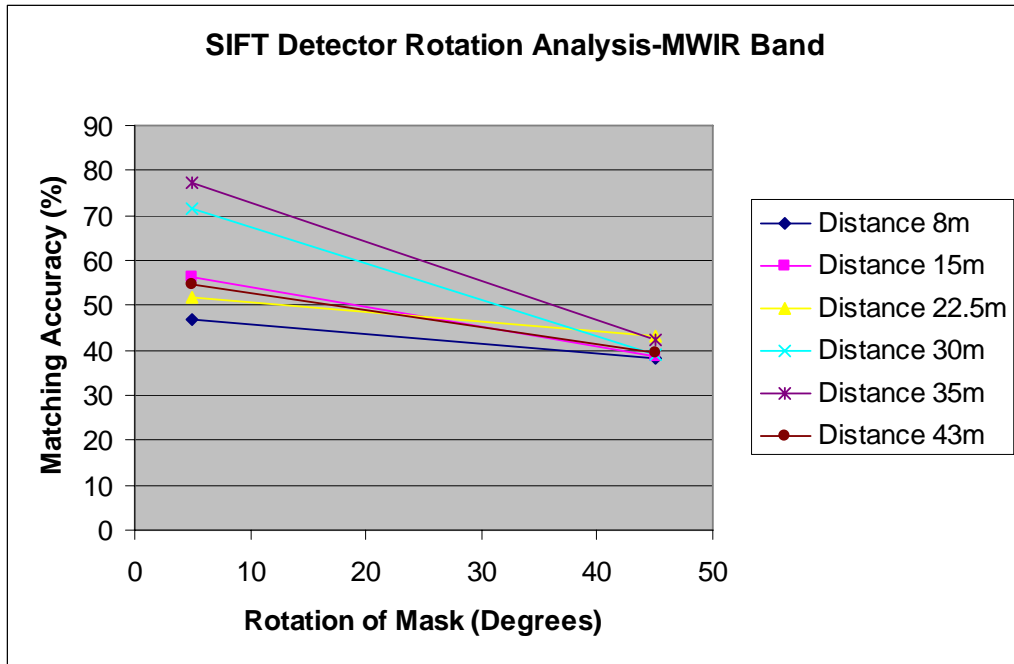


Figure 4-83 Matching Accuracy with Changing Rotation : MWIR Band, SIFT

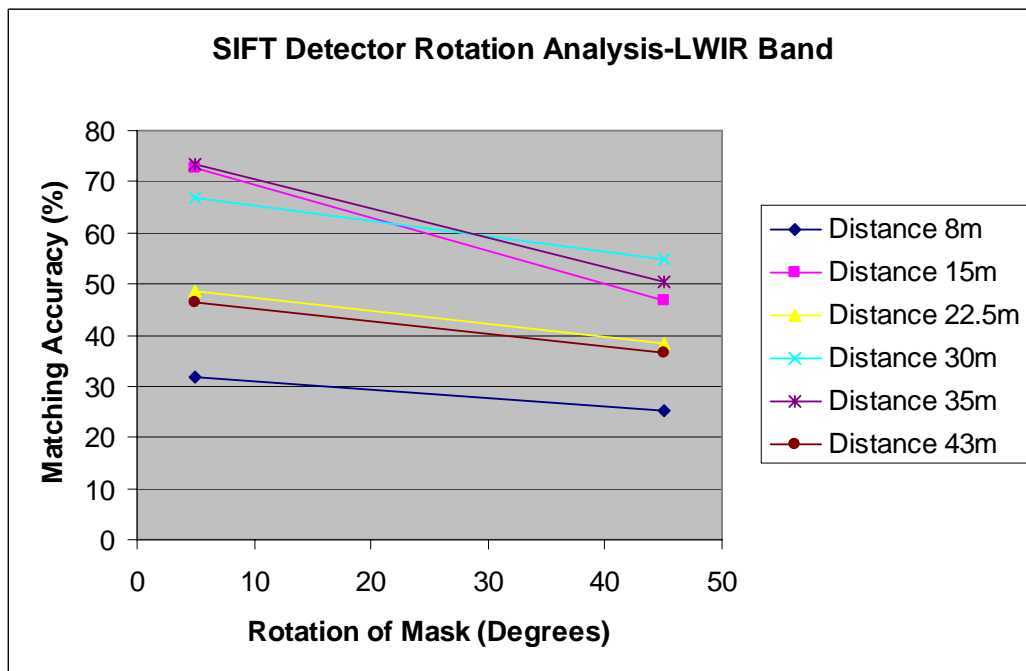


Figure 4-84 Matching Accuracy with Changing Rotation : LWIR Band, SIFT

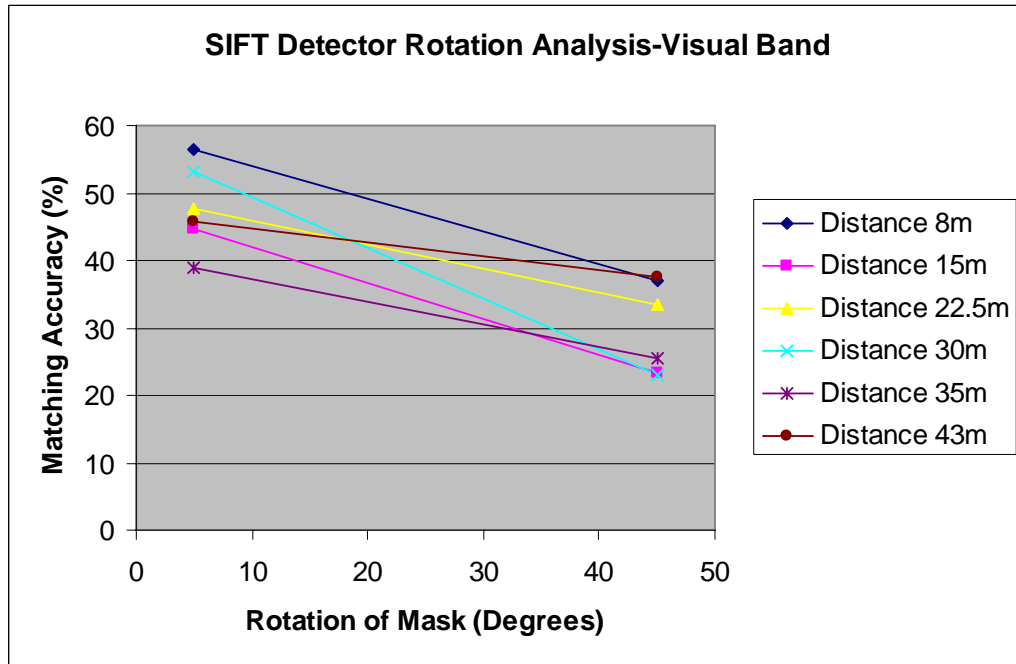


Figure 4-85 Matching Accuracy with Changing Rotation : Visual Band, SIFT

The matching accuracy percentage values obtained for SIFT detector are summarized below, for each distance and band combinations.

Table 4-2 Matching Accuracy Values, SIFT Detector, Experiment 3

Matching Accuracy	Rotation Angles (degrees)	Distance (m)					
		8	15	22.5	30	35	43
MWIR Band	5	47	56	52	71	77	55
	45	38	39	43	39	42	39
LWIR Band	5	32	73	49	67	73	46
	45	25	47	38	55	51	37
Visual Band	5	56	45	48	53	39	46
	45	37	23	33	23	25	38

Effect of rotations of the mask on the matching accuracy is presented in Figure 4-86, Figure 4-87, and Figure 4-88 below for Harris-Laplace detector in MWIR, LWIR and visual-bands, respectively.

