

Timescape Image Panorama Registration Techniques

by

Heider K. Muhsin Ali, M.Sc

A thesis submitted to the

Faculty of Graduate and Postdoctoral Affairs

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Electrical and Computer Engineering

Ottawa-Carleton Institute for Electrical and Computer Engineering (OCIECE)

Department of Systems and Computer Engineering

Carleton University

Ottawa, Ontario

November 2016

© Heider K. Muhsin Ali, 2016

Abstract

Panoramic image viewing in its common form simulates a comprehensive spatial or temporal view. The term panorama indicates the concepts of space (or place) and time. The space sense of panorama has dominated as the default concept since it has been scientifically and commercially applied for decades. However, the temporal concept of panorama, which has been recently investigated by researchers and historians, offers a new context of the definition of panorama.

A novel application in the field of panorama display has been realized and presented by this work which aims to participate effectively in cultural heritage preservations around the world and to provide an easy access tool to demonstrate a timelined panorama view of landmarks.

Consequently, a new concept of panorama has been accomplished through a timelined display of registered historic and modern images for many famous landmarks around the world. This is achieved by merging geographical and historical information into a single attractive temporal and spatial panoramic view; also known as a timescape. This comprehensive view is achieved from a collection of historic and modern images of such landmarks available on the Internet.

Hundreds of thousands of landmark images covering more than one century of time have been collected using websites like Flickr and Google Images. We use Gabor filters and neural networks to select a subset of images that have roughly the same view for each landmark in order to keep the most suitable images that will be used to build and demonstrate the temporal panorama. The processing of the selected set of historic and modern images has to contend with radiometric differences, geometric distortions, and

severely differing capture systems and environments in order to prepare them for alignment.

Finally, the aligned images; warped or registered, were prepared for presentation in a timelined image display. The image registration is done using optimized optical flow of extracted SIFT features after applying a set of similarity measure standards to select the highest precision subset of aligned images and to develop the timescape display.

Acknowledgements

I am so grateful to the GOD (ALLAH) for his uncountable bounties on me including the ability and help to perform and write this thesis.

I would like to express my deep and sincere appreciation to my thesis supervisor, Professor Anthony Whitehead, for his precious guidance throughout the period of my graduate study. This thesis would not have been possible without his direction and assistance. His vast knowledge and continuous guidance have helped shape the contributions in this thesis.

I express my great appreciation to my wife, Zahra who dedicated her never ending love for me and was my partner towards achievement, and to my adorable daughters and son, who were patient with me while being always busy. I also appreciate the love and support of my brothers and sisters who are always the source inspiration and encouragement for me.

Table of Contents

Abstract.....	ii
Acknowledgements	iv
Table of Contents	v
List of Table.....	xi
List of Illustrations.....	xiii
List of Appendices.....	xxii
Chapter 1: Introduction	1
1.1 Cultural Heritage Preservation	1
1.2 Panoramas.....	2
1.3 Historic Photos Challenges.....	3
1.4 Thesis Objectives.....	4
1.5 System Structure.....	6
1.6 Thesis Contributions.....	8
1.7 Thesis Organization.....	9
Chapter 2: Background and Related Work	10
2.1 Historical Photos and Documents.....	11
2.2 Photographic Technology Evolution.....	14
2.2.1 Technical Aspects of Photography.....	17
2.2.2 Exposure and Rendering.....	18
2.3 Image Subset Selection.....	19
2.3.1 Image classification approaches.....	19
2.3.2 Image Classification Techniques.....	21

2.3.3	Image Selection Literature Review.....	23
2.4	Automatic Image Registration.....	24
2.4.1	Registration Methods and Classification.....	25
2.4.1.1	Automation Degree.....	25
2.4.1.2	Single and Multi-Modal Methods.....	26
2.4.1.3	Linear and Nonlinear Transformation.....	26
2.4.1.4	Spatial and Frequency Domain Methods.....	26
2.4.1.5	Parametric and Nonparametric Image Registration.....	26
2.4.1.5.1	Parametric Geometric Transformations.....	27
2.4.1.5.1.1	Rigid Transformations.....	27
2.4.1.5.1.2	Affine Transformations.....	28
2.4.1.5.1.3	Phase Correlation.....	29
2.4.1.5.2	Nonparametric Geometric Transformations.....	29
2.4.2	Similarity Measure Criteria.....	29
2.4.2.1	Similarity Measures for Linear Relations.....	30
2.4.2.1.1	Sum of Squared Differences.....	30
2.4.2.1.2	Histogram Difference Measure.....	30
2.4.2.1.3	Cross Correlation.....	31
2.4.2.2	Similarity Measures for Nonlinear Relations.....	31
2.4.2.2.1	Correlation Ratio.....	31
2.4.2.2.2	Structural Similarity Index (SSIM).....	31
2.4.2.2.3	Root-Mean Square Error (RMSE).....	32
2.4.2.2.4	Mutual Information.....	32
2.5	Feature Extraction and Matching.....	33
2.5.1	Low-Level Feature Extraction.....	33
2.5.1.1	Edge Detection.....	33

2.5.1.2	Phase congruency.....	34
2.5.1.3	Localized feature extraction.....	35
2.5.1.4	Describing Image Motion.....	36
2.5.2	Feature Extraction Standard Techniques.....	37
2.5.3	High-Level Feature Extraction.....	45
2.5.3.1	Deformable Templates.....	46
2.5.3.2	Parts-Based Shape Analysis.....	47
2.5.3.3	Active Contours/Snakes.....	47
2.5.3.4	Shape Skeletonization.....	48
2.6	Optimization Process in Image Registration.....	49
2.6.1	Gauss-Newton (GN) Optimization Algorithm.....	51
2.6.2	Levenberg-Marquardt (LM) Optimization Algorithm.....	51
2.6.3	Quasi-Newton (QN) Optimization Algorithm.....	52
2.6.4	Nonlinear Conjugate Gradient (NCG) Optimization Algorithm.....	52
2.6.5	Markov Random Fields (MRF).....	52
2.6.6	Belief Propagation.....	54
2.7	Panorama Types.....	56
2.7.1	Spatial Panoramas.....	56
2.7.1.1	2-D Spatial Panoramas.....	57
2.7.1.2	3-D Spatial Panoramas.....	58
2.7.2	Temporal Panoramas.....	60
2.7.2.1	Time-lapse Videos.....	60
2.7.2.2	Historic Perspective Panorama (HPP).....	61
2.7.3	Timelined Panorama.....	62
2.7.3.1	Rideau Timescapes.....	62

Chapter 3: Image Subset Selection.....	63
3.1 Timeline Based Image Data Set.....	64
3.2 Gabor Functions and Filters for Feature Extraction.....	65
3.3 Neural Networks.....	71
3.3.1 Neural Network Training.....	73
3.4 Experimental Results and Analysis.....	82
Chapter 4: Automatic Feature Extraction	89
4.1 Feature extraction: Optical Flow of Extracted Features.....	90
4.1.1 Dense SIFT Features and Optical Flow.....	92
4.1.2 Results and Discussion.....	94
4.2 Feature extraction: Standard Techniques.....	96
4.2.1 Feature Extraction and Matching Techniques.....	96
4.2.1.1 Matching Procedure.....	98
4.2.1.2 Disparity Gradient Filter.....	99
4.2.2 Harris Corner Detector.....	99
4.2.3 SIFT/SIFT.....	101
4.2.4 SURF/SURF.....	103
4.2.5 BRIEF/BRIEF.....	103
4.2.6 BISK/BRISK.....	104
4.2.7 ORB/ORB.....	105
4.2.8 Hybrid ORB Detector /SURF Descriptor.....	105
4.2.9 Results and Analysis.....	106
4.2.9.1 Technique Accuracy.....	106
4.2.9.2 Runtime.....	109
4.2.9.3 ORB/SURF Effectiveness.....	109
4.3 Dense SIFT of Optical Flow and Standard Techniques Evaluation.....	110

Chapter 5: Aligning Images	111
5.1 Automatic Image Registration (Optimized Optical Flow of Detected Features).....	112
5.1.1 Loopy Priority Belief Propagation for Optical Flow.....	117
5.1.2 Results and Discussion.....	119
5.1.3 Computation Complexity.....	128
5.2 Image Warping (Standard Techniques).....	130
Chapter 6: Timelined Panorama Display	132
6.1 Interface Design.....	132
6.1.1 Web-Based Graphics and Informatics (WBGI) Module.....	133
6.1.2 Image Subset Selection (ISS) Module.....	134
6.1.3 Automatic Image Registration (AIR) Module.....	134
6.1.4 Timescapes Panorama Display (TPD) Module.....	135
6.2 Panoramic Display.....	137
6.3 Results and Discussion.....	139
6.4 Timescape Panorama and Existing Spatial-Temporal Panoramas Evaluation.....	143
Chapter 7: Conclusions and Future Work	144
7.1 Conclusions.....	144
7.2 Future Work.....	147
Appendices.....	150
Appendix A: Dense SIFT descriptors for landmark images.....	150
Appendix B: Landmark matching tables for HARRIS/HARRIS feature matching technique	154
Appendix C: Landmark matching tables for SIFT/SIFT feature matching technique	157
Appendix D: Landmark matching tables for SURF/SURF feature matching technique.....	159
Appendix E: Automatic image registration applied on landmark images	163

Appendix F: User Experience	172
References.....	174

List of Tables

Table 2.1:	Comparison of different techniques of Feature Detectors and Descriptors..	43
Table 2.2	Set of 2D coordinate transformations	58
Table 3.1	Applied algorithms results on landmark images collected from the image websites	84
Table 3.2	Accuracy table of the algorithm applied on all landmarks image sets.....	85
Table 4.1	Eiffel tower landmark matching using HARRIS/HARRIS matcher.....	100
Table 4.2	Eiffel tower landmark matching using SIFT/SIFT matcher.....	102
Table 4.3	Taj-Mahal landmark matching using SIFT/SIFT matcher.....	102
Table 4.4	Eiffel tower landmark matching using SURF/SURF matcher.....	103
Table 4.5	Empire State building landmark matching using ORB/SURF matcher.....	106
Table 4.6	Accuracy of high performance image matching techniques for each landmark, with highest accuracy rates in bold.....	107
Table 4.7	Runtime of high performance matching techniques.....	109
Table 5.1	Comparison of the related registration approaches.....	129
Table 6.1	Comparison of panorama types	143
Table B.1	Pyramids of Giza landmark matching using HARRIS/HARRIS matcher...	154
Table B.2	Coliseum of Rome landmark matching using HARRIS/HARRIS matcher.	154
Table B.3	Dome Rock landmark matching using HARRIS/HARRIS matcher.....	154
Table B.4	Empire State building landmark matching using HARRIS/HARRIS matcher.....	155
Table B.5	Hagia-Sophia landmark matching using HARRIS/HARRIS matcher.....	155
Table B.6	Machu-Picchu landmark matching using HARRIS/HARRIS matcher.....	155

Table B.7 Stonehenge landmark matching using HARRIS/HARRIS matcher.....	156
Table B.8 Taj-Mahal landmark matching using HARRIS/HARRIS matcher.....	156
Table B.9 White House of Washington landmark matching using HARRIS / HARRIS.....	156
Table C.1 Pyramids of Giza landmark matching using SIFT/SIFT matcher.....	157
Table C.2 Coliseum of Rome landmark matching using SIFT/SIFT matcher.....	157
Table C.3 Dome Rock landmark matching using SIFT/SIFT matcher.....	157
Table C.4 Empire State building landmark matching using SIFT/SIFT matcher.....	158
Table C.5 Hagia-Sophia landmark matching using SIFT/SIFT matcher.....	158
Table C.6 Machu-Picchu landmark matching using SIFT/SIFT matcher.....	158
Table C.7 Stonehenge landmark matching using SIFT/SIFT matcher.....	159
Table C.8 Whitehouse of Washington landmark matching using SIFT/SIFT matcher	159
Table D.1 Pyramids of Giza landmark matching using SURF/SURF matcher.....	159
Table D.2 Coliseum of Rome landmark matching using SURF/SURF matcher.....	160
Table D.3 Dome Rock landmark matching using SURF/SURF matcher.....	160
Table D.4 Empire State building landmark matching using SURF/SURF matcher....	160
Table D.5 Hagia-Sophia landmark matching using SURF/SURF matcher.....	161
Table D.6 Machu-Picchu landmark matching using SURF/SURF matcher.....	161
Table D.7 Stonehenge landmark matching using SURF/SURF matcher.....	161
Table D.8 Taj-Mahal landmark matching using SURF/SURF matcher.....	162
Table D.9: Whitehouse of Washington landmark matching using SURF/SURF matcher	162

List of Illustrations

Figure 1.1 Pinhole camera (a) Geometry of pinhole camera without lens (b) Pinhole camera with lens	4
Figure 1.2 System structure of timelined-based image panorama	7
Figure 1.3 System structure of timelined-based image panorama shows papers contribution in construction or addressing issues in different system subsystems.....	9
Figure 2.1 Ancient line pattern images.....	11
Figure 2.2 Various aspects of image.....	14
Figure 2.3 Digital Camera elements.....	17
Figure 2.4 Venn diagram for Shannon entropy and mutual information.	32
Figure 2.5 For each octave of scale space, the initial image is repeatedly convolved with Gaussians to produce the set of scale space images shown on the left.	38
Figure 2.6 To build the descriptor, an oriented quadratic grid with 4x4 square sub-regions is laid over the interest point (left).	40
Figure 2.7 Twelve-point segment test corner detection in an image patch.	41
Figure 2.8 Eye template.....	46
Figure 2.9 Finding face features by parts-based analysis.....	47
Figure 2.10 Applying distant transform.....	48
Figure 2.11 MRF network.....	53

Figure 2.12 Bayesian network.....	55
Figure 2.13 An example of factor graph.....	55
Figure 2.14 An illustration of the message passed in Belief Propagation.....	56
Figure 2.15 Basic set of 2D planar transformation.....	57
Figure 2.16 2D Spatial panoramas for the Parliament Hill building of Canada.....	58
Figure 2.17 3D projection onto 2D plane.....	59
Figure 2.18 Central projection, showing the relationship between the 3D and 2D coordinates.....	59
Figure 2.19 3D panorama picture of the center courtyard of Sony Center at the Potsdamer Platz in Berlin.....	60
Figure 2.20 Timelined multimedia components exploring the history of Steve Jobs in a historic perspective panoramic view.....	61
Figure 2.21 Screenshots of a timescapes location and timeline interface.....	62
Figure 3.1 Samples of the image data sets for the 11 landmark images containing mix of historic and modern images.....	65
Figure 3.2 Wavelets generated with three control parameters: numbers above each wavelet arranged as scale, frequency, and orientation respectively.....	68
Figure 3.3 Gabor family.....	69
Figure 3.4 Feature extraction process.....	71
Figure 3.5 Single perceptron neuron.....	72
Figure 3.6 Most popular transfer functions.....	72
Figure 3.7 Multi-layer feed-forward neural network.....	73

Figure 3.8 Coliseum of Rome image (a) The candidate feature area (b) Sample of training set for area 1.....	76
Figure 3.9 Dome of the Rock of Jerusalem image (a) The candidate feature area (b) Sample of training set for area 1.....	76
Figure 3.10 Whitehouse of Washington image (a) The candidate feature area (b) Sample of training set for area 1.....	77
Figure 3.11 Empire State Building of New York image (a) The candidate feature area (b) Sample of training set for area 1.....	77
Figure 3.12 Eiffel Tower image (a) The two candidate feature areas (b) Sample of training set for area 1(c) Sample of training set for area 2.....	79
Figure 3.13 Hagia-Sophia of Turkey image (a) The two candidate feature areas (b) Sample of training set for area 1(c) Sample of training set for area 2.....	78
Figure 3.14 Taj-Mahal of India image (a) The two candidate feature areas (b) Sample of training set for area 1(c) Sample of training set for area 2.....	79
Figure 3.15 Machu-Picchu of Peru image (a) The two candidate feature areas (b) Sample of training set for area 1(c) Sample of training set for area 2.....	79
Figure 3.16 Parliament Hill of Canada image (a) The three candidate feature areas (b) Sample of training set for area 1(c) Sample of training set for area 2 (d) Sample of training set for area 3.....	80
Figure 3.17 Pyramids of Giza image (a) The three candidate feature areas (b) Sample of training set for area 1(c) Sample of training set for area 2 (d) Sample of training set for area 3.....	81

Figure 3.18 Stonehenge of the United Kingdom image (a) The three candidate feature areas (b) Sample of training set for area 1(c) Sample of training set for area 2 (d) Sample of training set for area 3.....	82
Figure 3.19 Correctly detected features (a) One candidate feature images (b) Two candidate features images(c) Three candidate feature images.....	83
Figure 3.20 False positive detected features.....	84
Figure 3.21 Flow chart of operation sequence in image subset selection subsystem	88
Figure 4.1 Dense SIFT descriptor calculation for every image pixel of the original landmark image.....	93
Figure 4.2 Dense SIFT descriptor for the Coliseum landmark images	94
Figure 4.3 Dense SIFT descriptor for the Eiffel Tower landmark images	95
Figure 4.4 Dense SIFT descriptor for the Parliament Hill of Canada landmark images.	95
Figure 4.5 Eiffel tower (a) Image taken in 1889 (b) Image taken in 2013 (c) SURF/SURF matching of left keypoints with those on right image using Brute Force algorithm.....	98
Figure 4.6 Applying disparity gradient filter (DGF) on features extracted using SURF/SURF (a) Before applying DGF (b) After applying DGF.....	99
Figure 4.7 Applying HARRIS/HARRIS feature extraction on (a) images of 1880 with image of 2006 and (b) image of 1880 with itself for the Pyramid landmark	100
Figure 4.8 Keypoint matching of images of Eiffel tower (a) Bad matching in the pair 1889-2005 and (b) Successful matching in the pair of 1945-1967b.....	102