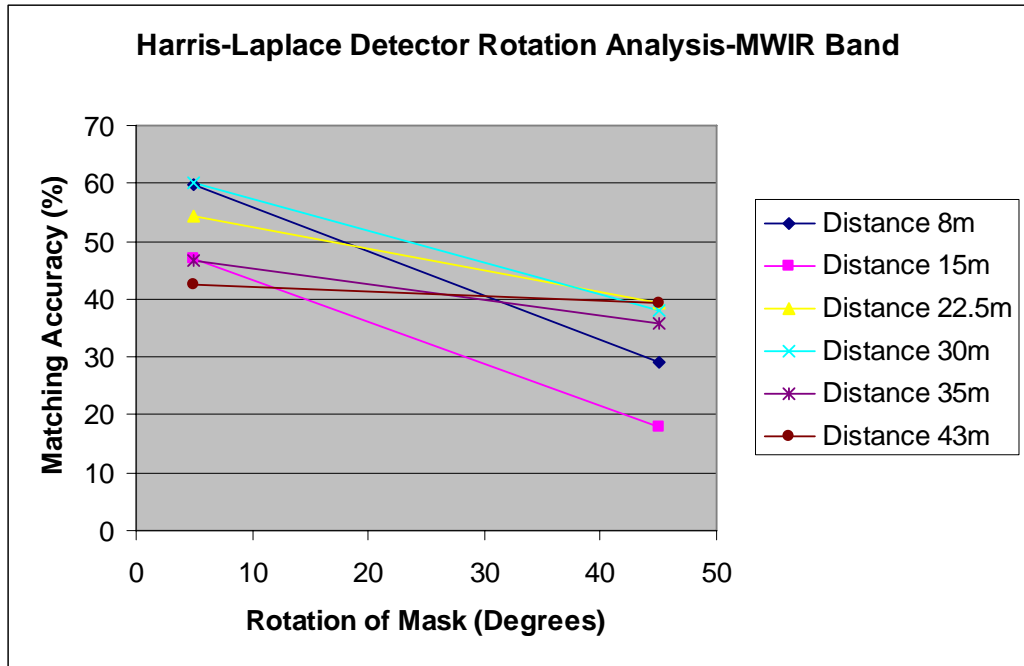


Figure 4-86 Matching Accuracy with Changing Rotation : MWIR Band, Harris-Laplace

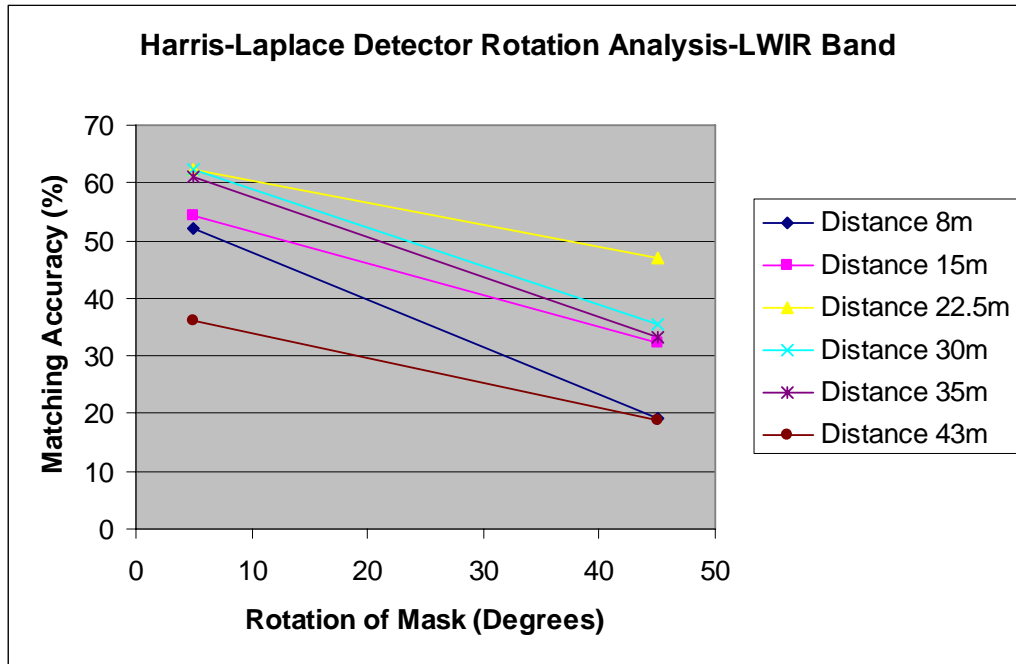


Figure 4-87 Matching Accuracy with Changing Rotation : LWIR Band, Harris-Laplace

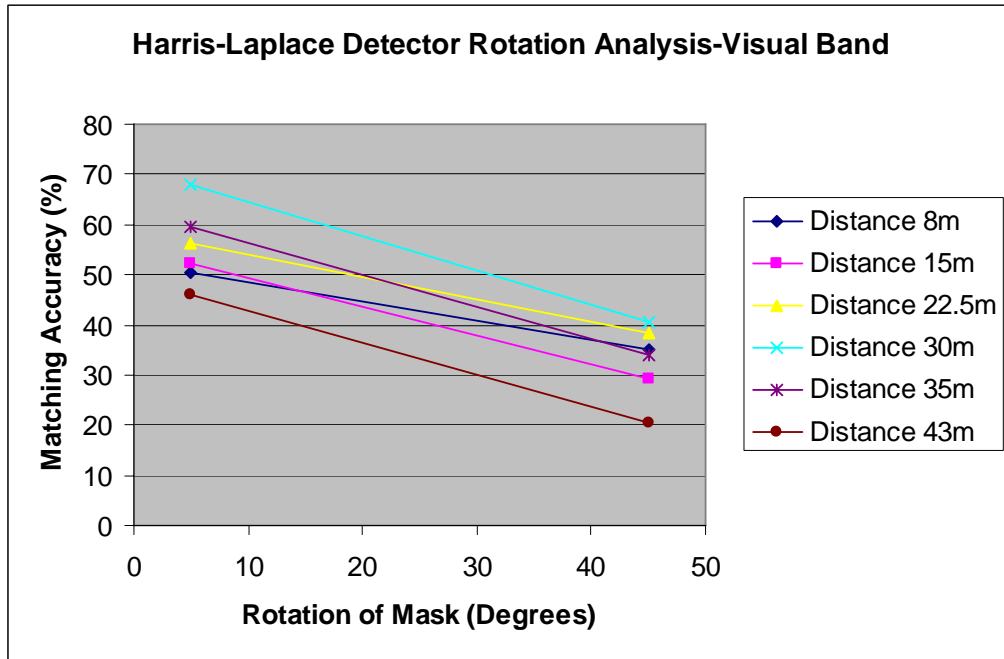


Figure 4-88 Matching Accuracy with Changing Rotation : Visual Band, Harris-Laplace

Matching accuracy percentage values obtained for Harris-Laplace detector are summarized below, for each distance and band combinations.

Table 4-3 Matching Accuracy Values, Harris-Laplace Detector, Experiment 3

Matching Accuracy	Rotation Angles (degrees)	Distance (m)					
		8	15	22.5	30	35	43
MWIR Band	5	60	47	54	60	47	43
	45	29	18	39	38	36	39
LWIR Band	5	52	54	62	62	61	36
	45	19	32	47	35	33	19
Visual Band	5	50	52	56	68	60	46
	45	35	29	38	41	34	20

Effect of rotations of the mask on the matching accuracy is presented in Figure 4-89, Figure 4-90, and Figure 4-91 below for Hessian-Laplace detector in MWIR, LWIR and visual-bands, respectively.

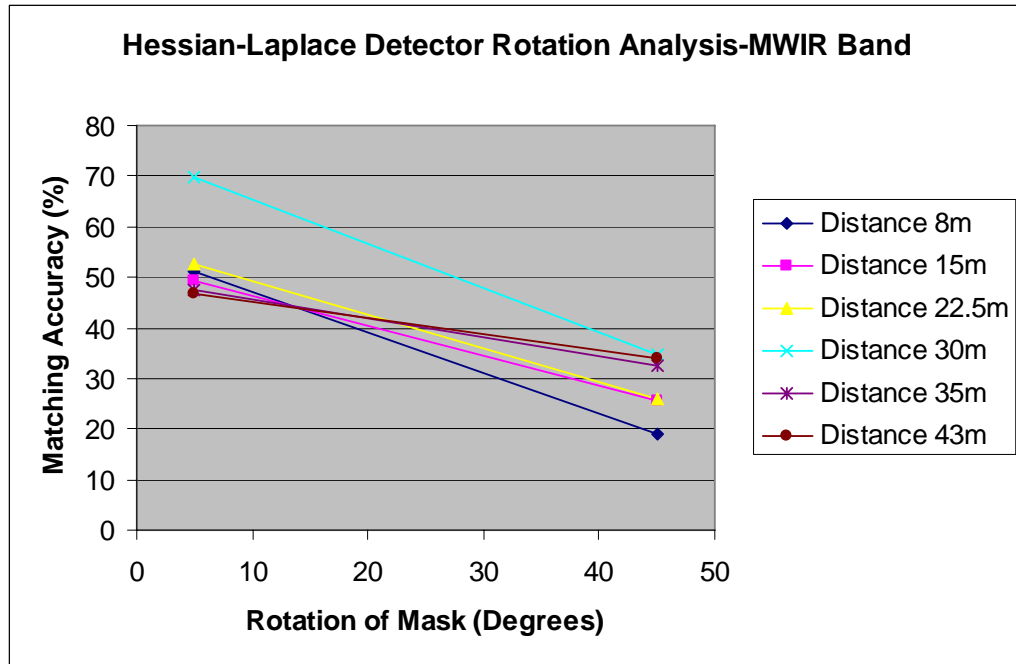


Figure 4-89 Matching Accuracy with Changing Rotation : MWIR Band, Hessian-Laplace

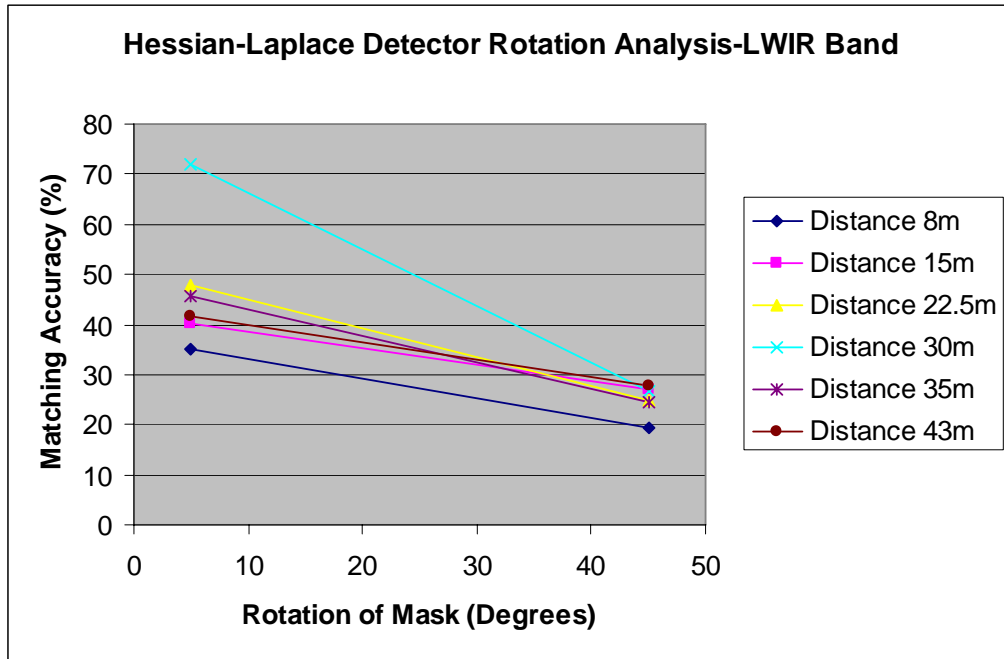


Figure 4-90 Matching Accuracy with Changing Rotation : LWIR Band, Hessian-Laplace

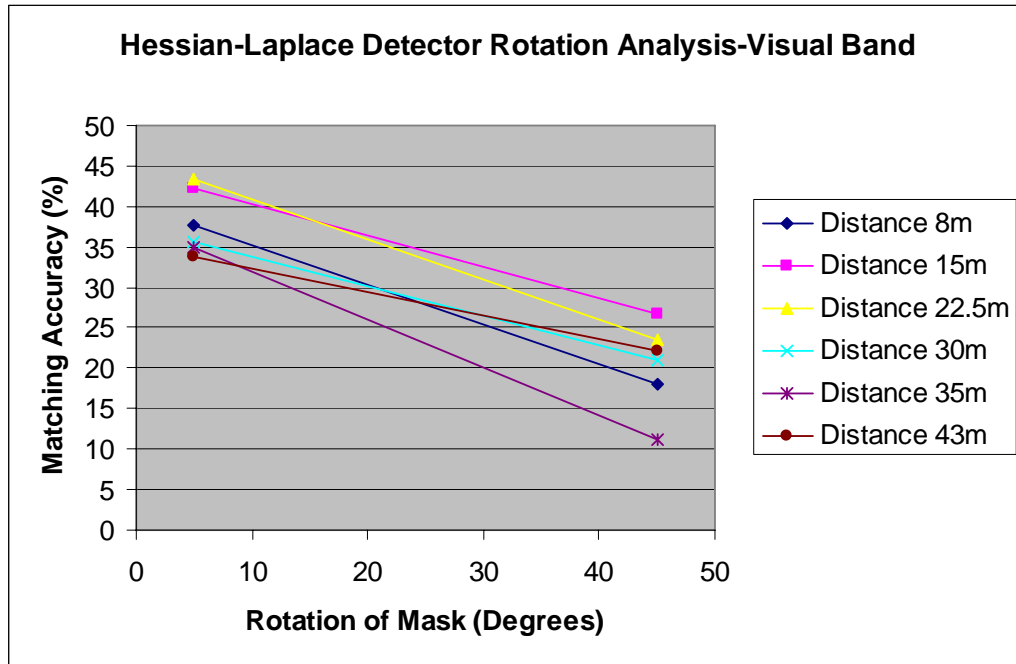


Figure 4-91 Matching Accuracy with Changing Rotation : Visual Band, Hessian-Laplace

Matching accuracy percentage values obtained for Hessian-Laplace detector are summarized below, for each distance and band combinations.

Table 4-4 Matching Accuracy Values, Hessian-Laplace Detector, Experiment 3

Matching Accuracy	Rotation Angles (degrees)	Distance (m)					
		8	15	22.5	30	35	43
MWIR Band	5	51	49	53	70	48	47
	45	19	26	26	35	33	34
LWIR Band	5	35	40	48	72	46	42
	45	19	27	25	26	25	28
Visual Band	5	38	42	43	36	35	34
	45	18	27	24	21	11	22

Effect of rotations of the mask on the matching accuracy is presented in Figure 4-92, Figure 4-93, and Figure 4-94 below for SURF detector in MWIR, LWIR and visual-bands, respectively.