Figure 4.9 Applying BRIEF/BRIEF feature extraction on (a) images of 1993 and 2012 and (b) on images pair 1948 - 2009 for the Whitehouse landmark.....	104
Figure 4.10 Applying BRISK/BRISK feature extraction on the pair image of 1955-2009 for the Stonehenge landmark.....	104
Figure 4.11 High performance feature matching techniques accuracy comparison.....	108
Figure 4.12 Low performance feature matching techniques accuracy comparison.....	108
Figure 5.1 Image histograms for Coliseum landmark images (a) Image of 1910 (b) Image histogram (c) Image of 1913 (d) Image histogram.....	114
Figure 5.2 Dense SIFT images for images of Fig. 5.1 of Coliseum landmark images (a) Dense SIFT of image of 1910 (b) Image histogram (c) Dense SIFT of image of 1913 (d) Image histogram.....	115
Figure 5.3 Image histograms for: column (a) rows (1:1890 and 2:2007) Eiffel Tower (3:2011 and 4:2015) Dome Rock (5:1910 and 6:1959) Taj-Mahal landmarks with their histograms in column (b), their dense SIFT features in column (c), and histograms of dense SIFT in column (d).....	116
Figure 5.4 Image completion structure of MRF (a) Labels represented by w x h patches (b) Nodes and edges used during image completion.....	117
Figure 5.5 Messages propagation between nodes p and q (a) Node p makes use of all its neighboring nodes r,s,t to propagate a message to node q (b) Node p calculates its belief by collecting messages coming from its neighboring nodes.....	118
Figure 5.6 Node weight calculation to decide message priority.....	119

Figure 5.7 Modern to historic image registration (a) Historic image from 1889 (b)	120
Modern image from 2010 (c) Registered image. (d) Alignment energy....	120
Figure 5.8 Registered images of the Eiffel Tower landmark image set.....	121
Figure 5.9 Registered images of the Coliseum of Rome landmark image set.....	122
Figure 5.10 Registered images of the Taj-Mahal landmark image set.....	123
Figure 5.11 Texture Distribution (a) No texture image (b) Simple and easy transition	
texture image (c) Dense and overlapped texture image.....	124
Figure 5.12 Texture segmentation for an Eiffel Tower landmark image.....	125
Figure 5.13 Texture segmentation Row (1) Stonehenge landmark image, Row (2)	
Pyramids of Giza landmark image, Row(3) Empire State Building landmark	
image with Columns (a) Original image (b) Entropy texture image (c)	
Binarized texture image and (d) Holes and small areas removing image...126	
Figure 5.14 Warping images applied on the timeline of Eiffel Tower images from 1889	
to 2013.....	130
Figure 6.1 Screenshot of a WBGI module.....	133
Figure 6.2 Screenshots of an ISS module.....	134
Figure 6.3 Screenshot of an AIR module.....	135
Figure 6.4 Screenshots of a TPD module.....	136
Figure 6.5 Panorama software used on modern and historic images (a) PTGui Panorama	
Software used with images of Parliament Hill of Canada (b) ArcSoft	
Panorama maker used with images of Coliseum of Rome (c) Autopano Giga	
Panorama software on the Whitehouse images.....139	

Figure 6.6 Timescape panorama for Eiffel Tower landmark exploring the registered historic and modern images starting with reference image of 2002.....	141
Figure 6.7 Percentage ratio of SSIM and RMSE for the registered landmark images.	142
Figure 7.1 The Web-Based Geographic and Informatics Subsystem can be an expansion core for a general landmark timescape panorama.....	148
Figure A.1 Dense SIFT descriptor for the Dome Rock landmark images	150
Figure A.2 Dense SIFT descriptor for the Empire Building landmark images	150
Figure A.3 Dense SIFT descriptor for the Hagia-Sophia landmark images	151
Figure A.4 Dense SIFT descriptor for the Mach Picchu landmark images	151
Figure A.5 Dense SIFT descriptor for the Pyramids of Giza landmark images	152
Figure A.6 Dense SIFT descriptor for the Stonehenge landmark images	152
Figure A.7 Dense SIFT descriptor for the Taj Mahal landmark images	153
Figure A.8 Dense SIFT descriptor for the Whitehouse landmark images	153
Figure E.1 Registered images of the Dome Rock landmark image set.....	163
Figure E.2 Registered images of the Pyramids of Giza landmark image set.....	164
Figure E.3 Registered images of the Machu-Picchu landmark image set.....	166
Figure E.4 Registered images of the Empire State Building landmark image set.....	167
Figure E.5 Registered images of the Hagia-Sophia landmark image set.....	168
Figure E.6 Registered images of the Parliament Hill of Canada landmark image set...	169
Figure E.7 Registered images of the Stonehenge landmark image set.....	170
Figure E.8 Registered images of the Whitehouse landmark image set.....	171
Figure F.1 Screenshot for a user request to select a landmark for timescape display....	172
Figure F.2 Screenshot shows the reference image of the Coliseum of Rome landmark	172

Figure F.3 User navigation to older images, system responds showing registered images
with geographic information, historic information, and landmark coordinates

.....173

Figure F.4: User navigation to newer images, system responds showing registered images
with geographic information, historic information, and landmark coordinates

.....173

List of Appendices

Appendix A: Dense SIFT descriptors for landmark images.....	150
Appendix B: Landmark matching tables for HARRIS/HARRIS feature matching technique	154
Appendix C: Landmark matching tables for SIFT/SIFT feature matching technique	157
Appendix D: Landmark matching tables for SURF/SURF feature matching technique.....	159
Appendix E: Automatic image registration applied on landmark images	153
Appendix F: User Experience	172

Chapter 1

Introduction

1.1 Cultural Heritage Preservation

Cultural heritage is the legacy of sites, groups, or societies through tangible or intangible attributes which are inherited from past generations, maintained in the present, and safely entrusted to future generations [13]. All tangible attributes (such as buildings, landscapes, books, and artefacts) and intangible attributes (such as folklore, traditions, language, and knowledge) have to be kept from the past and presented to the future in reasonably acceptable prominence; this is known as cultural heritage preservation [13]. Since these objects; tangible and intangible, represent part of the study of history, the preservation process determines the importance of the past and the cultural objects that can validate the history [14].

As a part of the UNESCO interest in the world cultural heritage urged five things that highly support the preservation of cultural heritage [15]:

- Partner: any person or organization can be a UNESCO partner and get involved in preservation activities.
- Volunteer: participate in the volunteering or interning programs of the UNESCO.
- Travel: respecting global or local culture sites and customs by participating sustainable tourism that do not damage sites through visiting.
- Spread Awareness: Participate in the preservation process by introducing and sharing multimedia, news, or links through social networking or public sharing websites.

- Donate: financial support through donations is important to support these projects.

Our participation is to spread awareness and to stimulate virtual visitation opportunities by presenting a cultural heritage site or object through a technology. We do this specifically by introducing and implementing a timelined display of many famous landmarks around the world. Moreover, we have created a precise processing of historic images of these landmarks (captured with old analog cameras and digitized) in order to combine them with modern images (captured with modern digital cameras) to produce timescapes displays.

1.2 Panoramas

A panorama is a wide-ranging view or representation of a space or location. The formal definition of the term panorama [1] reveals two main concepts: space and time. The first concept as “an unbroken view of the whole region surrounding an observer” [1] harmonizes the spatial sense of the viewer, while the second concept as “a complete survey or presentation of a subject or sequence of events” refers to both the space sense by referring to the subject in addition to the temporal sequence of events presented to the viewer.

The panoramic image space (spatial panoramas) has had an extensive amount of research [1, 2, 3, 4] to provide a continuous, unbroken view of a place through image stitching techniques [2, 5, 6, 7] for images under varying different conditions of scale [8], illumination [9], rotation [8,9,10,11] and/or viewpoints [8,9]. However, all of this research uses a common camera, or restricts itself to modern camera data which greatly simplifies the problem compared to what the thesis attempts to solve.

The temporal sense of a panorama in terms of an evocation of history and

experiencing its details is conventionally available through the traditional methods of history lessons and the facts from history textbooks. This process may be facilitated now through the ubiquity of multimedia with its associated rich presentation capabilities. In essence, the ability for a person to experience the timeline of events for a given place brings more depth to the historical significance. So, sending a person standing at a specific place to a specific point in time may be a better learning experience through the contextualization of what has happened in that exact place by bringing hundreds of years of history associated with a specific location [12] directly to the user. This thesis realizes a portion of this lofty goal by implementing a timelined panoramic view that allows users to explore decades of history of a location or space through a panoramic display of registered (-or warped) images with both modern and (digitized) historical data coming from a variety of different cameras with different physical setups and parameters.

1.3 Historic Photos Challenges

Our landmark images dataset is comprised of a large number of historic and modern photos covering a period of time extending from the 1800's to modern time. Working on photos of widely different natures is a difficult challenge since the historic photos captured by an obsolete technology in which camera information is missing, used imprecise lenses, and were captured under poor capturing conditions [16]. Some of the factors that lead to poor quality historic photos are [16]:

- Historic photos captured by photographers using cameras of the pinhole era had very few control factors were not controllable at all.
- The aperture adjustments of pinhole cameras with or without lenses (Figure 1.1) differ in a way in which they provide a small control over light exposure.

- Because of the pinhole cameras of infinite depth of field, the pinhole size is a problematic issue in terms of size selection. Smaller size is assumed to produce sharper image resolution while an extremely small size leads to less clear images and diffraction conditions due to light wave properties [17]. As well, as the hole size reaches the material thickness in which it punched the vignetting phenomena becomes more apparent.
- The pinhole cameras in their basic formulation has no lenses (Figure 1.1-a) or the newest generation with lenses (Figure 1.1-b) in which the aperture adjustment differs in a manner which provides a little control on the light exposure [16].

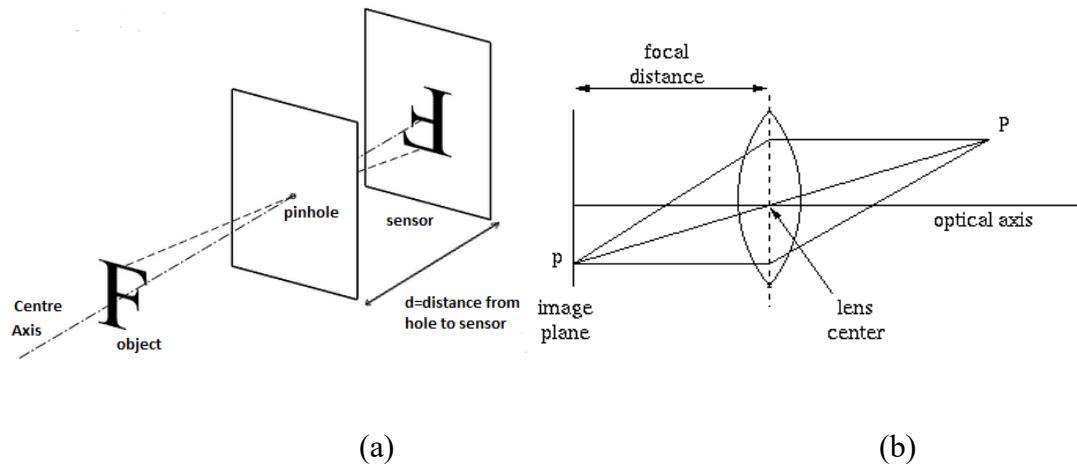


Figure 1.1: Pinhole camera (a) Geometry of pinhole camera without lens (b) Pinhole camera with lens [16]

1.4 Thesis Objectives

In this thesis, our main objective is to introduce and develop techniques that facilitate the new timelined panorama concept known as a timescape. Our contribution is to participate effectively in cultural heritage preservation via the implementation of a

timelined image display of historic and modern images of many landmarks around the world with a temporal panoramic view of these landmarks that merges geographical and historical information. The automated alignment of historical and modern photos has not been adequately addressed in the literature to facilitate such an objective.

The achievement of this goal requires:

- (1) Hundreds of thousands of images of famous landmarks from around the world including both historic and modern images have to be collected from the Internet through websites like Flickr and Google Images.
- (2) Image subset selection: From the available data set, which is organized as subsets of landmarks, we have to select a roughly same view for each landmark, and for each landmark we have to select images of good quality. This is the input conditioning problem.
- (3) Aligning images through:
 - Image Warping: the mapping between images is performed depending on all features matched so that warping to a single view point is possible. It is expected that this problem is effectively solved; traditional image warping techniques will suffice if accurate feature matching can occur, or
 - Image Registration: modern to historic image registration by overcoming challenging problems of using different various generations of technologies of cameras under different capturing conditions.
- (4) Timescape Viewing: the time-based panorama presentation achieved through a multi-layer viewing mechanism allowing the navigation from the past to

the present by offering an opportunity for the end users to scroll through time. The added dimension of time to a panorama viewer creates an extra dimension of control for the user that needs contemplation and study.

The realized system provides an opportunity for the user to navigate from the past up to the present for many landmarks around the world through a high quality registered images display. This navigation illustrated in a detailed system performance showed by screenshots of the system with the interaction steps between the user and the system are shown in Appendix F: User Experience.

1.5 System Structure

To realize these objectives, a system has been designed, implemented, and tested. The system structure is shown in Figure 1.2. The first step collects the historic and modern images of many landmarks around the world from the Internet. Then, the system classifies these images to select a roughly singular viewpoint for each landmark. In the next stage, feature extraction is completed for these images to prepare them for the feature correspondence step. The images are then aligned through image warping or image registration to demonstrate the best correspondence between each pair of images.

Next, the image registration step is performed in a pairwise image registration method that attempts to minimize visual artefacts. The final stage in the process is the use of a display system for viewing temporal panoramas.

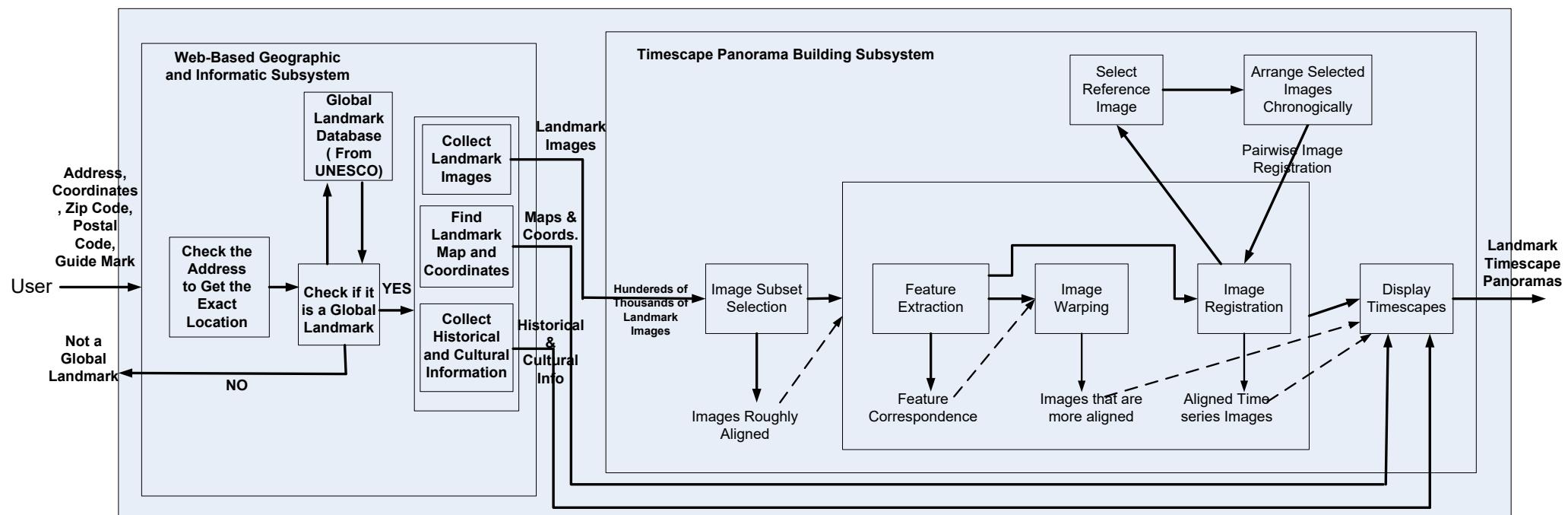


Figure 1.2: System structure of timescape image panorama

1.6 Thesis Contributions

The main contributions of this thesis are summarized as follows:

- We develop and implement a new timelined panorama concept. The temporal panorama that displays a timeline series of spatially aligned images in a chronological order comprised from historic and modern images of famous landmarks around the world in a timescape.
- An effective participation in the world efforts of cultural heritage preservation by providing and sharing our work presenting many world landmarks as temporal panoramas.
- We propose a modern to historic image matching by exploiting the facilities of standard image feature detection and extraction techniques and present the ORB/SURF image matching technique.
- We introduce a highly accurate historic to modern image registration technique that is based on finding the optical flow of the extracted SIFT features which outperforms the standard techniques and available state-of-the-art panorama software packages.
- The following papers are the outcome of this research: [16, 26, 31, 76, 98, and 104]. Parts of this thesis have been published in these papers. These papers contributed in the construction or addressing issues in different sections and subsystems of our system as illustrated in Fig. 1.3.