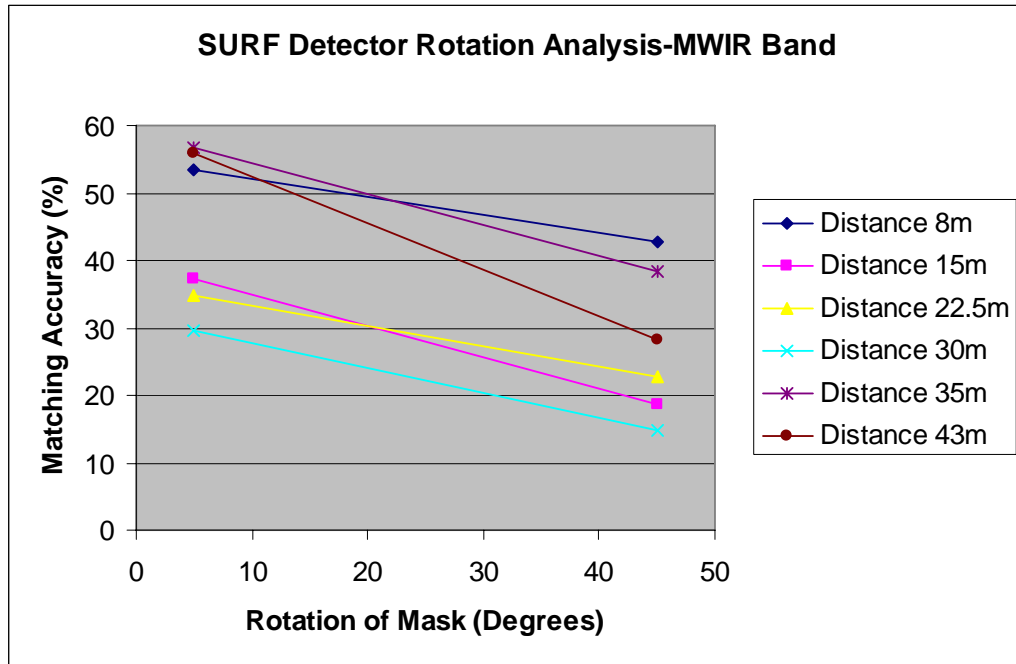


Figure 4-92 Matching Accuracy with Changing Rotation : MWIR Band, SURF

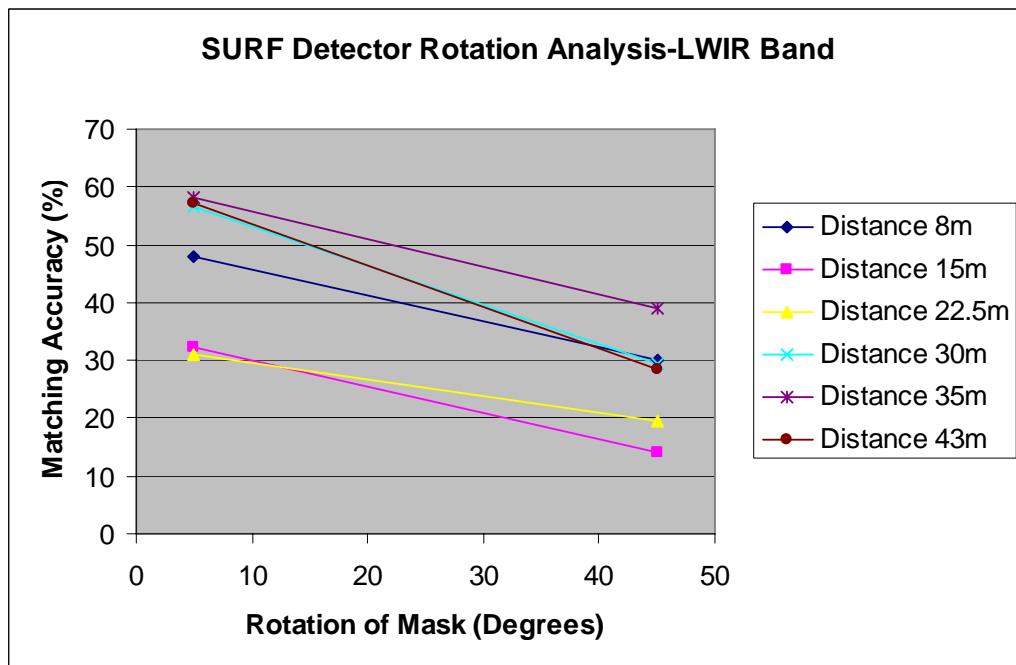


Figure 4-93 Matching Accuracy with Changing Rotation : LWIR Band, SURF

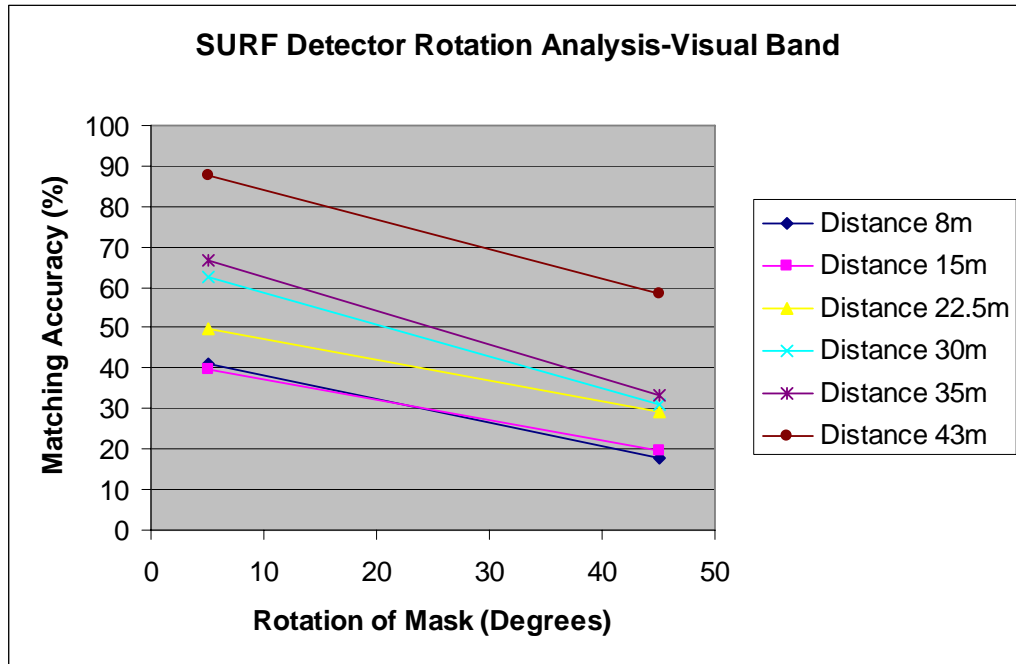


Figure 4-94 Matching Accuracy with Changing Rotation : Visual Band, SURF

Matching accuracy percentage values obtained for SURF detector are summarized below, for each distance and band combinations.

Table 4-5 Matching Accuracy Values, SURF Detector, Experiment 3

Matching Accuracy	Rotation Angles (degrees)	Distance (m)					
		8	15	22.5	30	35	43
MWIR Band	5	53	37	35	30	57	56
	45	43	19	23	15	38	28
LWIR Band	5	48	32	31	57	58	57
	45	30	14	20	29	39	29
Visual Band	5	41	40	50	63	67	88
	45	18	20	29	31	33	58

Matching accuracy is higher when rotation angle is lowest at 5 degrees at all distances in all of three bands for SIFT, Harris-Laplace, Hessian-Laplace, and SURF detectors. As rotation angle increases, matching accuracy decreases.

4.5 Matching Performance with Changing Distance Experiment

Matching performance with changing distance experiment is performed in order to understand the effect of distance of the blackbody device with mask on the matching accuracy in each of the corresponding band with all the feature detectors. This experiment also evaluates the scale invariance property of detectors as well.

4.5.1 Experimental Setup

Experimental studies for matching performance with changing distance experiment are performed by taking images in all three bands.

The experimental setup used in this experiment is the same with the previous one. The images are captured at 8m, 15m, 22.5m, 30m, and 43m distances. The images at 8m distances are taken as reference, whereas the ones at other distances are taken to be transformed images.

The obtained image pairs, that are analyzed, are presented in Figure 4-95, Figure 4-96, and Figure 4-97 below, for LWIR-band at 8m distance and 22.5m distance, MWIR-band at 8m distance and 22.5m distance, visual-band at 8m distance and 22.5m distance, respectively. These images are shown for comparison. The image pairs at other distances are not presented.

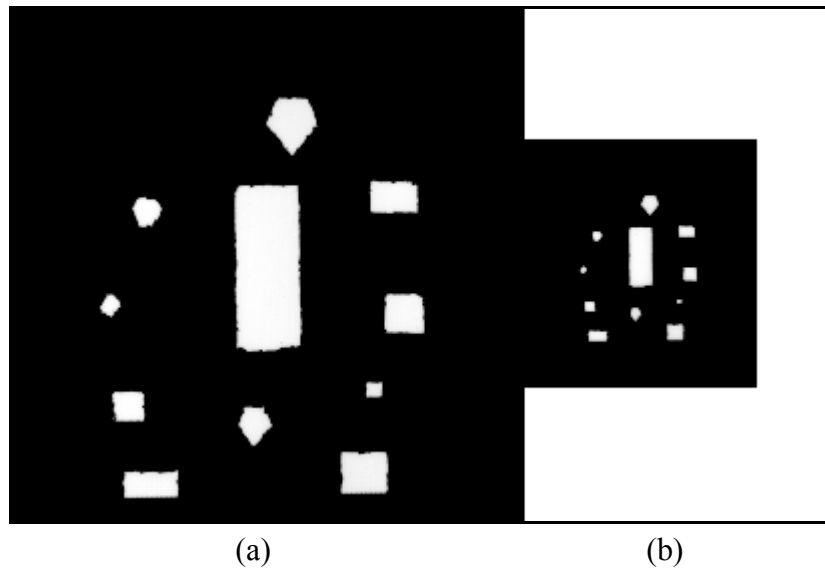


Figure 4-95 LWIR Band Images with 150°C Blackbody at 8m and 22.5m.

(a) Reference image at 8m, (b) Transformed image at 22.5m

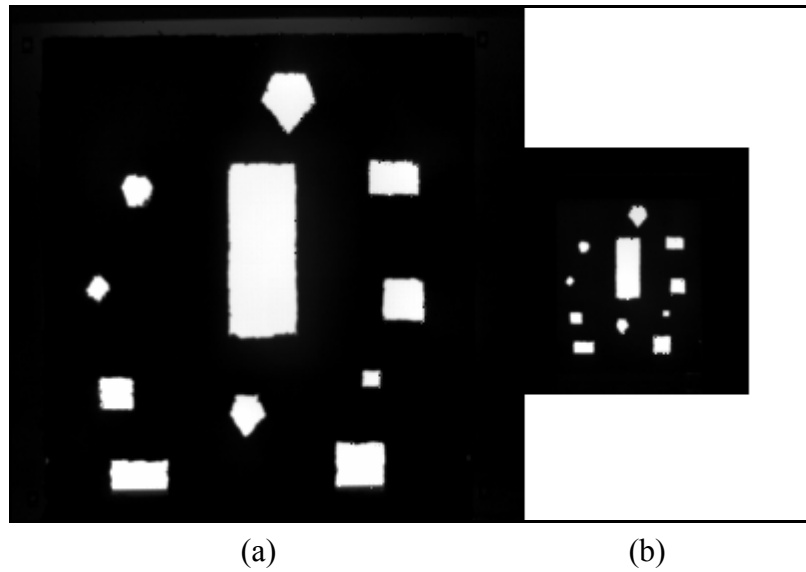


Figure 4-96 MWIR Band Images with 150°C Blackbody at 8m and 22.5m.

(a) Reference image at 8m, (b) Transformed image at 22.5m

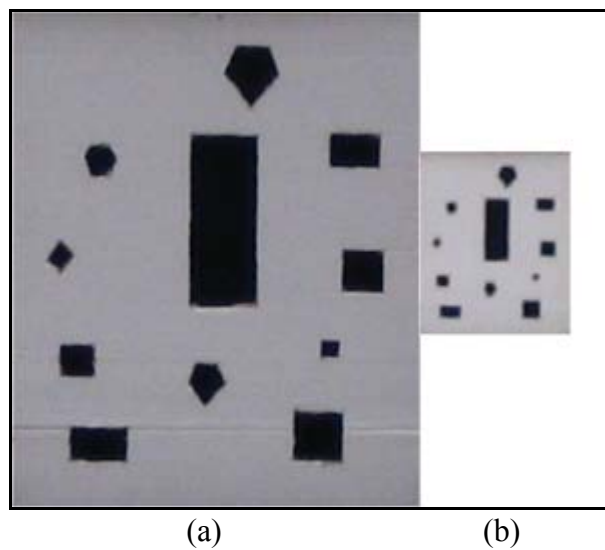


Figure 4-97 Visual Band Images with 150°C Blackbody at 8m and 22.5m.

(a) Reference image at 8m, (b) Transformed image at 22.5m

The results of this experiment are presented in the next section.

4.5.2 Results of Matching Performance with Changing Distance Experiment

Feature points of MWIR-band image pairs belonging to 150°C blackbody are presented in Figure 4-98, Figure 4-99, Figure 4-100, and Figure 4-101 below, extracted by SIFT, Harris-Laplace, Hessian-Laplace and SURF detectors, respectively.