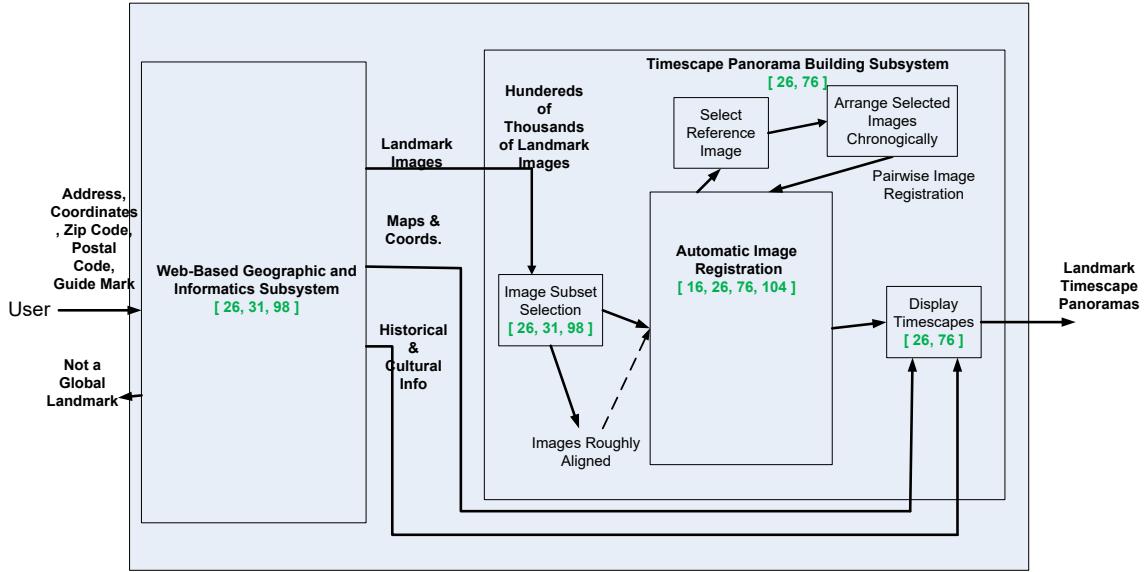


Figure 1.3: System structure of timeline-based image panorama shows papers contribution in construction or addressing issues in different system subsystems

1.7 Thesis Organization

The rest of the thesis is arranged as follows. In Chapter 2 we cover the background of our work and the related work. Chapter 3 emphasises the image subset selection and the process of image classification of all the landmarks. In Chapter 4, our two methods in feature extraction and matching which are ORB/SURF and optimized optical flow of the extracted SIFT features is examined. We explore all the steps and explain the effectiveness of these proposed methods by reviewing our results by comparing them with standard techniques and methods. The two aligning techniques examined in this thesis, which are warping and registrations are presented in Chapter 5 showing their applicability to our dataset. In addition, we examine the quality of the results of the image registration process based on image similarity measures. Chapter 6 integrates all the steps and shows examples of the temporal panoramas of landmarks. Chapter 7 concludes the thesis and discusses the future prospects of the work contained herein.

Chapter 2

Background and Related Work

In this chapter, we will review the importance of the historic photos and documents, as well as the researchers, work to collect, process, and analyze these documents. In addition, we will demonstrate the effective technological attempts to digitize these documents and preserve them in an appropriate form as part of the international cultural heritage.

Then, the photographic technology evolution will be reviewed showing the significant happened photographic developments and their effects on the historic and modern photos.

The automatic image registration process will be explored later. Classification of registration techniques based on different criteria will be shown. This implies the study of the similarity techniques used in conjunction with and as part of image registration methods. Feature extraction and matching will be reviewed thoroughly through the classification of feature extraction techniques. A comprehensive comparison of the standard feature extraction will explore the basic properties of these techniques as well as powerful, and weakness points of these techniques will be shown.

Optimization process in image registration will be described and the widely used optimized techniques in image registration will be explained showing their important features.

Study of the image subset selection and classification is presented through the exploring the basic techniques and a literature review.

Finally, we demonstrate the systems, techniques, and related technology that interpret the different panorama concepts. This reviews the panorama concept evolution from pure spatial panoramas, to temporal panoramas leading to our newest concept; timescapes, which merges the spatial and temporal aspects in a single timeline based display. The multimedia and its role in panorama building will be discussed, as well, in this chapter.

2.1 Historical Photos and Documents

Interest in historical photos and documents including the ancient printed documents increased because of their precious value cultural heritage. These historical materials are challenging in terms of document image analysis and recognition because of their features of capturing technology and complex patterns [56]. Researchers struggle in exploiting all features of photos and documents branched into two main categories depending on the nature of the documents; line pattern images and historical photos contents. The line pattern images composed mainly from a combination of different line patterns to form shapes and backgrounds, as shown in Fig. 2.1.



Figure 2.1: Ancient line pattern images [56]

These historical assets are generally exhibited in museums and art galleries. Digital

libraries attempt to provide digitized copies of these images to historians and researchers by preparing searchable database. The early retrieval system offered is keyword-based. This represents the difficulty for users since they have to understand and be able to describe the characteristics of ancient images they are looking for, which may be different from the ones used to describe these images. Whereas the scientific ground truth was still inadequate for such complex ancient line pattern images, not much relevant indexing works exist. The work proposed in [57,58] primarily based on Zipf's law in indexing and classifying the images depending on predefined categories according to their line patterns. This approach was mainly performing the indexing and classification rather than retrieving. Thi Thuong et.al [59] relied on a bag of strokes approach which is a time consuming, while Uttama et.al. [60] based on texture feature to propose a full segmentation algorithm which requires high computing cost and retrieval lag.

Alternatively, a content-based image retrieval system (CBIRS) provided an interesting proposal to this problem. It comprises of two parts: indexing and retrieval. Indexing is primarily done for feature extraction and description while the retrieval performed based on feature matching (Extraction). Coustantay [56] based on partial density to partially segment line pattern images for CBIRS in indexing technique to compensate the computation complexity and lack of ground truth.

Historical images and maps represent an important information source for different disciplines. Historical maps are useful for surveyors, geographers, and historians who are interested in the background of an area [61]. Livieratos [62] explored the historical stages of geographical area representation starting from Ptolemy cartography who introduced the parameterization of maps in terms of spherical coordinates to Greek ideas and

practices based on rational geometry coordinates. Then, he proposed a unified, digital, computational, and graphical domain techniques mentioned the term “infographics” based on Ptolemy’s representations, the orientation of significant base lines or alignments, the shapes of coastlines, the best-fitting of regional areas, and the comparison of two maps belong to the same or different map family.

Aerial images and maps added an evolutionary source to be used in evaluating and analyzing historic images. Techniques for detection and interpretation objects have been developed to detect contours of man-made objects as linear segments [89], semantic labeling of images [63], and detection of high-level meaningful structural changes [64]. Many software packages and algorithms have been developed for historic and modern map matching and GPS tracking like Automatic Correlation Map (ACM), 2W, European Galileo system and the European Geostationary Overlay Service (EGNOS) [65].

Historically, images used to stimulate two things of audiences. First, is the incredible variety of things that go by image name. This may be pictures, statue, projections, poems, maps, optical illusions, diagrams, dreams, memories, and even ideas as images [66]. This list would impossible make any unified and systematic understanding. The second is that the calling all of these things by the name image does not granted mean that they have something in common [66]. This can be summarized in the following aspect of the image in Fig. 2.2 below:

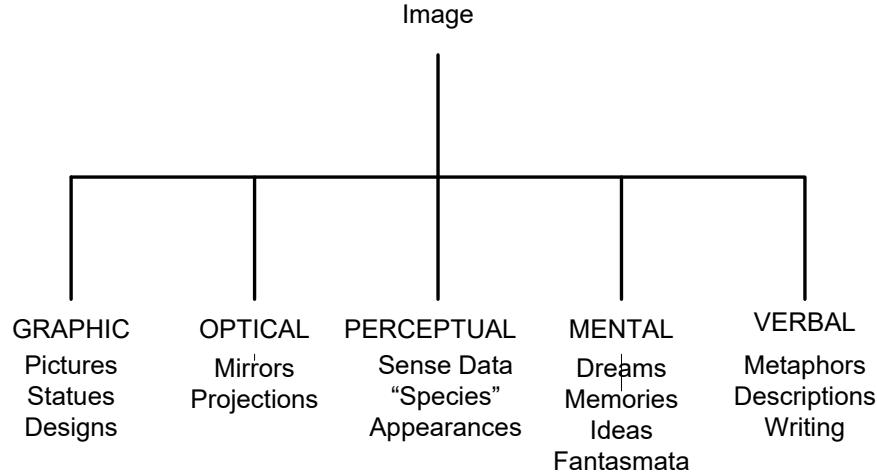


Figure 2.2: Various aspects of image [66]

The landmark historic image when centered as a milestone for a chronological line of images up to present view image may stimulate the graphical, perceptual, mental, and verbal aspects of an image. This view displayed as a timescape panorama presents in a pictorial image of the graphic aspect which retrieves memories and ideas about the landmark. It also may offer many appearances of the landmark in one harmonic panoramic view. Finally, this image view may reveal architectural developments occurred during the view timeline or shows defects may have happened due to the environmental or other conditions.

2.2 Photographic Technology Evolution

Our timeline of photos consists of images from the early 1800's up to recent time. These photos were captured by different photographers ranging from professionals, who professionally captured images using their skills and experience, to other images that may have been captured by amateurs. Both types of photographers used the available photography technology of their time. Cameras, which have undergone significant

technical changes, substantially affect the quality of the captured images.

The different stages of photographic technology evolution and camera development, the characteristics of each era, and the control factors that affect the photo quality will be investigated to evaluate the photo sets and to suggest the best criteria for dealing with poor quality images to extract their important features.

Photography and camera technology have passed through many stages of developments over generations of photographic evolution including:

- Camera obscura: an optical device dating back to ancient China and ancient Greece [32] based on the pinhole phenomenon to project an image of the scene upside-down on the viewing surface. This description is based completely on the pinhole camera (Fig. 1.1-a) which is a simple camera without a lens and a single small aperture, described by Ibn Al-Haytham (Alhazen) [33] who clearly reasoned the object image was appearing upside-down onto the viewing surface. In a pinhole camera, the method of calculating the pinhole diameter was derived by Jozef Petzval [34]. The pinhole size is determined by the formula:

Where d is the pinhole diameter, f is the focal length (distance from pinhole to image plane) and λ is the wave-length of light.

The next evolutionary step was the addition of lenses to pinhole camera which creates a much larger hole through which light can make it onto the film. The result of this was that more light could enter the camera and the film can be exposed faster (Fig. 1.1-b).

- Daguerreotypes cameras were introduced in 1839 and were the first publicly announced photographic process to come into widespread use [35].
- Calotypes were an early photographic process using paper coated with silver iodide introduced by W. Fox Talbot. It needed a very long exposure (for an hour or more) to produce a negative [36].
- Dry plate, also known as gelatin process was invented in 1871 by Dr. Richard L. Maddox. In 1879, the first dry plate factory was established [37].
- Photographic film is a sheet or strip of transparent plastic film base coated with a gelatin emulsion containing a very tiny light-sensitive silver halide crystal on one side. The characteristics of these crystals like the size determine the contrast, sensitivity, and resolution of the film [38].
- Digital cameras encode images and video digitally and store them for later projection or reproduction. Digital and film cameras share an optical system by using lenses with a variable diaphragm to focus light onto an image recording device [39]. The shutter and diaphragm permit the suitable amount of light to the imaging surface similar to film, but the image recording device is electronic rather than chemical (Fig. 2.3).

The processes of capturing images from as early as 1839 to as recent as 2016 involve a variety of image capture mechanism from physical, chemical and electronic. In all cases, the digitization process is affected by the capture process.

Because of the varying mechanisms for capturing an image inherent in a timeline-based data set, the underlying data that makes up these images (once

digitized) do not necessarily maintain the same properties. As such, it becomes difficult to match patterns in the image data using traditional techniques that imagine a homogenous capture system.

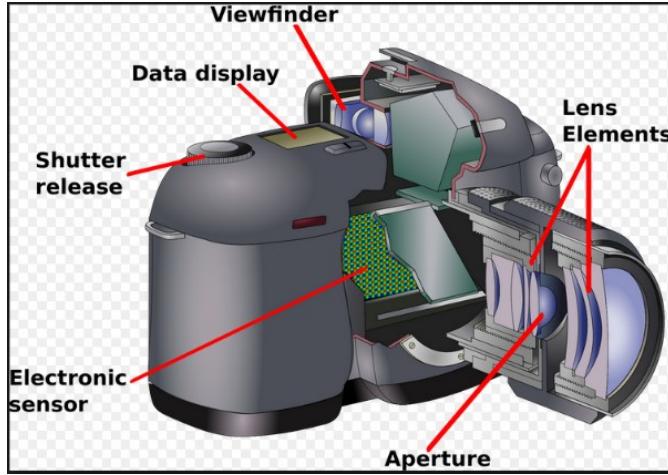


Figure 2.3: Digital Camera elements [40]

2.2.1 Technical Aspects of Photography

Photographers control the camera by controlling many factors and adjusting the amount of light exposure the imaging medium is exposed to; creating the latent image on the film, or via electronics (CCD or CMOS for example). In all cameras the process of obtaining a suitable exposure involves a few controls; manually or automatically, to ensure producing a clear, sharp and well-illuminated photo [38]. The controls include but are not limited to:

- Focal length: the distance from the lens center to the image plane.
- Aperture: The adjustment of lens opening and represented by the f-number, which controls the amount of light passing through the lens and is calculated by dividing the distance from the pinhole to the imaging plane (the focal length) by the diameter of the pinhole. Aperture affects on the depth of field, the higher the f-number, and

the smaller the opening, the greater the depth of field and hence more diffraction blur [41]. The effective aperture diameter is found by dividing the focal length by the f-number.

- Type of lens: normal, long, focus, wide angle, telephoto, macro, fisheye, or zoom [42], lenses affect the overall light hitting the imaging media.
- Focus: Adjusting the position of a viewed object relative to the camera as in focus or out focus. Focus affects the sharpness of a captured image.
- Shutter speed: The shutter controls the amount of time during which the imaging medium is exposed light for each exposure [42]. This affects the contrast and illumination of the resulting image.
- Metering: Controlling the highlights and shadows according to the photographer desire by the exposure measurement. Artistic desire will affect the image quality.
- ISO speed: Numerical indicator of the gain from light to adjust the automatic exposure system [42].
- Inherent sensitivity of the recording medium to the light color/wavelengths and intensity [42]. This affects required exposure times and affects the imaging processes with motion blur and other speed-based artifacts.

2.2.2 Exposure and Rendering

These camera control factors are interrelated and the exposure changes with:

- exposure duration
- aperture of the lens, and
- the effective focal length of the lens

To model these control factors mathematically, we would end up with an underdetermined system because we have multiple variables (such as those listed above) and only a single image with which to determine the exact configuration.

2.3 Image Subset Selection

Image subset selection or classification is an important and challenging task in various different applications including computer vision, biomedical imaging, surveillance systems, robot navigation, aerial and remote sensing, and robot navigation [90].

Classification images vary depending on the classification basis and the images nature. Therefore, many classification approaches and techniques exist.

2.3.1 Image classification approaches

Image classification approaches are classified depending on [90]:

- Basis of characteristic used:
 - Shape-based: common features used are points, primitive geometric shapes (rectangle or ellipse) skeleton, silhouette, and contour of 2D spatial objects [90].
 - Motion-based: relies on tracked temporal features of objects [90].
- Basis of training sample used:
 - Supervised classification: the process of using samples of training sets of known informational classes to classify unknown objects [90].
 - Unsupervised classification: a method in which a large number of unknown features examined and divided into a number of classes based on natural

grouping available in the image values with no extensive prior knowledge is required. K-means is an example [90].

- Basis of assumption of parameter on data:
 - Parametric classifier: based on an assumption of Gaussian distribution, the parameters like covariance matrix and mean vector are used. Maximum likelihood and linear discriminator are examples of parametric-based techniques [90].
 - Non-parametric classifier: what characterizes these classifiers is no use of statistical parameters to estimate a class and no prior assumption about the input data is required. Examples of such techniques are decision-tree classifier, artificial neural networks (ANN), support vector machine (SVM), and expert systems [90].
- Basis of pixel information used:
 - Per-pixel classifier: in this technique, a classifier signature generated by using the combination of the spectra of all training set pixels for a given feature and all the elements presented in the training set pixels will participate in the outcome signature of a feature. It can be parametric or non-parametric [90].
 - Sub-pixel classifier: the spectral value of each pixel is defined by pure elements called end numbers. This may be a linear or nonlinear combination of these end numbers, which provides a relative membership of each pixel to each end number [90].

- Object-oriented classifiers: pixels are merged to form objects to facilitate classification based on objects [90].
- Basis of number of outputs for each spatial element:
 - Hard classification: each pixel is enforced to show a membership to a single and only single class as in minimum distance, ANN, decision-tree, and SVM [90].
 - Soft classification: each pixel may show multiple and partial class membership. This is called a fuzzy classification [90].
- Basis of spatial information:
 - Spectral classifier: based on pure spectral information. Examples are ANN and maximum likelihood.
 - Contextual classifiers: uses the spatially neighboring pixel information. Frequency-based contextual is an example.
 - Spectral-contextual classifiers: uses both spectral and spatial information.

2.3.2 Image Classification Techniques

Image classification techniques can be classified as:

- Decision tree (DT): based on hierarchical rules methods and computes a class membership by recursive partitioning a dataset into subsets. This technique allows accepting or rejecting class labels at each intermediate repetition cycle. DT techniques based on partitioning the set into nodes, find the terminal nodes and the allocation of specific data set of class label to terminal nodes [90].
- Support vector machine (SVM): support vector machines or networks are supervised learning models with accompanying learning algorithms used in

machine learning for data classification and regression analysis. In SVM, a hyperplane or set of hyperplanes in a high- or infinite-dimensional space can be used. Most likely, a good separation is achieved by the hyperplane that has the largest distance to the nearest training-data point of any class (so-called functional margin), since in general the larger the margin, the lower the generalization error of the classifier.

- Artificial neural networks (ANN): ANNs are a family of models inspired by biological neural networks (the central nervous systems of animals, in particular, the brain) which are used to estimate or approximate functions that can depend on a large number of inputs and are generally unknown. To produce a classifier using an NN, a series of examples of input data and constructing a network that has one output for each array of inputs of its type [91]. The selection of the training (or learning) algorithm depends on many factors such as [99]:
 - the number of inputs to the networks
 - the number of weights and biases in the network
 - the complexity of the problem
 - the error goal
 - the network use in discriminant analysis like pattern recognition, or a function approximation like regression
- Moreover, the learning algorithms can be classified as [100]:
 - Supervised learning: the learning algorithms classified under this category if the desired output is known and provided with the input while training phase.

- Unsupervised learning: in this category of algorithms, only the inputs are given through the training phase, and the network is responsible for finding some kind of the output pattern with the support of inputs and no external aid.
- Reinforcement learning: similar to supervised learning with some feedback is given, and instead of providing an explicit output, a reward is given based on how well the system performed.
- Fuzzy measure: in fuzzy classification, stochastic relationships are calculated to describe the features of an image. The various types of stochastic relationships are incorporated (set of properties) in which the members of this set of properties are fuzzy in nature. There exist some different classes of fuzzy measures including plausibility/belief measures; possibility/necessity measures; and probability measures which are a subset of classical measures. [90].