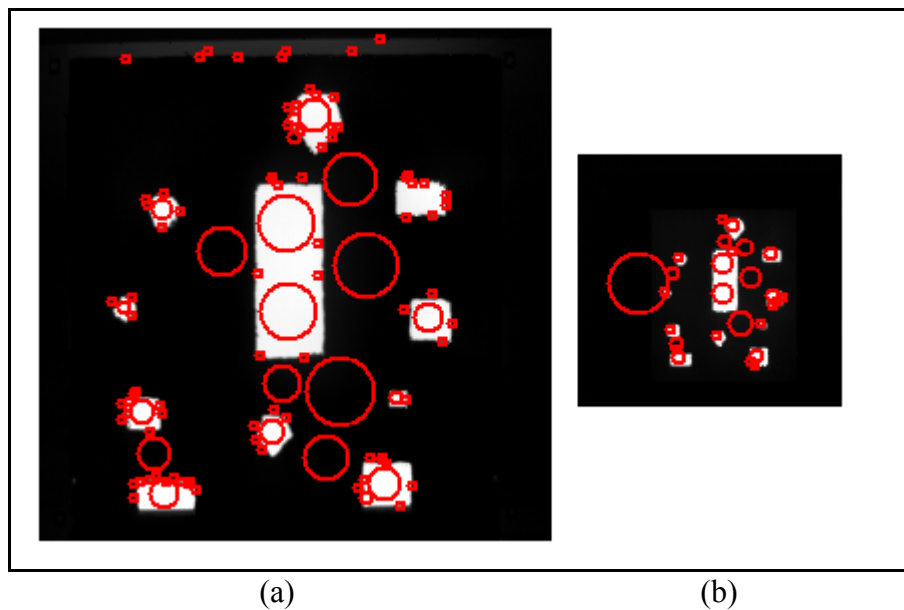
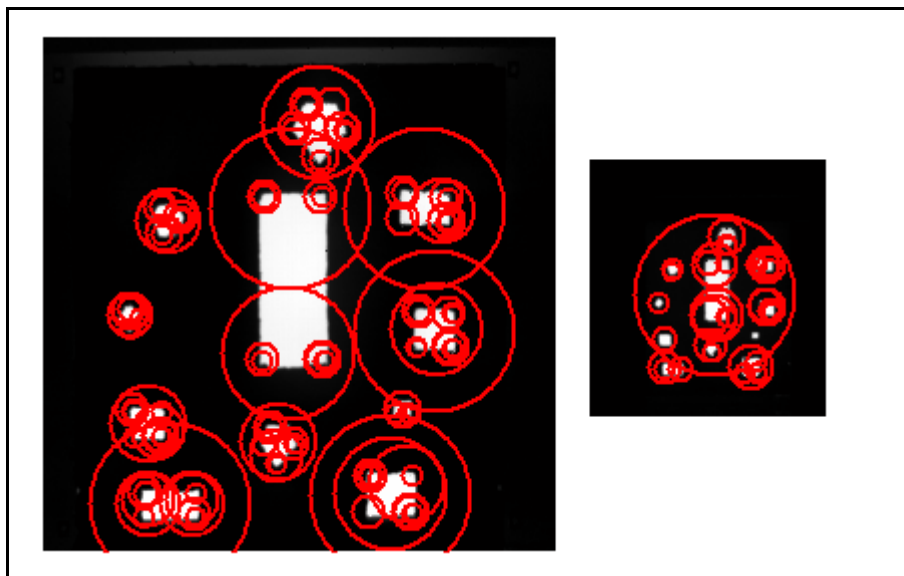


Figure 4-98 MWIR Band Image Feature Points Extracted by SIFT Detector, 150°C Blackbody.

(a) Reference image at 8m, (b) Transformed image at 22.5m

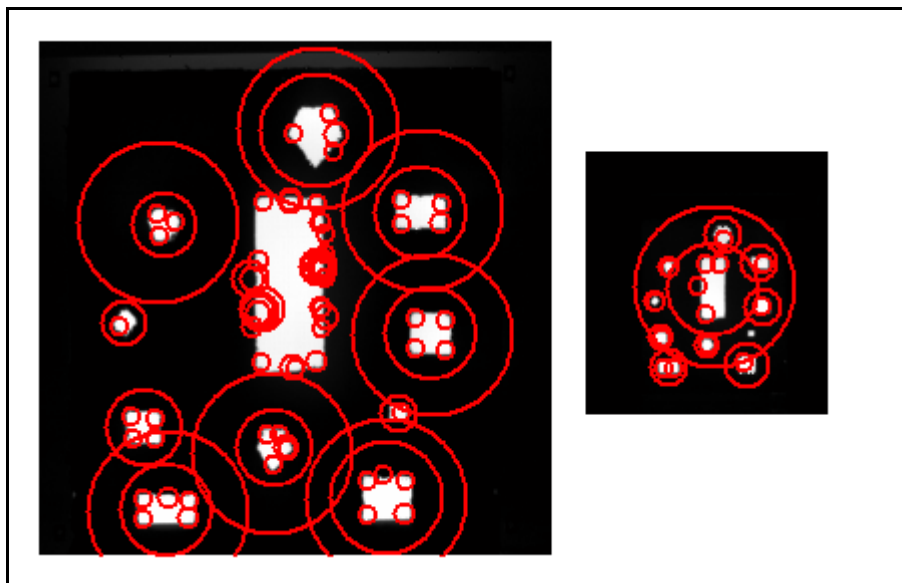


(a)

(b)

Figure 4-99 MWIR Band Image Feature Points Extracted by Harris-Laplace Detector, 150°C Blackbody.

(a) Reference image at 8m, (b) Transformed image at 22.5m



(a) (b)
Figure 4-100 MWIR Band Image Feature Points Extracted by Hessian-Laplace
Detector, 150°C Blackbody.

(a) Reference image at 8m, (b) Transformed image at 22.5m

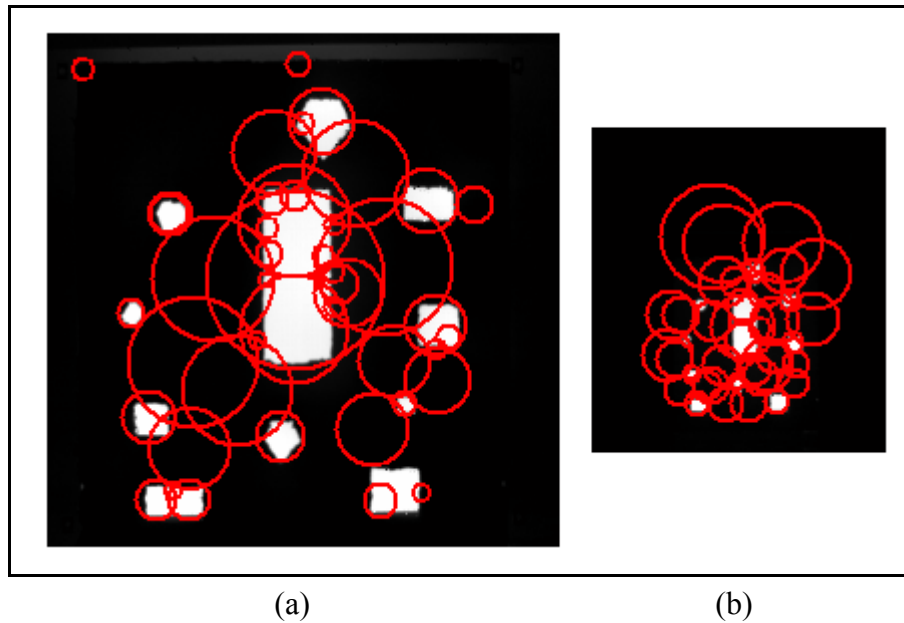


Figure 4-101 MWIR Band Image Feature Points Extracted by SURF Detector, 150°C Blackbody.

(a) Reference image at 8m, (b) Transformed image at 22.5m

It can be observed from these figures that feature points extracted are scale invariant, since almost the entire feature regions found in the reference images at 8m are observed in the transformed images as well.

Feature points of visual-band image pairs belonging to 150°C blackbody are presented in Figure 4-102, Figure 4-103, Figure 4-104, and Figure 4-105 below, extracted by SIFT, Harris-Laplace, Hessian-Laplace and SURF detectors, respectively.

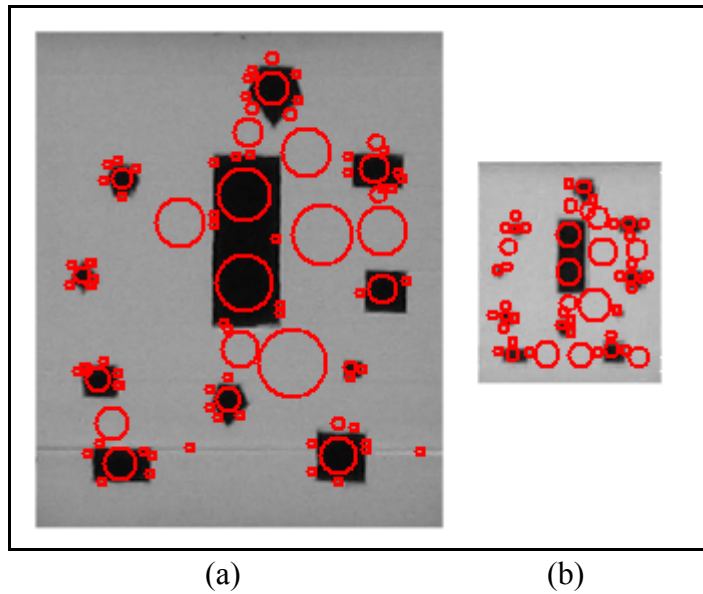


Figure 4-102 Visual Band Image Feature Points Extracted by SIFT Detector.