2.3.3 Image Selection Literature Review

The selection of image subsets from a larger set depends widely on the image type and contents. A thorough analysis of image contents has to be done to extract the salient features and characteristics that are common in an image set. This extraction depends on the conditions under which the image was taken [43]. Traditional image set selection and classification mainly rely on the analysis of the low-level features of the image to determine high-level content semantics [49]. To optimize the feature detection and extraction process, the researchers have worked to support their efforts using non-traditional techniques. Machine learning techniques like artificial neural networks (ANN) has obtained popularity in image subset selection [50].

Xiong et. al. [51] proposed a back propagation neural network to improve the performance of image selection by segmenting and clustering the image into several visual objects and building the total feature vector, while Xu and Qu [52] developed a method based on feature matching similarity and frequency-inverse document frequency (TF-IDF). Multi-scale features used by Li et. al. [53] in medical image classification where the classifiers use a set of complementary image features in various scales to compare the results of the classification process.

To optimize the feature selection process for a robust image classification, Al-Sahaf et. al. [54] introduced a Two- Tier Genetic Programming (GP) based image classification method that works on raw pixels rather than high-level features, while Kharrat et. al. [55] also uses a genetic algorithm for the selection of MR brain images using wavelet co-occurrence.

2.4 Automatic Image Registration

Image registration is the process of geometric transformation between two images such that one image is best aligned with the other after the transformation. This geometric transformation has to be optimally done to get the best registration. The optimality degree is oscillating depending on the images taken from different sensors, at different times, and/or different perspectives, but are taken of the same object or the same scene [67].

More precisely, for two images, reference or original image $I(x)$ and sensed image $S(x)$, image registration tends to calculate a spatial transformation $u(x)$ in the working space V such that the energy function is minimized:

Or correspondingly to solve the following optimization problem [67]:

$$\min_{u(x) \in V} J(I(x), S o h(x)) + \lambda \cdot R(u) \quad \dots \dots \dots \quad (2.3)$$

Where J is the dissimilarity term measure between the reference image $I(x)$ and the factor $S \circ h(x)$ which is the sensed image $S(x)$ after applying the transformation $h(x)$. λ is a parameter balances the regularization term R and the dissimilarity term J . The smaller λ is, the small weight will be on $R(u)$ and the less smoothness the deformation $u(x)$ will have. For the affine or rigid transformation, the term $\lambda \cdot R(u)$ will be removed from the energy function since the regularization already represented by the transformation parameters [67].

2.4.1 Registration Methods and Classification

Depending upon the given data and techniques utilized for registration, there are several diverse approaches to classify registration models [67].

2.4.1.1 Automation Degree

Registration methods might be characterized by the level of automation they give. Manual, interactive, semi-automatic, and automatic methods have been utilized. Manual registration develops tools enable the user to align images manually. Interactive methods diminish user inclination by performing certain operations automatically while still depending on the user to control the registration. Semi-automatic registration performs more of the registration steps automatically, however, relies upon the user to confirm the correctness of a registration process. Automatic methods perform all the steps and don't permit any user cooperation.

The highest level of automation; automatic registration, represents a challenge in

cases of multisensor, multitemporal, and multispatial image set, which is the case of our research of historic to modern image registration [56].

2.4.1.2 Single and Multi-Modal Methods

Images captured using the same capturing devices or cameras are single-sensor or single-modality images. They can be in linear relation after registration.

2.4.1.3 Linear and Nonlinear Transformation

Image registration process can be classified depending on the transformation type as linear or nonlinear. The linear transformation may include scaling, rotation, and affine transformation. The nonlinear transformations characterize the nonrigid or elastic changes and allow the local warping to align the reference image with the target image. These transformations include the thin-plate or surface splines and large deformation models [56].

2.4.1.4 Spatial and Frequency Domain Methods

Spatial and frequency domain transformations are characterized by the image source used in the registration process. Spatial methods work in matching features or intensity patterns, image domain between the reference image and sensed image. While in frequency domain methods, the original image transformed into K-space data are transferred by Fourier transform. Spatial domain facilitates similarity measures used for image registration such as cross correlation, correlation ratio, and mutual information. The registration process in K-space domain used to determine the transformation parameters between images in simple transformation such as translation, scale, and rotation [56].

2.4.1.5 Parametric and Nonparametric Image Registration

The geometric transformations nature, and hence, the possibility of the precise acquisition of this transformation will determine the image registration as parametric or nonparametric. The parametric registration approach aims to find all the optimal parameters which define the transformations used in image registration, while the nonparametric approach works to determine dense deformation field in nonparametric fashion required in image registration process.

2.4.1.5.1 Parametric Geometric Transformations

Parametric geometric transformations mainly comprise a rigid transformation, affine transformation, and projective transformations. These transformations are relatively easier to be calculated compared to nonparametric transformations [67].

2.4.1.5.1.1 Rigid Transformations

Rigid transformations are characterized by the geometrical transformations that retain distances for every pair of points, as well as, retain the straightness of lines and all nonzero angles between straight lines [67]. Rigid transformations include translations, rotations, reflections, and/or their combinations.

Rigid registration problems can be figured out by two components, a translation followed by rotation. To formulate the three dimensional case in rigid transformations, let R be the rotation matrix and let t be the translation vector of the reference image x and the sensed image y :

The rotation R can be represented by three angles of rotation about the Cartesian axes θ_x , θ_y , and θ_z :

$$R = \begin{pmatrix} \cos\theta_z & -\sin\theta_z & 0 \\ \sin\theta_z & \cos\theta_z & 0 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} \cos\theta_y & 0 & \sin\theta_y \\ 0 & 1 & 0 \\ -\sin\theta_y & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_x & -\sin\theta_x \\ 0 & \sin\theta_x & \cos\theta_x \end{pmatrix} \quad (2.5)$$

The translation t can be represented by three components:

$$t = \begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix} \quad \dots \quad (2.6)$$

So, the registration process can be performed by finding the optimum solution to the following problem:

$$\min_{\theta_x, \theta_y, \theta_z, t_x, t_y, t_z} J(I(x), S(Rx + t)) \quad \dots \dots \dots \quad (2.7)$$

2.4.1.5.1.2 Affine Transformations

Affine transformation is a linear mapping method that preserves points, straight lines, and planes. Sets of parallel lines remain parallel after an affine transformation. It retains the planarity of surfaces and parallelism, but it allows angles between lines to change.

All rigid and linear transformations are affine transformations, but rigid transformations which comprise a translation are not linear transformations.

Affine transformation is composed of two functions: a translation and a linear map, if the translation is represented as an addition of vector b and the linear map as a multiplication by a matrix A , the affine transformation acting on image x can be represented by:

So, the translation and the linear map can be represented using augmented matrix and augmented vector in a single matrix multiplication as:

$$\begin{pmatrix} y \\ 1 \end{pmatrix} = \begin{pmatrix} A & b \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ 1 \end{pmatrix} \dots \quad (2.9)$$

The registration process can be obtained by solving the following optimization problem:

$$\min_{a_{ij}, b_1, b_2, b_3} J(I(x), S(Ax + b)) \quad \dots \quad (2.10)$$

2.4.1.5.1.3 Phase Correlation

Image registration using the phase correlation method based on the frequency domain.

For a pair of two images to be compared, the phase correlation finds a third image contains a single peak. The location of this peak refers to the relative translation between these two images. Phase correlation is resilient to noise, occlusions, and other defects of multi-modal images [67].

2.4.1.5.2 Nonparametric Geometric Transformations

In deformable image registration as each image pixel is transformed independently, the nonparametric transformation technique is applied. Nonparametric transformation relies on a global objective function which composed of two terms: an alignment measure and an external regularization term which shows the expectations by compensation unpredictable transformations.

Nonparametric transformations applied in deformable image registration are more flexible in applications and images than those used for rigid or affine transformation. The deformable image registration methods are characterized by more computational complexity than parametric transformations [67].

2.4.2 Similarity Measure Criteria

Image similarity measure evaluates the similarity level of the intensity patterns in two images [56]. The selection of an image similarity measure highly relies on the nature of the scene or the location of captured images, the modality of the images to be registered, and on the application in which the registration process is applied. Also, the similarity

measures of images may depend on several relevant properties [67]:

- Reflexivity: the similarity measure for identical images supposed to be 1.
- Symmetry: the outcome of the similarity measure of two images has to be independent of the order in which the images are compared.
- Reaction to lightening and darkening: if an image has been lightened or darkened with a constant value, the similarity measure should return a high value for identical images.
- Reaction to noise: a practically trusted similarity measure should not be affected too much due to noise, and two same environment images should be decreased proportionally to an increasing noise percent.

So, image similarity measure can be classified depending on the images to be registered as linear relation or nonlinear relation.

2.4.2.1 Similarity Measures for Linear Relations

The simple correlation measures can be applied readily to images in linear relation. Cross correlation (CC) has been extensively applied as a robust matching function: It has been used to measure the linear dependency between image intensities depending on variances, means, and covariances of the images. The most direct and conditional similarity measure is the sum of squared differences which is applied to directly compare intensities of the images [56].

2.4.2.1.1 Sum of Squared Differences

This similarity measure can be used with images have the same range of intensity values and these intensities can be compared directly.

2.4.2.1.2 Histogram Difference Measure

This similarity measure is calculated by the application of suitable weighting function into the values of a difference-image grey-value histogram and finds the summation over all bins of the histogram. This measure is ideally suited for measures motivated by the fact that the optimal registration of mask and contrast image the difference image shows low contrast variation while in misregistration case, the contrast variation is larger [68].

2.4.2.1.3 Cross Correlation

It is a widely used measure as a robust comparison function in image similarity measure. Cross correlation being used as a linear dependency measure between image densities relied on the means, variances, and covariances of the images to be matched [56].

2.4.2.2 Similarity Measures for Nonlinear Relations

Multimodal images applications relied on a nonlinear relation of image intensities required to measure image similarity are directed toward nonlinear relation measures. Historic to modern registered images similarity are of nonlinear nature.

2.4.2.2.1 Correlation Ratio

Correlation ratio is defined as the ratio of two standard deviations representing the relationship between the statistical dispersion within individual categories and the dispersion within the whole samples [56].

2.4.2.2.2 Structural Similarity Index (SSIM)

The structural similarity index measures the similarity of two images depending on the degradation of structural information. The SSIM file is a full reference metric; which

provides an estimation or forecast of picture quality depends on an underlying uncompressed or contortion-free picture as reference. SSIM is intended to enhance traditional techniques, for example, peak signal-to-noise ratio (PSNR) and mean squared error (MSE), which have turned out to be conflicting with human visual recognition [75].

2.4.2.2.3 Root-Mean Square Error (RMSE)

The Root Mean Square Error (RMSE) which calculates the differences between the predicted values (reference image) and the observed values (sensed image) and it is widely used as a measure of accuracy [76]

2.4.2.2.4 Mutual Information

Mutual information is a measure of statistical dependence between two data sets; specifically, two images. This concept has been derived from information theory. It measures the amount of shared information in two images, or in other words, how the intensity distribution of two images is independent. For two images, reference image X and sensed image Y , the mutual information is defined as:

$$MI(X, Y) = H(X) + H(Y) - H(X, Y) \quad \dots \quad (2.11)$$

Where H is Shannon Entropy. MI can be also defined in terms of joint probability density function (PDF) and its marginal PDF for images X and Y which are $p_x(x)$ and $p_y(y)$ [56]:

$$MI(X, Y) = \int p(x, y) \log \frac{p_{x,y}(x,y)}{p_x(x) \cdot p_y(y)} dx dy \quad \dots \quad (2.12)$$

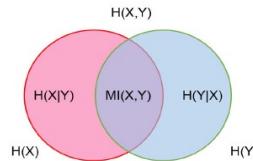


figure 2.4: Venn diagram for Shannon entropy and mutual information. The left disc represents the entropy of X and the right disc represents the entropy of Y . The mutual information is the overlapped area of the two discs.

2.5 Feature Extraction and Matching

Feature extraction and matching are essential in many computer vision applications. The extraction process concerns object or scene recognition while feature matching finds the correspondence between a pair or set of images. To get a robust matching, we generally seek invariant properties [43] so that the extraction result does not vary according to chosen (or specified) conditions. As an essential invariant, we are looking for immunity to changes in illumination level, scale, location, rotation, or viewpoint [43]. However, there will always be noise in the images at varying levels. Before we can extract features, we have to detect them. The extraction is more complex than detection since it implies a description of the detected shape such as its position, size, and direction [43].

Feature extraction process can be classified as low-level feature extraction and high-level feature extraction depending on the basic features to be extracted. Low-level features are those features that can be extracted automatically without any shape information, i.e. no information about the spatial relationships, while high-level feature extraction concerns mainly with finding shapes and objects [43].

2.5.1 Low-Level Feature Extraction

The basic approaches used in low-level feature extraction can be summarized as:

2.5.1.1 Edge Detection

- First-order edge detection: Image processing techniques based on edge detection exploit the fact that edge detection is insensitive to change in the overall illumination level. Edge detection highlights image contrast, which is a difference in intensity and can highlight the boundaries of the features within an

image. First order differentiation can be used to detect the edge position. Differencing adjacent points horizontally will detect vertical changes in intensity, while the horizontal edge is detected by differencing vertically adjacent points. For an image P , the vertical edges E_x can be found as [43]:

$$E_{x,y} = |P_{x,y} - P_{x+1,y}| \dots \quad (2.13)$$

While the horizontal edges E_y can be found:

$$E_{y,x} = |P_{x,y} - P_{x,y+1}| \dots \quad (2.14)$$

The first-order edge detection operation detects the peak where the rate of the change of the original signal occurrence. There are many edge operators used for edge detection like Prewitt, Sobel, and Canny edge operators.

- Second-order edge detection operators: The higher order derivatives are applied to the same cross section of data where the rate of change of the signal is greatest and zero when the rate of change is constant [43]. The zero crossing in the second order derivative occurred at the peak of the first-order derivative when the rate of change is constant. The basic second derivative operator is the Laplacian operator:

$$f''(x) = f'(x) - f'(x+1) \dots \quad (2.15)$$

Edge features matching reveals some of inherent problems such as incomplete contours, the need for selective thresholding, and their response to noise.

2.5.1.2 Phase congruency

The phase congruency is a feature detector technique which is derived by a frequency domain and operating on the considerations of phase. It offers two main advantages: it is invariant to local and smooth illumination change, and it can detect a broad range of

features [43]. The original concept of phase congruency which is the local energy was derived with links to the human visual system [69]. The phase congruency PC measure can be derived as:

$$PC(x) = \max_{\emptyset_u(x) \in 0, 2\pi} \left(\frac{\sum_u |F_{P_u}| \cos(\emptyset_u(x) - \bar{\emptyset}(x))}{\sum_u |F_{P_u}|} \right) \dots \quad (2.16)$$

Where $\emptyset_u(x)$ represents the local phase of the component F_{P_u} at position x . The value of $\bar{\emptyset}(x)$ that maximizes this equation is the amplitude weighted mean local phase angle of all the Fourier terms at the point being considered [43].

2.5.1.3 Localized feature extraction

Detecting image curvature or corner extraction represents other low-level feature that can be used in computer vision. Curvature can be described as the rate of change in edge direction and this rate characterizes the points in a curve. Points where edge direction changes rapidly are corners while points where there is little change in edge direction represent straight lines [43]. The normal way to measure curvature in the digital image is to compute the difference in edge direction between connecting pixels forming a discrete curve. This can be described as:

$$k(t) = \varphi_{t+1} - \varphi_{t-1} \dots \quad (2.17)$$

where the sequence $\dots, \varphi_{t-1}, \varphi_t, \varphi_{t+1}, \varphi_{t+2}, \dots$ represents the gradient direction of a sequence of pixels defining a curve segment.

Also, a curvature can be measured by changes in intensity based on the measure of angular changes in the discrete images. This can be represented as $\varphi'(x,y)$, i.e. the curvature is measured as the derivative of the function $\varphi(x,y)$ along a particular direction in the image P by measuring the average change in image intensity when a window is shifted in several directions. This idea is applied in Moravec's corner detection algorithm

[70] as:

$$E(u,v) = \sum_{x,y} w(x,y) [I(x+u, y+v) - I(x, y)]^2 \quad \dots \quad (2.18)$$

In this algorithm, four shifts: $(u,v) = (1,0), (1,1), (0,1), (-1, 1)$ are performed to look for local maxima in $\min\{E\}$. Moravec corner detector suffers from the problems of noisy response due to a binary window function, set of shifts limited at every 45 degrees, and the only minimum of E is taken into account.

Harris and Stephen [44] worked to address these problems as follows: they suggested to solve the noisy response due to a binary function by using a Gaussian function:

$$w(x,y) = \exp\left(-\frac{(x^2+y^2)}{2\sigma^2}\right) \quad \dots \quad (2.19)$$

and instead of shifts at 45 degrees they considered all small shifts by Taylor's expansion as:

$$E(u,v) = Au^2 + 2Cuv + Bv^2 \quad \dots \quad (2.20)$$

where,

$$A = \sum_{x,y} w(x,y) I_x^2(x,y) \quad \dots \quad (2.21)$$

$$B = \sum_{x,y} w(x,y) I_y^2(x,y) \quad \dots \quad (2.22)$$

$$C = \sum_{x,y} w(x,y) I_x(x,y) I_y(x,y) \quad \dots \quad (2.23)$$

While the problem of minimum of E is taken into account was solved by suggestion a new corner measurement depending on eigenvalues in all directions.

2.5.1.4 Describing Image Motion:

To consider the low-level features in case of motion, more than one image; at least two have to be obtained at different times. The features of the moved image can be detected in the simplest way by image differencing through the tracking of every pixel intensity values. The no motion pixel will show a zero difference value, while the moved

pixel will return the intensity difference values of two different locations or states.

Optical flow; which is the pattern of apparent motion of objects, surfaces, and edges caused by the relative motion between an observer and the scene, is used for the detect and percept motion. The optical flow approaches try to estimate the motion between two images or two image frames taken at times t and $t + \Delta t$ at every voxel position. Optical flow principle is used by roboticists, object segmentation, luminance, motion compensated encoding, stereo disparity measurement, and many other fields.

2.5.2 Feature Extraction Standard Techniques

The completed feature extraction requires a successful build of detectors and descriptors. A wide range of feature detectors and descriptors has been presented in the literature. A widely known and used detector was the Harris detector [44], which is based on the eigenvalues of the second-moment matrix, But Harris corners are not scale invariant.

Lowe [8] proposed the well-known technique SIFT for extracting invariant features from images which imply both detector and descriptor. The detection of keypoints depends on a cascade filter approach while the descriptor based on the orientation of keypoints through histograms of gradient with 128-D feature vector. SIFT is invariant to scale, rotation, illumination and viewpoint changes. The scale invariant realized through the application of scale space principle by detecting stable keypoint locations using scale-space extrema in the difference-of-Gaussian (DoG) function convolved with image $I(x, y)$, which can be computed from the difference of two nearby scales separated by a constant multiplicative factor k :

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \quad \dots \quad (2.24)$$

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

and $I(x,y)$ is the input image

as shown in figure 2.5 below:

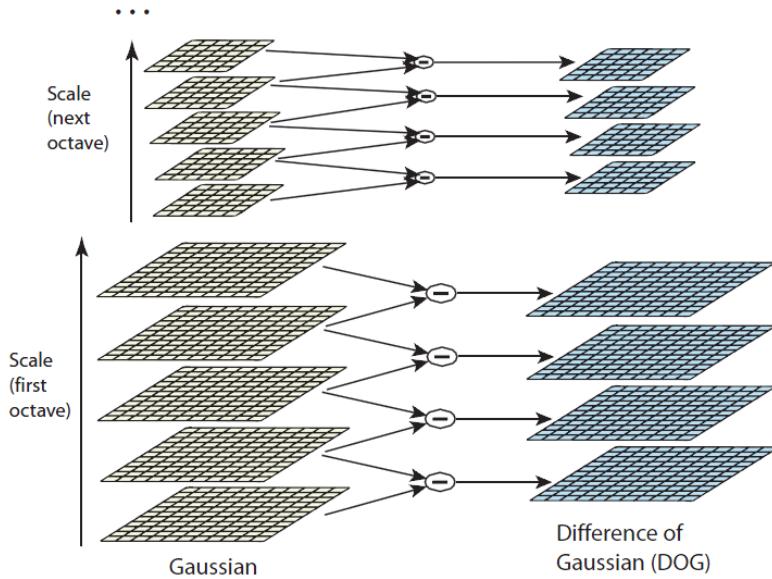


Figure 2.5: For each octave of scale space, the initial image is repeatedly convolved with Gaussians to produce the set of scale space images shown on the left. Adjacent Gaussian images are subtracted to produce the difference-of-Gaussian images on the right. After each octave, the Gaussian image is down-sampled by a factor of 2, and the process repeated. [8]

The descriptor is computed by assigning a consistent orientation to each keypoint based on image properties. The Gaussian smoothed image L with the closest scale is used in the computation of gradient magnitude, $m(x,y)$ and orientation, $\theta(x,y)$ are computed using pixel difference:

$$m(x,y) = \sqrt{(L(x+1,y) - L(x-1,y))^2 + (L(x,y+1) - L(x,y-1))^2} \dots (2.25)$$

$$\theta(x,y) = \tan^{-1} ((L(x,y+1) - L(x,y-1)) / (L(x+1,y) - L(x-1,y))) \dots \dots \dots (2.26)$$

The orientation histogram is formed from the gradient orientations of sample points within a region around the keypoint [8].

Mikolajczyk and Shmid [77] presented in their paper a comparison of several local descriptors including steerable filters [78], differential invariants [79], moment invariants [80], cross-correlation, GLOH [77] and SIFT [8]. They showed that the ranking of the accuracy for different algorithms was insensitive to the method used to find keypoints but depend on the representation to model these keypoints. So, Ke and Sukthankar [45] introduced a local image descriptor by applying Principal Components Analysis (PCA) to the normalized gradient patch to get more distinctive, more robust to image deformation, and more compact than the standard SIFT representation.

Bay et al [9] proposed the scale- and rotation –invariant interest point (SURF) detector and descriptor. The detector is based on the Hessian matrix and relies on the integral images to reduce the computation time. The integral images allow the fast computation of the box type convolution filters, the entry of the integral image $I\Sigma(x)$ at a location $a=(x, y)^T$ represents the sum of all pixels of the input image I within a rectangular region formed by the origin and a :

$$I\Sigma(x) = \sum_{i=0}^{i \leq x} \sum_{j=0}^{j \leq y} I(i, j) \quad \dots \quad (2.27)$$

While the Hessian matrix based interest points, for a point $a=(x, y)$ in an image I , the Hessian matrix $H(a, \sigma)$ in a at scale σ is defined as follows:

$$H(a, \sigma) = \begin{bmatrix} L_{xx}(x, \sigma) & L_{xy}(x, \sigma) \\ L_{yx}(x, \sigma) & L_{yy}(x, \sigma) \end{bmatrix} \quad \dots \quad (2.28)$$

Where $L_{xx}(x, \sigma)$ is the convolution of the Gaussian second order derivative $\frac{\partial^2}{\partial x^2} g(\sigma)$ with the image I in point a .

The descriptor describes a distribution of Haar-wavelet responses within the interesting point neighborhood to introduce the descriptor of 64-D length. The first step to constructing the descriptor consists of constructing a square region centered on the interest point and oriented along the orientation. The region then split up regularly into smaller 4x4 square sub-regions. For each sub-region, the Haar wavelet responses computed at 5x5 regularly spaced sample points. The Haar response in horizontal direction called d_x and the Haar wavelet response in vertical direction is called d_y :

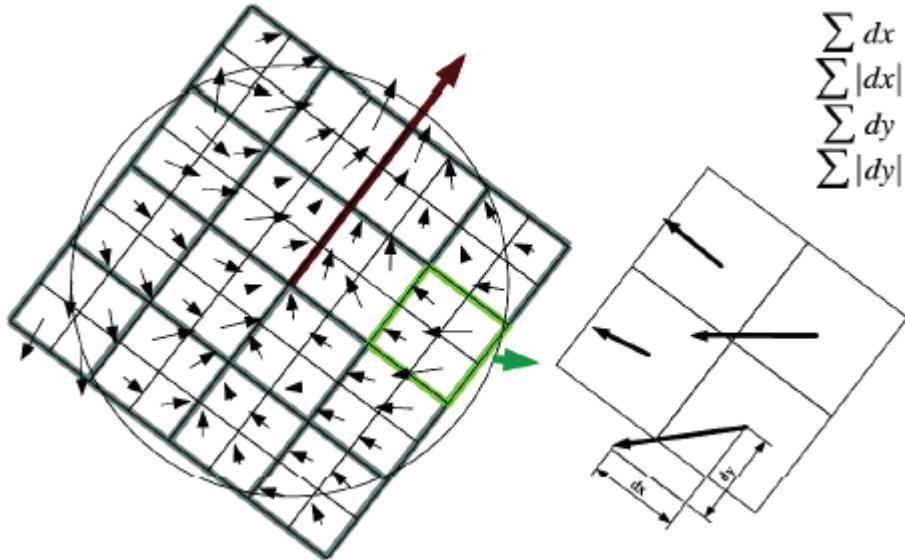


Figure 2.6: To build the descriptor, an oriented quadratic grid with 4x4 square sub-regions is laid over the interest point (left). For each square, the wavelet responses are computed. The 2x2 sub-divisions of each square correspond to the actual fields of the descriptor. These are the sums dx , $|dx|$, dy , and $|dy|$, computed relatively to the orientation of the grid (right). [9]

The wavelet responses d_x and d_y are summed over each sub-region and form a first set of entries in the feature vector as shown in figure 2.6. So, each sub-region has a four-dimensional descriptor vector \mathbf{v} for the interested structure:

$$\mathbf{v} = (\sum dx, \sum |dx|, \sum dy, \sum |dy|) \quad \dots \quad (2.29)$$

All of the aforementioned descriptors are vector- (or –word) based descriptors.

Calonder et al [46] proposed to use binary strings as an efficient feature point descriptor and presented a BRIEF binary descriptor of 256 bits length. The BRIEF suffers of high sensitivity to image rotation and scale changes. This approach classifies the image patches on the basis of relatively small number of pairwise intensity comparisons to create a bit vector out of test responses. For test τ on patch \mathbf{p} of size $S \times S$ can be computed as:

$$\tau(p; x, y) := \begin{cases} 1 & \text{if } p(x) < p(y) \\ 0 & \text{otherwise} \end{cases} \quad \dots \quad (2.30)$$

Where $p(x)$ is the pixel density.

Rosten et al [34] proposed a feature detector (FAST) which is based on machine learning and depending on a segment test criterion. This is achieved by considering a circle of 16 pixels around the corner candidate p which classified as a corner if there exists a set of n continuous pixels in the circle which are brighter than the intensity of the candidate pixel I_p plus the threshold t or all darker than I_p-t (Fig. 2.7):

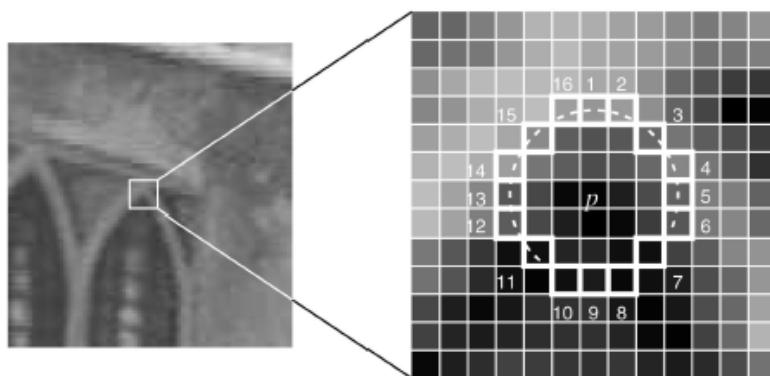


Figure 2.7: Twelve-point segment test corner detection in an image patch. The highlighted squares are the pixels used in the corner detection. The pixel at p is the center of a candidate corner. The arc is indicated by the dashed line passing through 12 contiguous pixels which are brighter than p by more than the threshold [34]

This detector is improved by using machine learning to address the problem of the ordering and distribution of corner appearance. Rublee et al [47] based on FAST keypoint detector and BRIEF descriptor to propose a very fast binary descriptor called ORB (Oriented FAST and Rotated BRIEF) which is rotation invariant and resistant to noise. The ORB detector which is called oFAST as FAST keypoint Orientation is based on the FAST detector; which takes one parameter, the intensity threshold between the center pixel and those in a circular ring about the center. oFAST uses a simple measure of corner orientation, the intensity centroid [47] which assumes that a corner's intensity is offset from its center, and this vector used to refer an orientation. The moments of the patch are defined as [81]:

$$m_{pq} = \sum_{x,y} x^p y^q I(x,y) \quad \dots \quad (2.31)$$

with these moments, the centroid is found as [84] :

$$C = \left(\frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}} \right) \quad \dots \quad (2.32)$$

So, the vector from the corner's center, O , to the centroid can be constructed. The orientation of the patch is:

$$\theta = \text{atan}(m_{01}, m_{10}) \quad \dots \quad (2.33)$$

To improve the rotation invariance of the measure, the moments are computed with x and y remaining within a circular region of radius τ .

The ORB descriptor; Rotation-Aware BRIEF (rBRIEF) is an improved version of the BRIEF descriptor which is a bit string description from a set of binary intensity tests. The first step to improve this descriptor is to steer BRIEF according to the orientation of keypoints. For any feature set of n binary test at location (x_i, y_i) , define the $2 \times n$ matrix:

$$\mathbf{S} = \begin{pmatrix} x_1, \dots, x_n \\ y_1, \dots, y_n \end{pmatrix} \quad \dots \quad (2.34)$$

Using the patch orientation Θ and the corresponding rotation matrix \mathbf{R}_Θ , the “steered” version \mathbf{S}_Θ of \mathbf{S} constructed as:

$$\mathbf{S}_\Theta = \mathbf{R}_\Theta \mathbf{S} \quad \dots \quad (2.35)$$

The steered BRIEF suffered from the loss of variance and the correlation among binary test. To recover from these two deficiency signs, Rublee et al [47] developed a learning method for choosing a good subset of binary tests a method like PCA to reduce the dimensionality, and eventually to get the ORB descriptor; r BRIEF.

Alahi et al [41] proposed a keypoint descriptor Fast Retina Keypoint (FREAK) inspired by the retina of the human visual system by computing a cascade of binary strings through the comparing of image intensities over a retina sampling pattern. This descriptor is dedicated towards the embedded applications. FREAK is a binary descriptor based on coarse-to-fine technique in which a descriptor F is a binary string formed by a sequence of one-bit Difference of Gaussians (DoG):

$$F = \sum_{0 \leq a < N} 2^a T(P_a) \quad \dots \quad (2.36)$$

where P_a is a pair of receptive fields, N is the desired size of the descriptor.

BRISK [82] is also a binary-based method for keypoint detection, description, and matching. Tola et al [83] presented DAISY descriptor which applied to wide-baseline stereo and inspired from earlier ones like SIFT and GLOH, while Fan et al [84] proposed two descriptors MROGH which is based on gradients and MRRID based on intensities, both are pool local features based on their intensity orders in multiple support regions and rotation invariant without relying on a reference orientation.

The following table summarizes the different feature detectors and descriptors:

Table 2.1: Comparison of different techniques of Feature Detectors and Descriptors

Technique	General Features	Invariance Features	Descriptor Length	Descriptor Type
SIFT (1999)	<ul style="list-style-type: none"> Describe orientation of keypoint through histograms of gradient 	<ul style="list-style-type: none"> ▪ Scale Invariance ▪ Rotation Invariance ▪ Illumination invariance ▪ Viewpoint invariance 	128-D (floating-point words)	Vector-based
PCA-SIFT (2004)	<ul style="list-style-type: none"> Apply Principal Component Analysis (PCA) technique for dimensionality reduction Applied on INRIA Graffiti dataset covered walls taken from different viewpoints 	<ul style="list-style-type: none"> ▪ Scale Invariance ▪ Rotation Invariance ▪ Illumination invariance ▪ Viewpoint invariance 	36-D	Vector-based
GLOH (2005)	<ul style="list-style-type: none"> SIFT-like descriptor that considers more spatial regions for the histograms The higher dimensionality of the descriptor is reduced to 64 through PCA 	<ul style="list-style-type: none"> ▪ Scale Invariance ▪ Rotation Invariance ▪ Illumination invariance ▪ Viewpoint invariance 	64-D	Vector-based
SURF (2006)	<ul style="list-style-type: none"> Describe orientation of keypoint through approximation by block patterns 	<ul style="list-style-type: none"> ▪ Image rotation ▪ Scale changes ▪ Illumination change ▪ Small change in Viewpoint 	64-D	Vector-based
CS-LBP (2006)	<ul style="list-style-type: none"> Combine the strengths of SIFT and LBP by using CS-LBP as local feature which is new texture feature 	<ul style="list-style-type: none"> • Invariance to illumination 	72-, 128-, 144-, or 256-D	Vector-based
BRIEF (2010)	<ul style="list-style-type: none"> Descriptor is a bit description of image patch constructed from set of binary intensity tests Very sensitive to image rotation and scale changes 	<ul style="list-style-type: none"> • Illumination invariance 	256 bit binary	Binary-based
FAST (2010)	<ul style="list-style-type: none"> Finds reasonable corner keypoints Does not include orientation operator Does not produce multi-scale feature Used Oxford data set 	<ul style="list-style-type: none"> • Viewpoint invariance • Rotation invariant • Intensity inversion of the image 	Only Detector	Binary-based
ORB (2011)	<ul style="list-style-type: none"> Measure corner orientation by intensity centroid Performs very well with scale, viewpoint, and lighting changes 	<ul style="list-style-type: none"> • Rotation invariance • Scale invariance • Viewpoint invariance • Lighting changes 	32-Bytes	Binary-based
FREAK (2012)	<ul style="list-style-type: none"> Retina-inspired keypoint descriptor Descriptor is a binary 	<ul style="list-style-type: none"> • Scale invariant • Rotation invariant 	16-bytes	Binary-based

	string formed by a sequence of one-bit of DoGs			
BRISK (2010)	<ul style="list-style-type: none"> • Scale-space FAST-based detector in combination with the assembly of a bit-string descriptor from intensity comparisons • Rotation as well as scale invariant to a significant extent, 	<ul style="list-style-type: none"> • Scale invariance • Rotation invariance 	64-bytes	Binary-based
MROGH (2012)	<ul style="list-style-type: none"> • Interest region description • Pools local features based on their intensity orders in multiple support regions • Based on gradients • Overcomes estimating a reference orientation, the major error source for most of SIFT, SURF, and DAISY 	<ul style="list-style-type: none"> • Rotation invariant without estimating a reference orientation 	192-D	Vector-based
MRRID (2012)	<ul style="list-style-type: none"> • Interest region description • Pools local features based on their intensity orders in multiple support regions • Based on intensities • Overcomes estimating a reference orientation, the major error source for most of SIFT, SURF, and DAISY 	<ul style="list-style-type: none"> • Rotation invariant without estimating a reference orientation 	256-D	Vector-based

2.5.3 High-Level Feature Extraction

High-level feature matching based on finding shapes and objects implies knowledge of a mathematical model or template of a target shape. This technique is a model-based in which the shape is extracted by searching for the best correlation between a known model and the pixels of the image. Hough transform represents an efficient implementation of computing the correlation between the template and the image by extracting simple shapes such as lines, circles, and ellipses.

Generally, Hough transform and related techniques relied on model-based idea are perfectly working on the fixed shape in that it is flexible only in terms of the parameters which characterize the shape or the parameters that define template's appearance [43].

However, in the cases that the exact shape is unknown or it might be that the degree of non-uniformity of the shape is impossible to be expected and hence to be parameterized, such model-based techniques fail to extract features. This leads to search for techniques that can evolve to target solution or adapt their result to the data [43]. This represents the milestone of deformable shape analysis. There are many techniques used in the manipulation of flexible shapes in images; most applicable techniques will be discussed in the following subsections.