(a) Reference image at 8m, (b) Transformed image at 22.5m

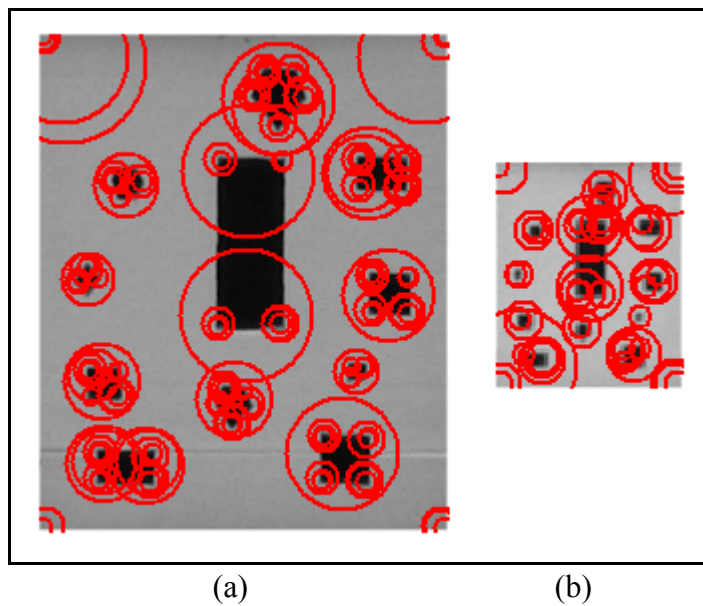
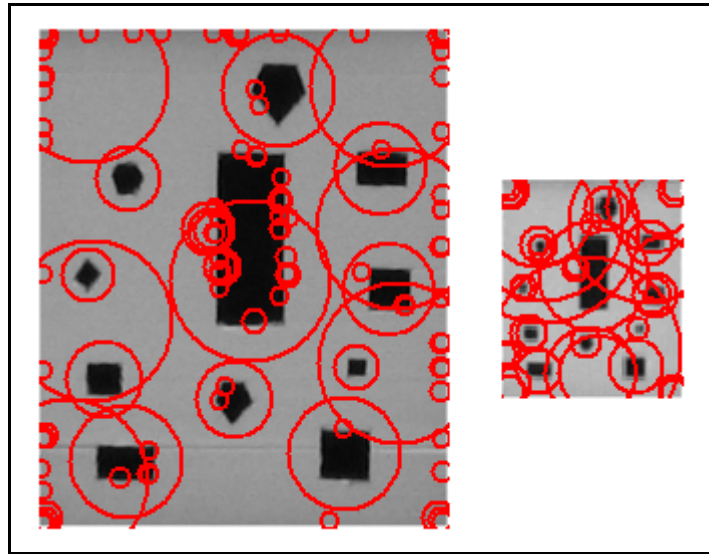


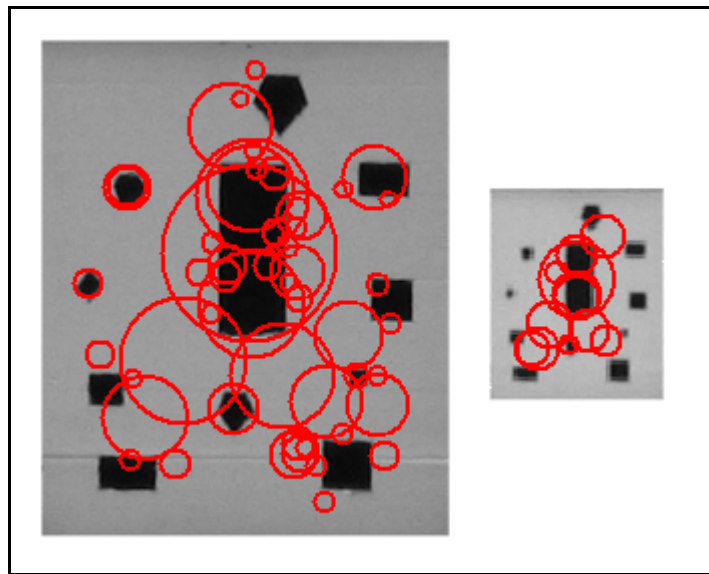
Figure 4-103 Visual Band Image Feature Points Extracted by Harris-Laplace Detector.

(a) Reference image at 8m, (b) Transformed image at 22.5m



(a) (b)
 Figure 4-104 Visual Band Image Feature Points Extracted by Hessian-Laplace Detector.

(a) Reference image at 8m, (b) Transformed image at 22.5m



(a) (b)
 Figure 4-105 Visual Band Image Feature Points Extracted by SURF Detector.

(a) Reference image at 8m, (b) Transformed image at 22.5m

Feature points extracted for LWIR-band image pairs of 8m and 22.5m distances are similar and not given. Moreover, feature regions of the 15m, 30m, and 43m image pairs are not presented for all the bands due to space limitations.

The same analysis steps are performed for this experiment too. Matching points are obtained, matched points outside the region of interest are not used in the calculations and matching accuracy is found.

The effect of distance of the measurement on the matching accuracy in all of the bands is presented in Figure 4-106, Figure 4-107, Figure 4-108, and Figure 4-109 below for SIFT, Harris-Laplace, Hessian-Laplace, and SURF detectors respectively.