2.5.3.1 Deformable Templates

The deformable templates technique is based on the analysis of a flexible image by modeling the image as composed of many basic shapes. For example, to find facial features, the eye is considered to be comprised of an iris which sits within the sclera which can be modeled as a combination of a circle that lies within a parabola [71]. This combination implies analysis complexities of the number of parameters that precisely describe each element of the component to be analyzed and the relationship between these parameters.

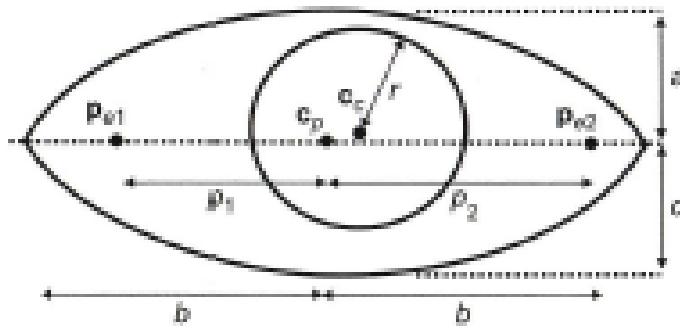


Figure 2.8: Eye template [43]

In the eye template Fig 2.8, there are eleven parameters eight for shapes and three for weighting coefficients. These parameters have to be combined with many possible values for each parameter (100 possible value for each parameter) to reach the best prediction for the eye shape. This is a very costly solution to be realized. Other alternative techniques have to be applied for practical solutions. Optimization techniques like genetic algorithms have been applied successfully by Benn et. al. [72].

2.5.3.2 Parts-Based Shape Analysis

The parts-based object analysis relied on representing objects as a collection of parts arranged in a deformable structure. These objects can be modeled as a network of masses which are connected by springs.

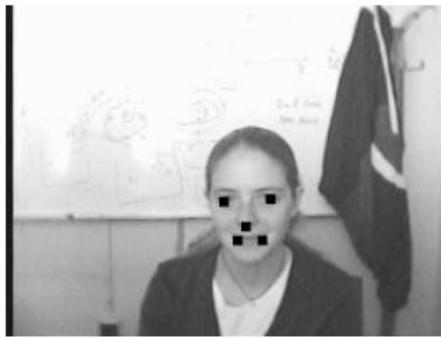


Figure 2.9: Finding face features by parts-based analysis [73]

In the face analysis of Fig. 2.9, two upper masses may represent the eyes, the two lower masses represent the mouth, and the middle one represents the nose. The springs constrain the relationships between the masses, like the mouth should be under the nose and hence, the nose under the eyes and between them [43].

2.5.3.3 Active Contours/Snakes

Active contour is a technique based on using a set of points to enclose a target object to extract required features. An initial contour is figured around the target feature and

then evolved through a repetitive process to get an optimized feature. These contours are expressed as energy minimization process, and the target feature is the minimum of best formulated energy function [43]. This function includes properties of controlling the way the contour can stretch and curve the target object.

Snakes represent a tradeoff between the contour self-properties; like the ability to bend and stretch, and image properties like the edge length. Hence, the snake contour energy function E_{snake} is comprised of contour's internal energy E_{int} , its constraint energy E_{con} , and the image energy E_{image} :

$$E_{snake} = \int_{s=0}^1 E_{int}(v(s)) + E_{image}(v(s)) + E_{con}(v(s)) ds \dots\dots\dots (2.37)$$

The internal energy E_{int} controls the natural behavior of the snake by arranging the snake points; the constraint energy E_{con} provides the higher level information controlling the contour evolution, and the image energy E_{image} pulls the snake to chosen features [43].

2.5.3.4 Shape Skeletonization

The shape can be described by its skeleton by specifying a central axis to a shape. This axis is equidistant from the borders of a shape and described by a distance transform (distance map or distance field). Distance transform will contain the essence of the shape and shows the distance from each point in an image shape to its central axis [43].

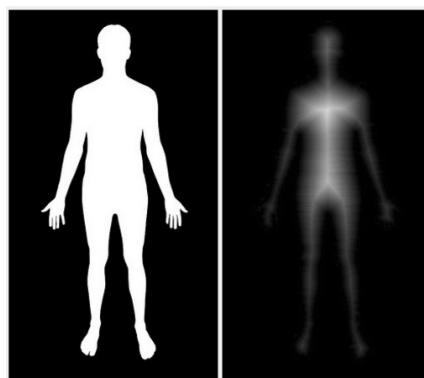


Figure 2.10: Applying distant transform

The shape skeleton of and image by the application of the distant transform is shown in Fig. 2.10.

2.6 Optimization Process in Image Registration

To register a reference image \mathbf{I} with a sensed image \mathbf{S} , a spatial transformation matrix $\mathbf{W}(\mathbf{p})$; where $\mathbf{p}^T = (p_1, p_2, \dots, p_n)$ is a set of parameters, has to be found to perform the matching of two images. There are different image registration algorithms aim to reduce the difference between the two images. The most widely used algorithm is Lucas-Kanade [48] which iteratively generates the transformation matrix $\mathbf{W}(\mathbf{p}_k)$ to fulfill this goal. Generating a set of a warp parameters \mathbf{p}_k at iteration k -th requires the gradient of the cost function \mathbf{g}_k and an appropriate descent parameter a_k [85],

$$\mathbf{p}_{k+1} = \mathbf{p}_k + a_k \mathbf{g}_k \quad \dots \quad (2.38)$$

As the image registration process aims to match $I(\mathbf{W}(\mathbf{p}))$ to S , this can be described as the minimization of:

$$\frac{1}{2} \sum_x [I(\mathbf{W}(\mathbf{p})) - \mathbf{S}]^2 \quad \dots \quad (2.39)$$

and for iteratively solving the problem for the increment $\Delta\mathbf{p} = a_k \mathbf{g}_k$, the cost factor F can be expressed as:

$$F = \frac{1}{2} \sum_x [I(\mathbf{W}(\mathbf{p} + \Delta\mathbf{p})) - \mathbf{S}]^2 \quad \dots \quad (2.40)$$

This minimization process will continue until F or \mathbf{p} converge.

Differentiating F with respect to \mathbf{p} will produce the gradient \mathbf{g}

$$\mathbf{g} = \sum_x \left[\nabla I \frac{\partial \mathbf{W}}{\partial \mathbf{p}} \right]^T [I(\mathbf{W}(\mathbf{p})) - \mathbf{T}] \quad \dots \quad (2.41)$$

where $\nabla I = (\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y})$ is the gradient of deformed image I evaluated with reference image T at $\mathbf{W}(\mathbf{p})$ and $\frac{\partial \mathbf{W}}{\partial \mathbf{p}}$ is the Jacobian of the transformation matrix.

Optimization process aims to iteratively minimize the cost function and update parameters. There are many popular optimization methods that vary in the number of parameters and required computation complexity to minimize the cost function acceptable in the application in hand. Generally, the optimization process can be summarized in Algorithm 2.1 below [85]:

Algorithm 2.1:

The optimization procedure can be outline as follows:

$$\hat{p} = \min_p \left\{ F = \frac{1}{2} \sum_x [I(W(p)) - T]^2 \right\}$$

Precompute : (1) $\frac{\partial W}{\partial p}$

Iterate : (2) Evaluate $W(p)$

(3) Warp I by $W(p)$ to obtain $I = I(W(p))$

(4) Evaluate ∇I

(5) Evaluate $\nabla I \frac{\partial W}{\partial p}$

(6) Compute the gradient $g = \sum_x \left[\nabla I \frac{\partial W}{\partial p} \right]^T [I(W(p)) - T]$

(7) Compute the Hessian H by an Optimization class *

(8) Find step-size τ

(9) Compute Δp

(10) Update $p_{k+1} = p_k - \tau \Delta p$

Until : i. $\|\Delta p\| \leq \epsilon_p$

ii. $\|\Delta F\| \leq \epsilon_F$

iii. $\|g\| \leq \epsilon_g$

iv. Exceeding maximum of iterations

* for NCG method: (7) is not used

** for LM method: use Trust Region instead of (8)

Step (1)-(6) and (10) are essential for every method, step (8) is negligible for good approximation. Therefore the comparable complexity now relies on step (7) and (9).

2.6.1 Gauss-Newton (GN) Optimization Algorithm

Gauss-Newton algorithm classified as a form of nonlinear least-square problem [86].

The objective function expressed as least-square problem:

$$F = \frac{1}{2} \sum_x f^2 \quad \dots \quad (2.42)$$

where $f_x = I_x(W(p)) - T_x$. Gauss-Newton method approximates the Hessian $\nabla^2 F$ to:

$$H \approx \sum_x \left[\frac{\partial f}{\partial p} \right]^T \left[\frac{\partial f}{\partial p} \right] \quad \dots \quad (2.43)$$

$$H = \sum_x \left[\nabla I \frac{\partial W}{\partial p} \right]^T \left[\nabla I \frac{\partial W}{\partial p} \right] \quad \dots \quad (2.44)$$

Gauss-Newton convergence to a local minimum can be rapid if the approximated Hessian term approaches the full Hessian value. The evaluation cost is $O(N^2 n^2)$ and the comparable cost is $O(N^2 n^2 + n^3)$ [85].

2.6.2 Levenberg-Marquardt (LM) Optimization Algorithm

Levenberg-Marquardt algorithm overcomes the weakness of Gauss-Newton of costly rounding through the error in approximating the Hessian by deciding whether the error in approximating Hessian gets better or worst after each iteration [85]. Levenberg-Marquardt form for registration problem is:

$$H \approx \sum_x \left[\nabla I \frac{\partial W}{\partial p} \right]^T \left[\nabla I \frac{\partial W}{\partial p} \right] + \delta \sum_x \begin{pmatrix} \left(\nabla I \frac{\partial W}{\partial p_1} \right)^2 & 0 & \dots & 0 \\ 0 & \left(\nabla I \frac{\partial W}{\partial p_2} \right)^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \left(\nabla I \frac{\partial W}{\partial p_n} \right)^2 \end{pmatrix} \dots \quad (2.45)$$

The convergence rate of Levenberg-Marquardt method can be estimated in same way as GN method and the comparable cost is same of GN as well.

2.6.3 Quasi-Newton (QN) Optimization Algorithm

Quasi-Newton method require the first-order derivative of the objective function at each iteration. The evaluation does not require the Jacobian of the warp matrix and the image gradient. Quasi-Newton method can be convergence in super linear way as [85]:

$$\lim_{k \rightarrow \infty} \frac{\|p_{k+1} - \hat{p}\|}{\|p_k - \hat{p}\|} \rightarrow 0 \quad \dots \quad (2.46)$$

The computational cost is $O(n^3)$ and the comparable complexity is same as computational cost.

2.6.4 Nonlinear Conjugate Gradient (NCG) Optimization Algorithm

In Nonlinear Conjugate Gradient method, the step Δp is defined as a linear combination of the gradient g_k and the previous search direction:

$$\Delta p = d_k = g_k - \beta_k d_{k-1} \quad \dots \quad (2.47)$$

The selection of β_k highly affect the convergence property of the method. The comparable cost of NCG is $O(n^2)$.

2.6.5 Markov Random Fields (MRF)

Markov random field is a set of random variables obey the Markov property characteristics and described by an undirected graph. MRF network is undirected and may be cyclic and represents variable dependencies within the network. Markov random field is used to model various low- to mid-level tasks in image processing and computer vision. These tasks are normally characterized as energy minimization problems on a

rectangular grid of pixels, an image as an example. The energy composed of a data term and a smoothness term:

$$E(u) = E_{data}(u) + E_{smoothness}(u) \quad \dots \dots \dots \quad (2.48)$$

The data term $E_{data}(u)$ describes the optimization model u to be logically related to the measurements while the smoothness energy $E_{smoothness}(u)$ is derived from our prior knowledge of persuasive results. MRF may be successfully applied in many fields of computer vision and image processing such as stereo disparity, segmentation, surface reconstruction, and denoising [87]. Fig. 2.11 shows an example of MRF in which Each edge represents dependency. In this example: A depends on B and D. B depends on A and D. D depends on A, B, and E. E depends on D and C. C depends on E.

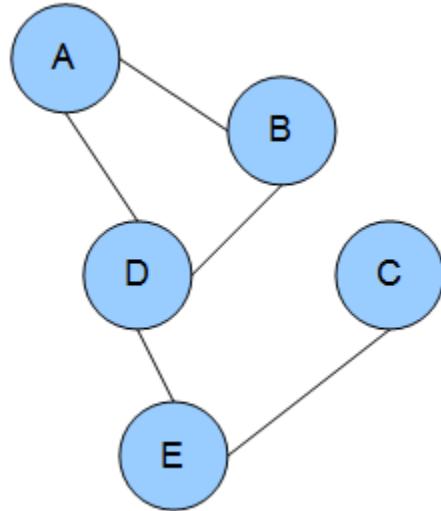


Figure 2.11: MRF network

A Markov random field is a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where

- $\mathcal{V} = \{1, 2, \dots, N\}$ is the set of nodes, each of which is associated with a random variable (RV), u_j , for $j = 1\dots N$.

- The neighbourhood of node i , denoted \mathcal{N}_i , is the set of nodes to which i is adjacent; i.e., $j \in \mathcal{N}_i$ if and only if $(i, j) \in \mathcal{E}$
- The Markov random field satisfies:

$$p(u_i | \{u_j\}_{j \in \mathcal{V} \setminus i}) = p(u_i | \{u_j\}_{j \in \mathcal{N}_i}) \quad \dots \quad (2.49)$$

The distribution over MRF, i.e. over RVs $u = (u_1, \dots, u_n)$ that satisfies 2.49 above can be expressed as the product of positive potential functions defined on maximal subgraphs of \mathcal{G} that are fully connected, this is known as cliques of \mathcal{G} . These distributions are sometimes expressed in terms of energy function E and clique potentials Ψ_c [87]:

$$p(u) = \frac{1}{Z} \exp(-E(u, \theta)), \text{ where } E(u, \theta) = \sum_{c \in \mathcal{C}} \Psi_c(\bar{u}_c, \theta_c) \quad \dots \quad (2.50)$$

Where \mathcal{C} is the set of maximal cliques of the graph and Z is the partition function that ensures the distribution sum to 1[87].

2.6.6 Belief Propagation

Belief propagation, also known as sum-product message passing, is a message passing algorithm for performing inference on graphical models, such as Bayesian networks and Markov random fields. It calculates the marginal distribution for each unobserved node, conditional on any observed nodes [88].

- A Bayesian network model or probabilistic directed acyclic graphical model is a probabilistic graphical model (a type of statistical model) that represents a set of random variables and their conditional dependencies via a directed acyclic graph (DAG). For example, a Bayesian network could represent the probabilistic relationships between diseases and symptoms. Given symptoms, the network can

be used to compute the probabilities of the presence of various diseases. An example of Bayesian network is shown in Fig. 2.12.

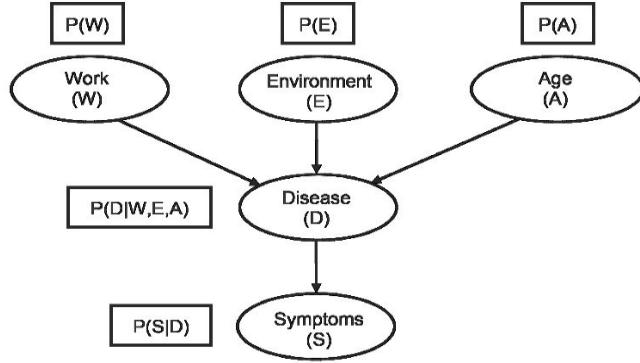


Figure 2.12: Bayesian network

- Factor graph: A factor graph is a type of probabilistic graphical model. A factor graph has two types of nodes: Variables, which can be either evidence variables when their value is known, or query variables when their value should be predicted. Factors, which define the relationships between variables in the graph. Fig. 2.13 shows an example of the factor graph.

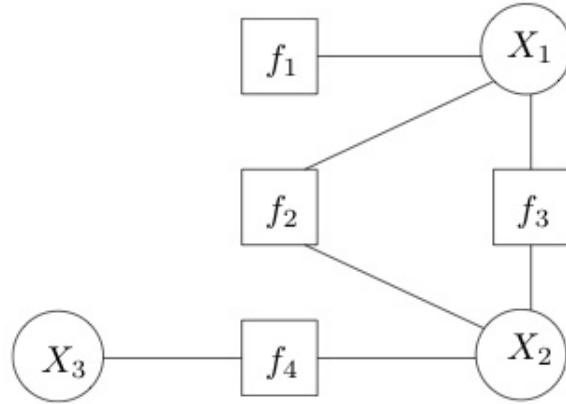


Figure 2.13: An example of factor graph

In belief propagation (BP), the variables $m_{ij}(x_j)$ are considered to be as a “message” from a hidden node i to the hidden node j in Fig. 2.14. In BP algorithm, the belief at a node i is

proportional to the product of the local evidence at that node ($\phi_i(x_i)$), the message passed to node I [74]:

where k is a normalization constant, and $N(i)$ denotes the nodes neighboring i .

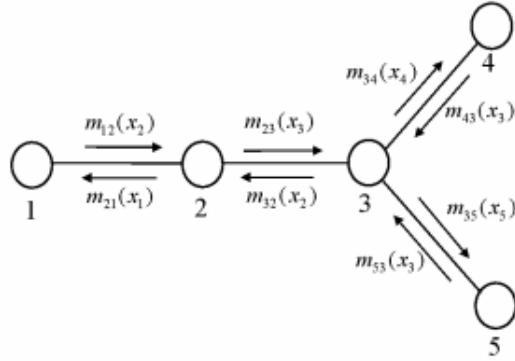


Figure 2.14: An illustration of the message passed in Belief Propagation

It is possible to apply BP to other models like Bayesian network, factor graph, or MRF network by first converting the graph into required model and then applying the required model rules.

2.7 Panorama Types

2.7.1 Spatial Panoramas

Panoramic image stitching based on image mosaicing has got an extensive interest with several commercial software packages produced. The process based on the basic geometry of the problem by estimating a 3×3 camera matrix or homography for each image. This estimation process requires an initial user input to approximately align the images or a fixed image overlaying. This yields a mosaicing panoramic view of spatial stitched images for certain location or scene in a specific time [6].

Panorama stitching system usually based on two main processes, image plane projection and image stitching. When two adjacent images are captured in different planes, it is difficult to get a correlated part from both adjacent images to stitch them together unless one of the images has to be projected from its image plane to the image plane of the adjacent image [7].

2.7.1.1 2-D Spatial Panoramas

An essential step of establishing the mathematical relationships that map pixel coordinates from one image to another is required. A variety of such parametric transformation models can be derived, from a simple 2D transforms to planar perspective models [23]. The mathematical model is required to establish image stitching completely depends on a specific image case. For example, to build a 2D panorama with a simplest transformations occur, which is a translation, as illustrated in Fig 2.15, the 2D translation model can be written as:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad \dots \dots \dots \quad (2.52)$$

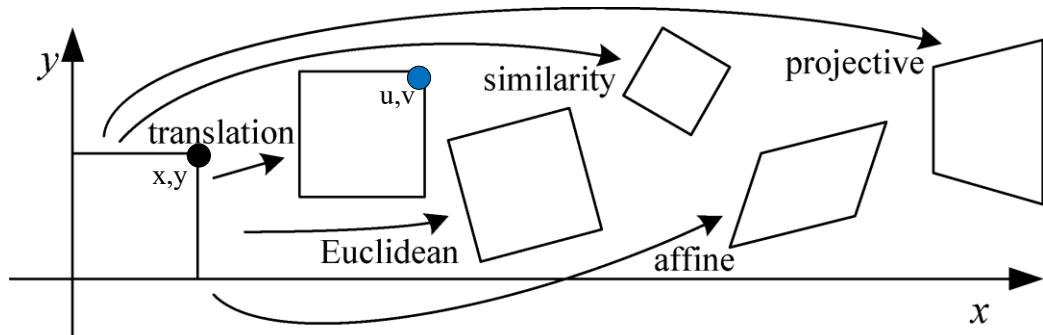


Figure 2.15: Basic set of 2D planar transformation [122]

The pixel x, y or x of an image is translated by t_x, t_y to u, v , the 2D translation can be written as $\mathbf{x}' = \mathbf{x} + \mathbf{t}$. The set of 2D transformations shown in Fig. 2.15 summarized in

Table 2.2 are potentially restricted 3×3 matrices work on 2D homogeneous coordinate vectors. Where $\mathbf{R} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$, s is arbitrary scale factor, \mathbf{A} is an arbitrary 2×3 matrix, and $\tilde{\mathbf{H}}$ is an arbitrary 3×3 matrix.

Table 2.2: Set of 2D coordinate transformations. The 2×3 matrices are extended with a third $[0^T \ 1]$ row to form full 3×3 matrix for homogeneous coordinate transformations. \mathbf{I} is (2×2) identity matrix [24].

Transformation	Matrix	Degree of Freedom	Preserves
Translation	$[\mathbf{I} \ \ \mathbf{t}]_{2 \times 3}$	2	Orientation +
Rigid(Euclidean)	$[\mathbf{R} \ \ \mathbf{t}]_{2 \times 3}$	3	Lengths +
Similarity	$[s\mathbf{R} \ \ \mathbf{t}]_{2 \times 3}$	4	Angles +
Affine	$[\mathbf{A}]_{2 \times 3}$	6	Parallelism +
Projective	$[\tilde{\mathbf{H}}]_{3 \times 3}$	8	Straight lines

The panorama creation is facilitated by stitching multiple images of overlapping views to build a continuous single view as shown in Fig. 2.16 for the Parliament Hill of Canada.



Figure 2.16: 2D Spatial panoramas for the Parliament Hill building of Canada

2.7.1.2 3-D Spatial Panoramas

3D coordinate transformations can be denoted using 4×4 transformation matrices

equivalent to translation, rigid body (Euclidean) and affine transformations. The central projection maps 3D coordinates $p = (X, Y, Z)$ to 2D coordinates $x = (x, y, l)$ at camera origin , Fig 2.17, through a pinhole onto a 2D projection plane a distance f along the z axis :

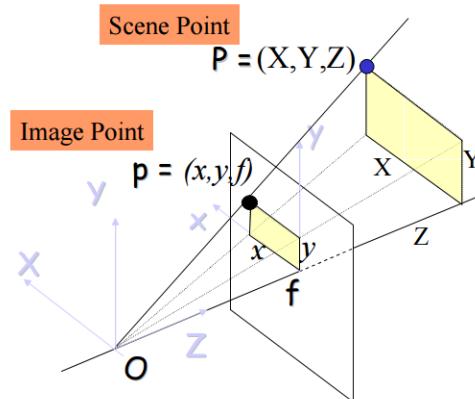


Figure 2.17: 3D projection onto 2D plane

The relationship between the focal length f and the field of view θ as shown in Fig. 2.18 is given by [24]:

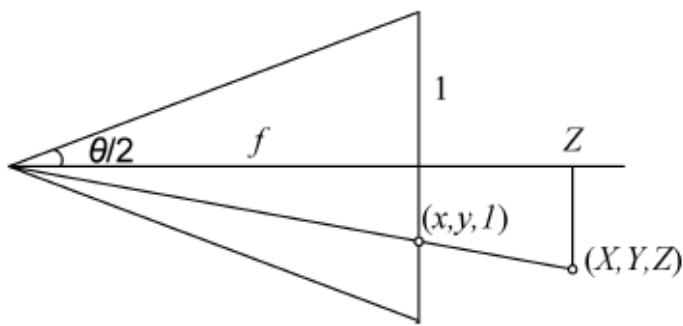


Figure 2.18: Central projection, showing the relationship between the 3D and 2D coordinates p and x , as well as the relationship between the focal length f and the field of view θ [24].

By representing the 2D point (x,y) by a 3D point (x,y,z) using the homogenous coordinates, the perspective projection can be represented as:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \dots \quad (2.55)$$

Fig. 2.19 shows a 360-degree panorama picture of the center courtyard of the Sony Center at the Potsdamer Platz in Berlin which was generated from 126 individual photos using autostitch.

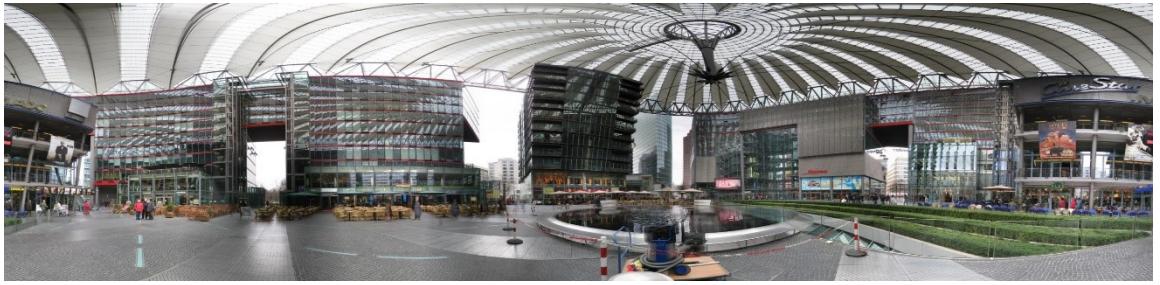


Figure 2.19: 3D panorama picture of the center courtyard of Sony Center at the Potsdamer Platz in Berlin [25]

2.7.2 Temporal Panoramas

Presenting timeline events or images in chronological order with a comprehensive coverage represents a valuable panorama type. This presentation may cover a few seconds or hundreds of years depending on the subject material. For example, presenting a 10-second time-lapse video of a flower's lifecycle may reflect days of growth, while 10 seconds of ordered images, text, and maps may highlight centuries of a landmark's history. When arranged and display properly, temporal panoramas may provide successful documentation of a place or objects history [26]. The temporal panoramas can be further classified as:

2.7.2.1 Time-lapse Videos

This is a technique based on the controlling of video or film frame rates. The low rate film is subsequently played back at the normal rate. The first use of time-lapse photography in a feature film was in Georges Méliès' motion picture Carrefour De L'Opera in 1897 [27]. This technique is used widely in many applications such as cloudscapes motions, plants growing and flower opening, evolution of project construction, and traffic [26].

2.7.2.2 Historic Perspective Panorama (HPP)

Provides a digital timeline of related events. HPP is used by dipity.com to allow the users to collect real-time contents from search services, social media, and more to integrate them in a single, easy to use interface [28]. These contents, organized by time and date, bring history to life with wonderful multimedia timelines. Fig. 2.20 shows a historical review of Steve Jobs as produced by dipity.com. It is important to notice that this timeline review consists of separated; unrelated geographically, images, text, and events arranged in chronological order [26].



Figure 2.20: Timelined multimedia components exploring the history of Steve Jobs in a historic perspective panoramic view [28].

2.7.3 Timelined Panoramas

Timelined panoramas are type of panoramas which merges both the spatial and temporal panorama concepts in a single presentation.

2.7.3.1 Rideau Timescapes

A historical media and GPS combination produce an Apple IOS application that allows the user to interact with the visual heritage of lock stations along the Rideau Canal; the oldest continuously operated the canal in North America [29]. This application was designed using a client/server standard that allows navigation through some available historical data sources and displaying different modes of timescapes data [12].

In this application, the client sends the GPS coordinates to the server as shown in Fig. 2.21 (a), and the server responds by presenting the timescapes visualization of the selected timescape pin chosen by the user as in Fig. 2.21 (b) [12]. This application was the first realization of the new timescapes concept [26].



Figure 2.21: Screenshots of a timescapes location and timeline interface [12].

Chapter 3

Image Subset Selection

A method for the selection of subsets of images, both modern and historic, out of a large set of landmark images collected from the Internet is presented in this chapter. This selection depends on the extraction of dominant features using Gabor filtering. Features are manually selected carefully from a preliminary image set and fed into a neural network as a training data. The method collects a large set of raw landmark images containing modern and historic landmark images and non-landmark images. The method then processes these images to classify them as landmark and non-landmark images.

Classification of a large set of images containing both modern and historic images allowing a selection of a subset of these images that meets certain technical criteria is based completely on the detection and extraction of image features. Feature extraction has received significant effort of researchers because of its great role in computer vision, image processing, and robotics fields [43]. Many standard feature extraction techniques have been applied to images to achieve optimal extraction performance and to maintain robustness by addressing the varying spatial resolutions, illumination, observer viewpoint, and rotation [8, 9, 46, 47]. However, matching historical images that have been digitized with those taken with modern digital cameras still proves an elusive problem [31].

In this chapter, firstly, we'll describe our data set, explain its nature, and declare the required processing steps so that both historic and modern images will be prepared for next processing steps in the timeline panorama implementation system. Then, the

theoretical basics supporting the image subset selection including both Gabor functions and neural networks will be explained. Finally, the results obtained are presented and analyzed, steps to improve this method are applied and tested.

3.1 Timeline Based Image Data Set

The method proposed in this chapter is created from a collection of photos from the Internet for eleven famous landmarks around the world. This dataset consists of historic images taken more than one century ago (and subsequently digitized) as well as modern images which have been captured by modern digital cameras. The pictures were acquired using various different technologies under very different capture environments [31]. Hundreds of thousands of images were collected from websites like Flickr and Google Images for landmarks. These images are automatically evaluated to ensure that they have a reasonable likelihood of being taken from roughly the same place, the core criterion leads to this chapter's contribution.

The primary collected images; which is a raw data huge set, consists of historic and modern landmark images and non-landmark images in folders of 50,000 to 350,000 images for each landmark collected from Flickr search and Google Image search. To set up these images for the following preparing steps, these images must be changed over into grayscale and resized to a common size.

The collected image sets are captured in different lighting conditions with considerable variations of brightness and contrast. Adaptive Histogram equalization [93] was used to address these issues in which many histograms are computed and each histogram; each histogram compares to a particular segment of the image and uses them

to redistribute the image brightness values [26]. The input values are finally normalized to the range [-1,1].

Image subset selection stage keeps many hundreds of images for each landmark in chronological order with many images belonging to the same year in the data set. The best image from each year that fit the required viewpoint was selected based on the highest image matching factor calculated in Algorithm 3.1. Samples of these images are shown in Fig. 3.1.

Eiffel Tower	Colliseum of Rome	Stone-henge	Parliament Hill of Canada	Dome Rock	Empire Building	Machu Picchu	Pyramids of Giza	Whitehouse of Washington	Taj Mahal	Hagia-Sophia

Figure 3.1: Samples of the image data sets for the 11 landmark images containing mix of historic and modern images

3.2 Gabor Functions and Filters for Feature Extraction

Consider that we have different landmark images for different landmarks. We are interested in highlighting those parts of the landmark images that are common for images of this landmark in roughly the same view. Of course, we need features that can successfully distinguish different landmark from each other.

Gabor features can do both and in this work have been used for both purposes. They remove useless and redundant information and what is left can be used for feature detection or feature recognition.

Automatic detection and extraction of features got a great attention of researchers in recent years due to the importance of this field in various different engineering, science, and arts applications such as surveillance, security, biomedical and computer vision. Research has shown that the local image features are robust against variations and distortions when a spatial-frequency analysis is used to extract such features. Gabor wavelets provide an optimized platform in both spatial and frequency domains [94].

Gabor wavelets seem to be a perfect solution to extract local features for the reasons [95]:

- Biological basis: the 2D Gabor theoretical framework developed above for understanding their receptive-field profiles are similar to properties of simple cells in the mammalian visual cortex.
- Mathematical basis: Gabor filters are successfully used for measuring local spatial frequencies.
- Empirical basis: Gabor wavelets are actively treated with the spaces of distortion tolerant feature like fingerprint recognition, handwritten numeral processing, and texture segmentation [94].

Gabor wavelets; which form the basics of Gabor functions are primarily constituted of two parts: 1) a complex sinusoidal carrier placed under a Gaussian envelope. Apart from the Gaussian envelope; and 2) sinusoidal carrier frequency and

orientation of its own that is similar to that of a human visual system known as the visual cortex.

D. Gabor presented the fundamental basics of Gabor wavelets in 1946 [94].

Based on these fundamentals, many schemes of the 2-D Gabor filters have been presented. One of the common forms is defined as [96]:

$$G(x, y) = \frac{f^2}{\pi\gamma\eta} e^{\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right)} e^{(2jfx' + \phi)} \quad \dots \quad (3.1)$$

where $x' = x \cos O + y \sin O$

$y' = -x \sin O + y \cos O$

f is the frequency of the sinusoid

O represents the orientation of the normal to the parallel stripes of a Gabor function

ϕ is the phase offset

σ is the standard deviation of the Gaussian envelope

γ is the spatial aspect ratio which specifies the ellipticity of the Gabor function

Gabor filters can be used effectively to classify the image features captured in different various conditions [97].

Figure 3.2 shows the wavelets generated with three control parameters: scale, frequency, and orientation respectively as shown above each kernel. The Gabor wavelets of Fig. 3.2 was applied on an Eiffel tower image (image shown at the top of Fig. 3.3), and the resultant Gabor family is shown in Fig. 3.3. If we look at the results of the Gabor filtering on an Eiffel tower image in the figure below, we can note that the most

distinguishable features of the tower are the vertical parts like tower bars. This is because the nature of the tower is that the vertical edge features are the dominant features of the images.

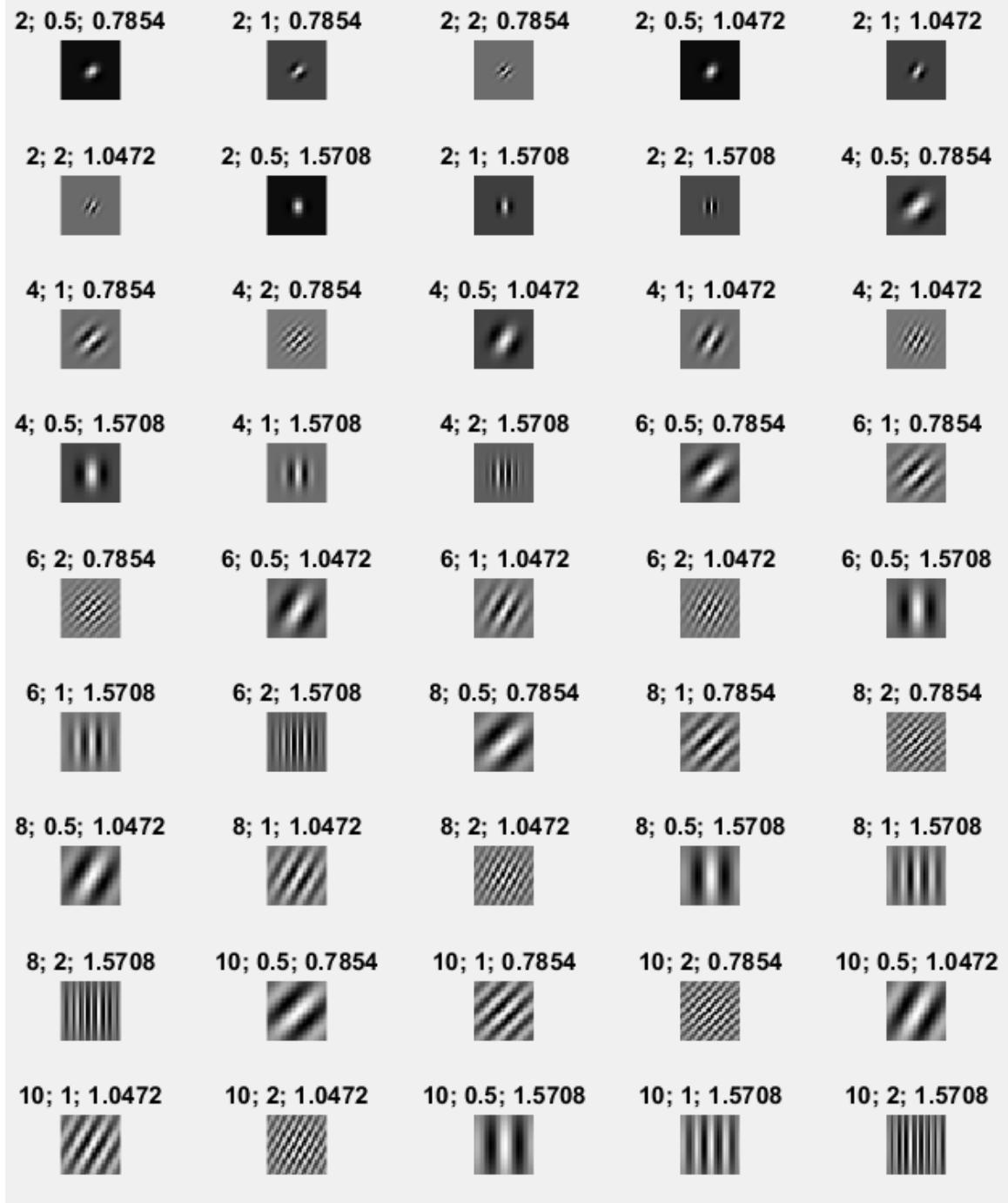


Figure 3.2: Wavelets generated with three control parameters: numbers above each wavelet arranged as scale, frequency, and orientation respectively