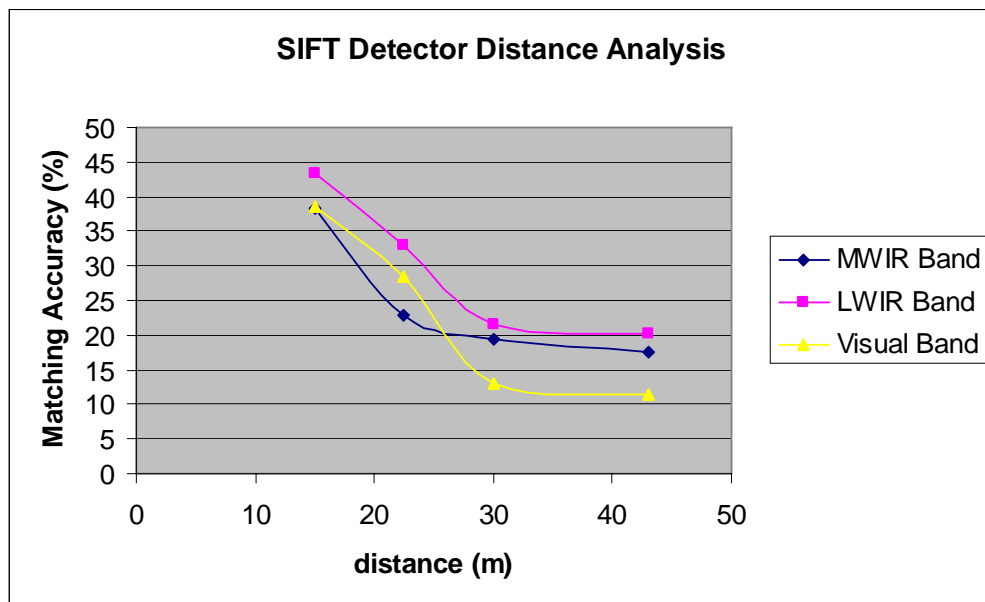


Figure 4-106 Matching Accuracy with Changing Distance : SIFT

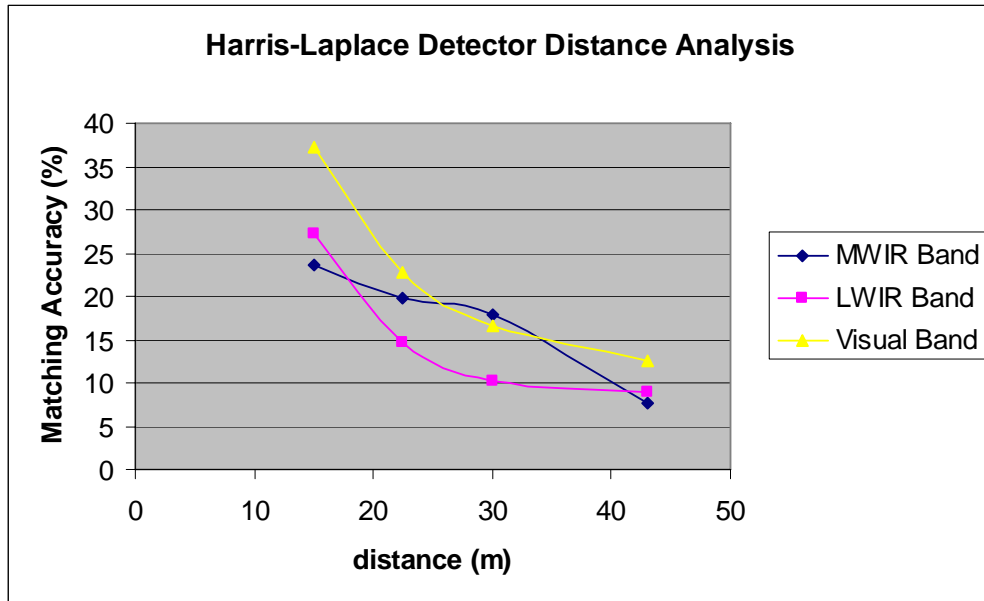


Figure 4-107 Matching Accuracy with Changing Distance : Harris-Laplace

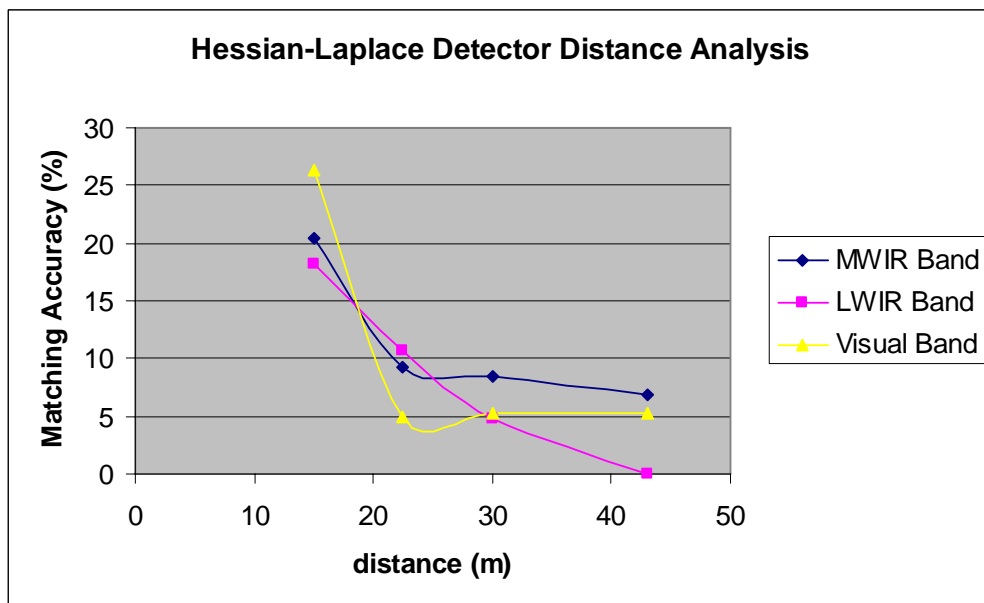


Figure 4-108 Matching Accuracy with Changing Distance : Hessian-Laplace

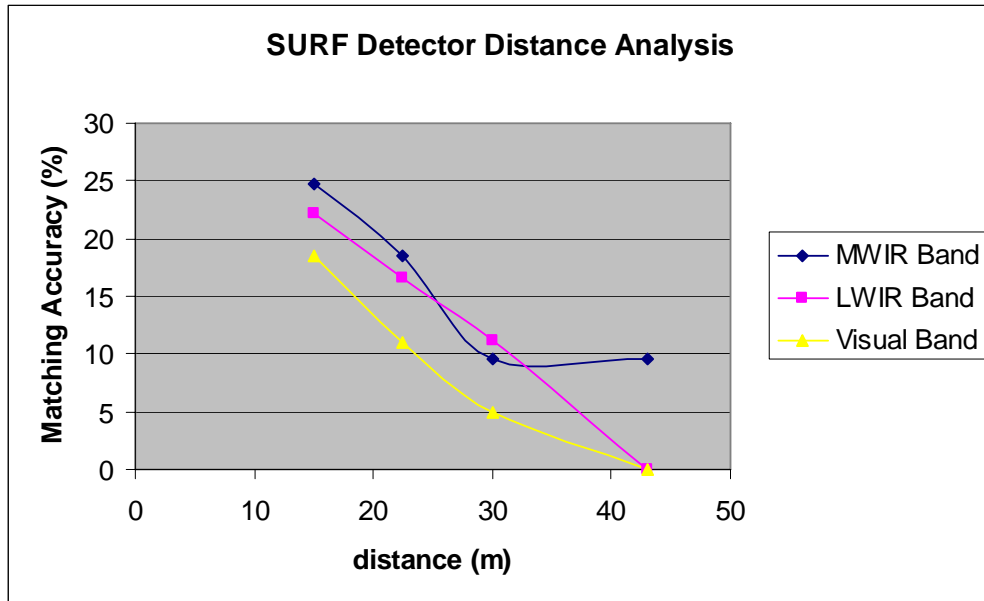


Figure 4-109 Matching Accuracy with Changing Distance : SURF

Matching accuracy is higher when distance is lowest at 15m in all of three bands for SIFT, Harris-Laplace, Hessian-Laplace, and SURF detectors. As distance increases, matching accuracy decreases.

Matching accuracy percentage values are summarized for each detector below.

Table 4-6 Matching Accuracy Values, SIFT Detector, Experiment 4

Matching Accuracy	Reference Distance (m)	Distance (m)			
		15	22.5	30	43
MWIR Band	8	38	23	20	18
LWIR Band	8	43	33	22	20
Visual Band	8	39	28	13	11

Table 4-7 Matching Accuracy Values, Harris-Laplace Detector, Experiment 4

Matching Accuracy	Reference Distance (m)	Distance (m)			
		15	22.5	30	43
MWIR Band	8	24	20	18	8
LWIR Band	8	27	15	10	9
Visual Band	8	37	23	17	13

Table 4-8 Matching Accuracy Values, Hessian-Laplace Detector, Experiment 4

Matching Accuracy	Reference Distance (m)	Distance (m)			
		15	22.5	30	43
MWIR Band	8	20	9	8	7
LWIR Band	8	18	11	5	0
Visual Band	8	26	5	5	5

Table 4-9 Matching Accuracy Values, SURF Detector, Experiment 4

Matching Accuracy	Reference Distance (m)	Distance (m)			
		15	22.5	30	43
MWIR Band	8	25	19	10	10
LWIR Band	8	22	17	11	0
Visual Band	8	19	11	5	0

The discussion of the results of all experiments is given in Conclusions chapter of this study.

CHAPTER 5

CONCLUSIONS AND DISCUSSIONS

Experimental studies are completed and results are obtained. The performance evaluation of SIFT, Harris-Laplace, Hessian-Laplace, and SURF feature detectors along with SIFT and SURF type descriptors are achieved based on matching accuracy of object images in infrared and visual-bands for changing viewpoints and distance conditions.

Four experimental studies are completed, that are listed below.

1. Experiment 1 : LWIR versus MWIR Infrared Images Experiment
2. Experiment 2 : LWIR versus Visual Band Images Experiment
3. Experiment 3 : Matching Performance with Changing Rotation Experiment
4. Experiment 4 : Matching Performance with Changing Distance Experiment

Important conclusions and results are obtained out of these experiments. Firstly, feature points extracted by all feature detectors are rotation invariant. This conclusion is supported by the results of experiments 1, 2, and 3. The feature points obtained by all detectors have orientation components providing invariance to rotation type of transformations. The results supporting this conclusion have some performance limitations.

Second important conclusion is that feature points extracted by all feature detectors are scale invariant. This is supported by the results of experiment 4. The reason for this result is due to the fact that feature points have scale components providing invariance to scale changes. The results supporting this conclusion have some performance limitations as well.

Matching accuracy of targets is better in MWIR-band for all feature detectors as the temperature of the targets increase. This conclusion is supported by the results of experiment 1. This result is due to the fact that MWIR-band is more successful for capturing the hot parts of the objects better than LWIR-band.

Matching accuracy of objects at higher temperature (112°C) in MWIR band is greater than or equal to the matching accuracy of objects at higher temperature (112°C) in LWIR band for all feature detectors used. This conclusion is supported by the results of experiment 1. The reason for this result is same with the previous one, which is MWIR-band camera captures the hot parts of the objects better than the LWIR camera.

Matching accuracy of targets at lower temperature (32°C) in LWIR-band is greater than the matching accuracy of targets at lower temperature (32°C) in MWIR-band for all detectors. This observation is obtained as a result of LWIR-band's strength in capturing the cool parts of the objects better than MWIR-band.

SIFT, Hessian-Laplace and SURF feature points are detected at round, curved and blob-like regions, whereas Harris-Laplace feature points are detected near corners or near regions having similar shape to corners in all of the experiments. The reason for this result is a result of the nature of each feature detector utilized. Hessian matrix based measures are good at detecting blob like structures. SIFT detector utilizes difference of Gaussian, which is an approximation to scale-normalized LoG. Laplacian is actually the trace of Hessian matrix, thus SIFT detector is an approximation to a Hessian matrix measure. Hessian-Laplace and SURF detectors

naturally utilize Hessian matrix based measures. Harris-Laplace detector uses cornerness measure, on the other hand.

Matching accuracy of face images decrease with increasing rotation angle in both LWIR and visual-bands, for all detectors used. This is due to the fact that viewpoint change of the face is a strong affine distortion and feature matching is simple and it is not affine invariant with some limitations. This reason is true for both LWIR and visual-bands. Illumination change of face images in visual-band is another reason. It affects the matching performance of all the detectors, when compared with face images in LWIR-bands. When there is a viewpoint change of face in Visual band, the illumination on the face surface affects the matching performance.

Matching accuracy of face and vehicle images decrease with increasing distance in both LWIR and visual-bands, for all detectors utilized. The reason for this result is that details of the face at far away distances decrease drastically as the distance of the measurement increases. In other words, there are less feature regions left for the matching.

Matching accuracy performance of SURF detector for face and vehicle images in terms of rotation angle is better in LWIR and visual-bands against other detectors. There are two reasons for this important result. Firstly, integral images provide not only very fast feature extraction of infrared images, but also they result in a high efficient matching performance for human tissue recognition in long wave infrared band. Secondly, the use of integral images provides such high performance in visual-band too. Moreover, the other steps of low dimension descriptor finding provides high probability of efficient matching between reference and training samples.

Matching accuracy performance of SURF detector for face images in terms of distance is also better in LWIR and visual-bands against other detectors. The reasons for this result are the same as the reasons of the previous result.

Matching accuracy of vehicle images decrease with increasing rotation angle in both LWIR and visual-bands, for all detectors utilized. This is due to the fact that vehicle rotation is a strong affine distortion and feature matching has some limitations in such cases, as it is previously stated.

The blackbody experiments for changing rotation and distance have similar results with the previous experiments. Matching accuracy decreases as rotation angle increases at all distances in MWIR, LWIR, and visual-bands for all feature detectors. Moreover, it decreases as distance of the measurement increases in MWIR, LWIR, and visual-bands for all feature detectors. This is a result of strong affine distortion of mask rotation and distance changes.

REFERENCES

- [1] D.G. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints”, International Journal of Computer Vision, vol. 60, no. 2, 2004
- [2] K. Mikolajczyk, and C. Schmid, “Scale & Affine Invariant Interest Point Detectors”, International Journal of Computer Vision, vol. 60, no. 1, 2004
- [3] H. Bay, T. Tuytelaars, and L.V. Gool, “SURF: Speeded Up Robust Features”, in Proceedings of the ninth European Conference on Computer Vision, Part I, Lecture Notes in Computer Science, vol 3951, pp. 404-417, 2006
- [4] P.A. Jacobs, “Thermal Infrared Characterization of Ground Targets and Backgrounds”, SPIE Press, Washington, 2006
- [5] A. Ryer, “Light Measurement Handbook”, <http://www.intl-light.com/handbook>, International Light, 1998
- [6] D.L. Adamy, “EW 102: A Second Course in Electronic Warfare”, Artech House Publishers, 2004
- [7] Prof.Dr. H. Geller, “Astr 402, Observational Astronomy Lecture Notes”, <http://physics.gmu.edu/~hgeller/astr402/ImageFormation.pdf>, 2006
- [8] F.J. Crawford, “Electro-Optical Sensors Overview”, IEEE AES Systems Magazine, October 1998
- [9] S.A. Miller and J.P. Harley, “Zoology”, Wm.C.Brown, 1991
- [10] C. Mueller and M. Rudolph, “Light and Vision”, Time Life International, 1972

- [11] İ. Bayram, “Interest Point Matching Across Arbitrary Views”, M.Sc. Thesis, Middle East Technical University, 2004
- [12] T. Lindeberg, “Feature Detection With Automatic Scale Selection”, *International Journal of Computer Vision*, vol. 30, no. 2, 1998
- [13] T. Lindeberg, “Scale-Space For Discrete Signals”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 3, 1990
- [14] D.G. Lowe, “Object Recognition from Local Scale-Invariant Features”, *International Conference on Computer Vision*, 1999
- [15] M. Brown, and D.G. Lowe, “Automatic Panoramic Image Stitching using Invariant Features”, *International Journal of Computer Vision*, vol. 74, no. 1, 2007
- [16] N. Snavely, S.M. Seitz, and R. Szeliski, “Modeling the World from Internet Photo Collections”, *International Journal of Computer Vision*, DOI 10.1007, 2007
- [17] A.S. Mian, M. Bennamoun, and R. Owens, “Keypoint Detection and Local Feature Matching for Textured 3D Face Recognition”, *International Journal of Computer Vision*, vol. 79, no. 1, 2008
- [18] P. Moreels, and P. Perona, “Evaluation of Features Detectors and Descriptors based on 3D Objects”, *International Journal of Computer Vision*, vol. 73, no. 3, 2007
- [19] J. Zhang, M. Marszalek, S. Lazebnik, and C. Schmid, “Local Features and Kernels for Classification of Texture and Object Categories: A Comprehensive Study”, *International Journal of Computer Vision*, vol. 73, no. 2, 2007
- [20] L. Yang, R. Jin, C. Pantofaru, and R. Sukthankar, “Discriminative Cluster Refinement: Improving Object Category Recognition Given Limited Training Data”, *IEEE Conference on Computer Vision and Pattern Recognition*, DOI 10.1109, 2007

- [21] M. Marszalek, and C. Schmid, “Semantic Hierarchies for Visual Object Recognition”, IEEE Conference on Computer Vision and Pattern Recognition, DOI 10.1109, 2007
- [22] F. Fraundorfer, and H. Bischof, “A novel performance evaluation method of local detectors on non-planar scenes”, IEEE Conference on Computer Vision and Pattern Recognition, vol. 3, pp. 33, 2005
- [23] P.R. Beaudet, “Rotationally invariant image operators”, International Joint Conference on Pattern Recognition, pp. 579-583, 1978
- [24] K. Mikolajczyk, and C. Schmid, “A performance evaluation of local descriptors”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, Issue 10, 2005
- [25] P. Viola, and M. Jones, “Rapid Object Detection using a Boosted Cascade of Simple Features”, IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 511, 2001
- [26] K. Pan, S. Liao, Z. Zhang, S.Z. Li, and P. Zhang, “Part-based Face Recognition Using Near Infrared Images”, IEEE Conference on Computer Vision and Pattern Recognition, DOI 10.1109, pp. 1-6, 2007
- [27] J. Heo, M. Savvides, and B.V.K. Vijayakumar, “Performance Evaluation of Face Recognition using Visual and Thermal Imagery with Advanced Correlation Filters”, IEEE Conference on Computer Vision and Pattern Recognition, vol. 3, pp. 9, 2005
- [28] D.A. Socolinsky, and A. Selinger, “Thermal Face Recognition in an Operational Scenario”, IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 1012-1019, 2004

- [29] D. Tan, K. Huang, S. Yu, and T. Tan, "Recognizing Night Walkers Based on One Pseudoshape Representation of Gait", IEEE Conference on Computer Vision and Pattern Recognition, DOI 10.1109, pp. 1-8, 2007
- [30] L. Zhang, B. Wu, and R. Nevatia, "Pedestrian Detection in Infrared Images based on Local Shape Features", IEEE Conference on Computer Vision and Pattern Recognition, DOI 10.1109, pp. 1-8, 2007
- [31] S.Y. Cheng, S. Park, and M.M. Trivedi, "Multiperspective Thermal IR and Video Arrays for 3D Body Tracking and Driver Activity Analysis", IEEE Conference on Computer Vision and Pattern Recognition, vol. 3, p.p 3, 2005
- [32] C.O. Conaire, E. Cooke, N.O. Connor, N. Murphy, and A. Smeaton, "Background Modelling in Infrared and Visible Spectrum Video for People Tracking", IEEE Conference on Computer Vision and Pattern Recognition, vol. 3, pp. 20, 2005
- [33] C.H. Pai, Y.P. Lin, G.G. Medioni, and R.R. Hamza, "Moving Object Detection on a Runway Prior to Landing Using an Onboard Infrared Camera", IEEE Conference on Computer Vision and Pattern Recognition, DOI 10.1109, pp. 1-8, 2007
- [34] S. Newsam, and Y. Yang, "Geographic Image Retrieval Using Interest Point Descriptors", Lecture Notes in Computer Science, vol 4842, pp. 275-286, 2007
- [35] C. Harris, and M. Stephens, "A combined corner and edge detector", In Alvey Vision Conference, pp. 147-151, 1988
- [36] Y. Wang, A. Camargo, R. Fevig, F. Martel, and R.R. Schultz, "Image Mosaicking from Uncooled Thermal IR Video Captured by a Small UAV", IEEE Southwest Symposium on Image Analysis and Interpretation, DOI 10.1109, pp. 161-164, 2008

- [37] Wikipedia, “IKONOS”, <http://en.wikipedia.org/wiki/IKONOS>, 16th July 2008, Last date accessed: 12th August 2008
- [38] P.J. Burt, and E.H. Adelson, “The Laplacian Pyramid as a Compact Image Code”, IEEE Transactions on Communications, vol. Com-31, no. 4, 1983
- [39] A. Vedaldi, “An open implementation of SIFT”, <http://vision.ucla.edu/~vedaldi/code/sift/sift.html>, 2005-2008, Last date accessed: 12th August 2008
- [40] V. Garcia, “Keypoint Extraction”, http://www.i3s.unice.fr/~garcia/goodies_matlab.php, 2008, Last date accessed: 12th August 2008
- [41] H. Bay, T. Tuytelaars, and L.V. Gool, “SURF: Speeded Up Robust Features”, http://www.vision.ee.ethz.ch/~surf/download_ac.html, 2006, Last date accessed: 12th August 2008
- [42] R. Cipolla, and M. Johnson, “Handout 3: Feature Descriptors”, Information Engineering Elective Paper 8 Image Search and Matching, mi.eng.cam.ac.uk/~cipolla/lectures/PartIB/IB-descriptors.pdf, 2006, Last date accessed: 13th August 2008
- [43] T. Cukur, Y. Altug, C. Uzunoglu, K. Kilic, and E. Emir, “IRLook: An advanced mobile infrared signature measurement, data reduction, and analysis system”, Proceedings of SPIE Preprint, vol. 6569, 2007
- [44] R. Deriche, and G. Giraudon, “Accurate Corner Detection : An Analytical Study”, International Conference on Computer Vision, pp. 66-70, 1990
- [45] Wikipedia, “Haar wavelet”, http://en.wikipedia.org/wiki/Haar_wavelet, 30th October 2008, Last date accessed: 22nd November 2008

- [46] Wikipedia, “Scale space”, http://en.wikipedia.org/wiki/Scale_space, 22nd June 2008, Last date accessed: 22nd November 2008
- [47] A. Vedaldi, “An implementation of SIFT detector and descriptor”, <http://vision.ucla.edu/~vedaldi/code/sift/assets/sift/sift.pdf>, 2005-2008, Last date accessed: 23rd November 2008
- [48] M. Brown, and D.G. Lowe, “Invariant Features from Interest Point Groups”, British Machine Vision Conference, pp. 656-665, 2002
- [49] O. Pele, “SIFT - The Scale Invariant Feature Transform”, www.cse.huji.ac.il/course/2006/compvis/lectures/SIFT.ppt, 2006, Last date accessed: 15th December 2008
- [50] Y.M. Wyman Wong, “SURF: Speeded Up Robust Features, ECCV 2006, Group Meeting Presentation”, www.cse.cuhk.edu.hk/~lyu/seminar/06fall/Wyman.ppt, 2006, Last date accessed: 15th December 2008
- [51] C. Kambhamettu, “Corner Detection, CISC 4/689: COMPUTER VISION Lecture Presentation”, vims.cis.udel.edu/~chandra/689/Spring06/Corners+Ransac.ppt, 2006, Last date accessed: 15th December 2008