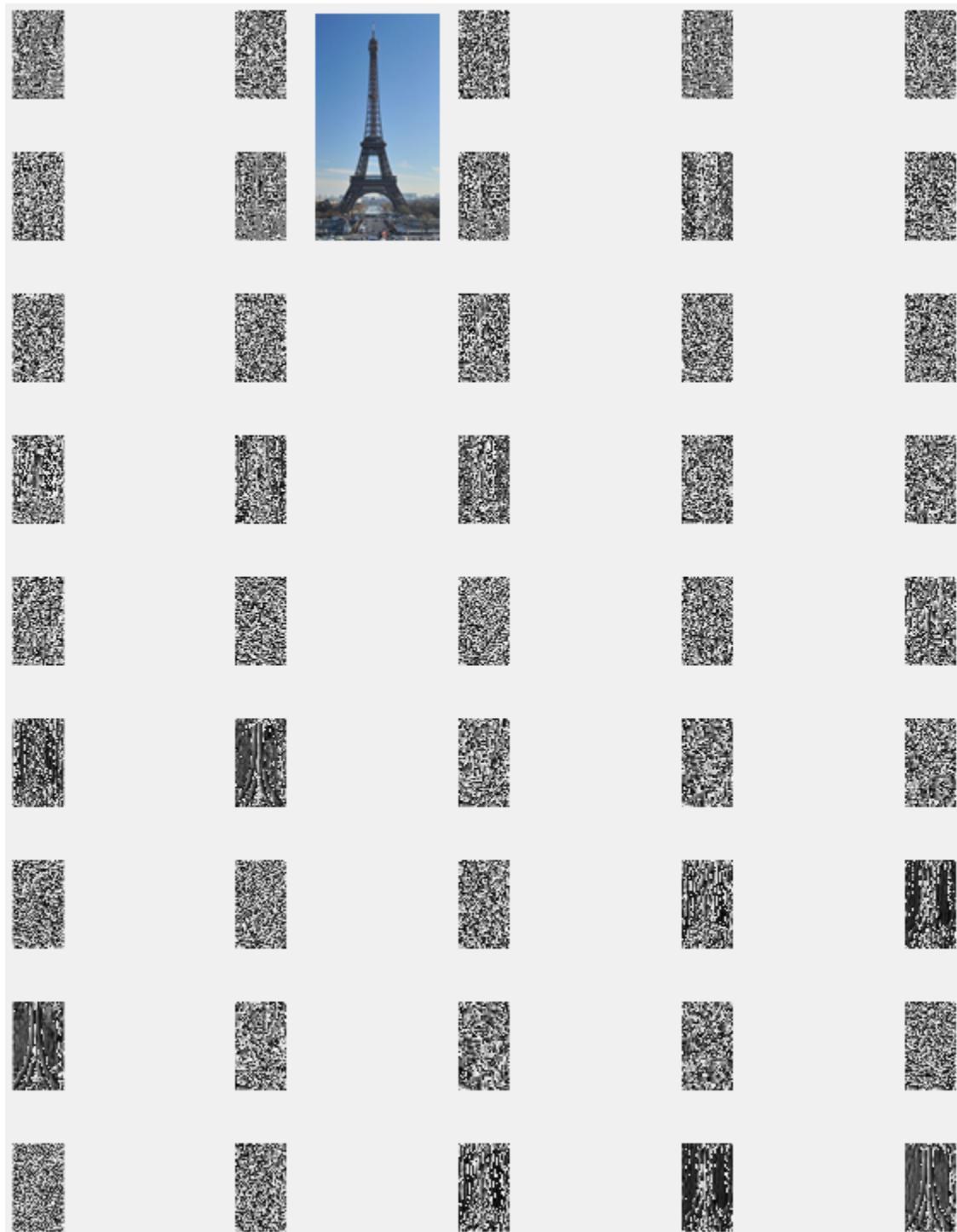


Figure 3.3: Gabor family

The Gabor filtering is performed by convolving images with Gabor kernels. After creating all required kernels, each kernel should be convolved with the window. Convolution process can be done using Fast Fourier Transform (FFT) by multiplying the frequency domain of Gabor kernels with all image components on a pixel-by-pixel basis to produce the dot product. To return back the results into the spatial domain, the inverse Fourier Transfer (IFFT) is performed. The feature vector is formulated finally from the convolved image data with Gabor kernels and then fed to be trained on the neural network.

Then, the Gabor features extracted from an image has to be converted to a vector which should be forwarded to the classifier for the task of classification. The feature extraction process is illustrated in Fig. 3.4. Generated Gabor wavelets were in five scales and eight orientations. The size of every landmark image used in our work is 256×256 pixels. Using 40 Gabor wavelets, the dimension of the feature vector is $256 \times 256 \times 40 = 2,621,440$ this is a very large vector that will complicate the computation time and occupies a large memory size. To reduce it, a technique based on the high correlated adjacent pixels of images and the fact that our main aim of feature extraction is the detection of prominent features of the image, we worked to reduce this information redundancy by downsampling the feature images by a factor related to the nature of the landmark image. For example, the Eiffel Tower images were downsampled by a factor of 8, which means that the feature vector will have the size of $2,621,440 / (8 \times 8) = 40,960$.

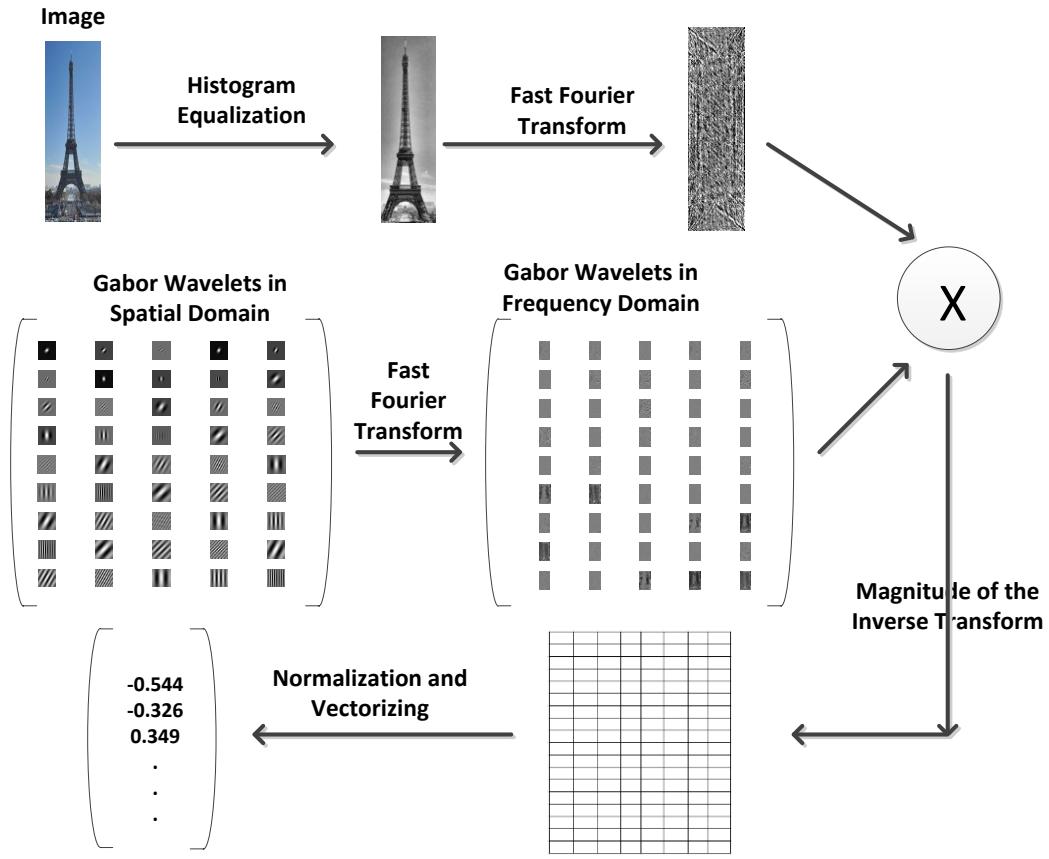


Figure 3.4: Feature extraction process

3.3 Neural Networks

The simplest element of a neural network is called a perceptron. It can be seen as a binary classifier which maps its input vector x to a single binary output value depending on the used transfer function $f(x)$:

$$f(x) = \begin{cases} 1 & \text{if } w \cdot x + b > 0 \\ 0 & \text{otherwise} \end{cases} \quad \dots \dots \dots \quad (3.2)$$

where w is a vector of weights which all its elements are real values, and b is called a bias which is a constant term that does not depend on any input value as shown in Fig. 3.5. Through a training phase, the weight w and its bias b are changing depending on the

required function. After the learning phase, the weights and bias are fixed as the training is over.

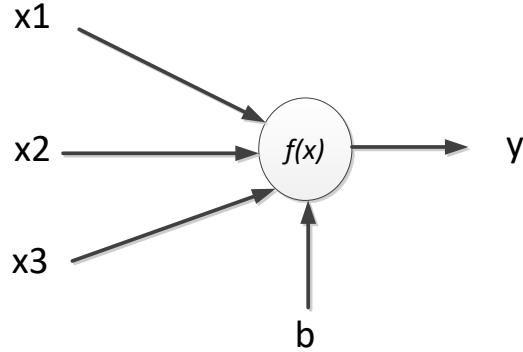


Figure 3.5: Single perceptron neuron

In the testing phase, the neuron can be given any input pattern even the perceptron has not seen it before. In nearly any supervised learning machine; like a neural network, the perceptron must be shown first patterns with correct answers (training phase), then a new unseen pattern should be tried in the testing phase and tested with the required specific transfer function. There are many functions may be used as transfer functions depending on a specific application. The most popular transfer functions are shown in Fig. 3.6.

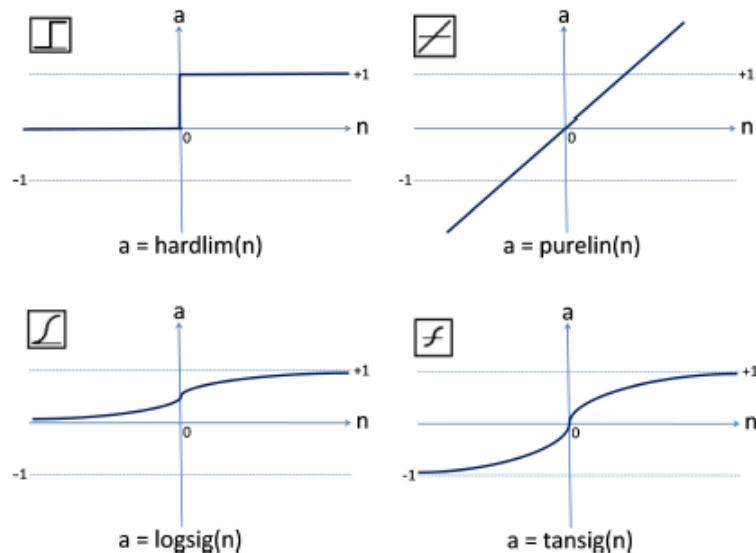


Figure 3.6: Most popular transfer functions

The neural networks can be configured to be comprised of:

- single layer or multi layers,
- feed-forward or back propagation form,
- single or multi transfer functions

in different various configurations which depend completely on the task to be performed, the network complexity degree, and output computation complexity.

A multi-layer feed-forward neural networks has been selection to classify the landmark image data sets which composed of both historic and modern images. The network structure is shown in Fig. 3.7.

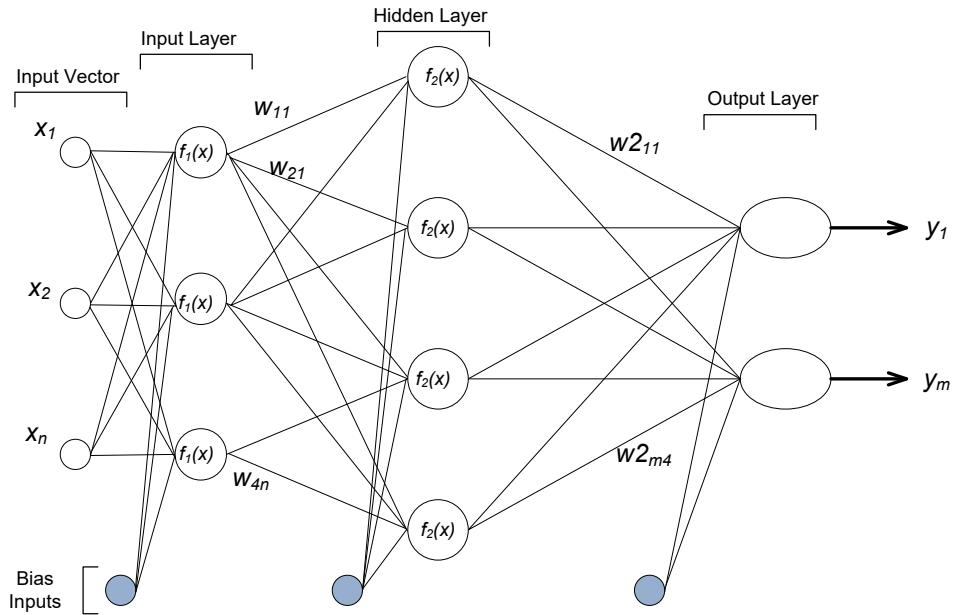


Figure 3.7: Multi-layer feed-forward neural network

3.3.1 Neural Network Training

For our system of feed-forward neural network, a back propagation training algorithm has been used. The error energy function is E calculated as follows [101]:

$$E = \frac{1}{2} \sum_{j,p} (e_j)^2 = \frac{1}{2} \sum_{j,p} (t_j - z_j)^2 \dots \dots \dots \quad (3.3)$$

where $j = 1, 2, \dots, N_L$, N : No. of Nodes, and L : state vector of nodes

$p = 1, 2, \dots, P$, P : No. of training sample sets

t is the largest output vector

Z is the actual output vector

The weights are updated to minimize this error function E according to the gradient-descent algorithm as follows:

$$\Delta W(k) = W(k+1) - W(k) = -\eta \frac{\partial E[W(k)]}{\partial W} \quad \dots \quad (3.4)$$

where η is the learning rate decreasing function or constant, and k is the iteration number.

The weights of the output layer are updated as:

$$\Delta w_{j,i}^{(L)} = \eta \delta_j^{(L)} x_i^{(l-1)} \quad \dots \quad (3.5)$$

where $\delta_j^{(L)} = (t_j - z_j^2)$, $j=1, 2, \dots, N_L$, $i=1, 2, \dots, N_{L-1}+1$ \dots (3.6)

The weights of the remaining layers are updated as:

$$\Delta w_{j,i}^{(l)} = \eta \delta_j^{(l)} x_i^{(l-1)} \quad \dots \quad (3.7)$$

A couple of image sets feature image set and non-feature image set are fed into the neural network in the training phase. Feature image set is comprised of 100-150 images contain the feature to be tested for every landmark, while the non-feature image set contains 50 images for the same specific feature under test. As a preprocessing step to facilitate the image subset selection process, every landmark image set is studied separately on a manual basis to select the highest detectable features with Gabor filtering depending on several visual and technical factors to select the candidate features for each landmark set. For example, a careful study of Hagia-Sophia landmark reveals that two regions are assumed to have accurately detectable features which are the dome and the internal bow. A dome is a unique object that is surrounded by a plain texture area while

the internal bow is a unique texture object in the landmark image. The neural network output neurons are set proportional to the number of candidate features of a landmark under investigating. For the Hagia-Sophia landmark, the output neurons will be set to two. These two outputs will be appeared in a mutually exclusive basis, i.e. when a feature exists in a test image, one of these output neurons will be greater than the threshold value of 0.8 or less than -0.5 and vice versa. The complete image classification process is described in algorithm 3.1 [98].

Algorithm 3.1:

```

Read the test feature-images.
Read the test non-feature images.
Set the number of candidate features.
Set the training parameters:
    Epoch=300;
    Performance: MSE=0.3 x 10-3;
    Gradient= 10-6;
Set neuron number to 1,
Set all Neuron Detection factors to zero.
Set Overall Matching factor to 1.
While (neuron number ≤ number of candidate features)
{
    Read neuron output;
    If (neuron output ≥ 0.8 )
        Set Neuron Detection factor to 1.
        Multiply Overall Matching factor by      Neuron Detection
factor.
        Increment neuron number.
    }
    If (Overall Matching factor =1)
        Classify image as Matched images;
    Else
        Classify image as Un-Matched image
}

```

To select the training sets of the eleven landmarks, a careful study of the image sets of these landmarks was done to select the most dominant features and depending on the image subset selection approaches and techniques of section 2.3, the eleven landmarks training sets have been categorized into three groups: one-feature training sets, two-features training sets, and three-features training sets as follows:

- One-feature landmarks: Coliseum of Rome, Dome of the Rock of Jerusalem, Empire State Building at New York, Whitehouse of Washington DC.
- Two-features landmarks: Eiffel Tower at Paris, Hagia-Sophia of Turkey, Taj Mahal of India, Machu-Picchu of Peru.
- Three-features landmarks: Parliament-Hill of Canada, Pyramids of Giza, Stonehenge of the United Kingdom.

The training set features are extracted from the landmarks images and fed into the network. For the one feature group, the extracted feature training set images are shown in Figures 3.8 through 3.11.

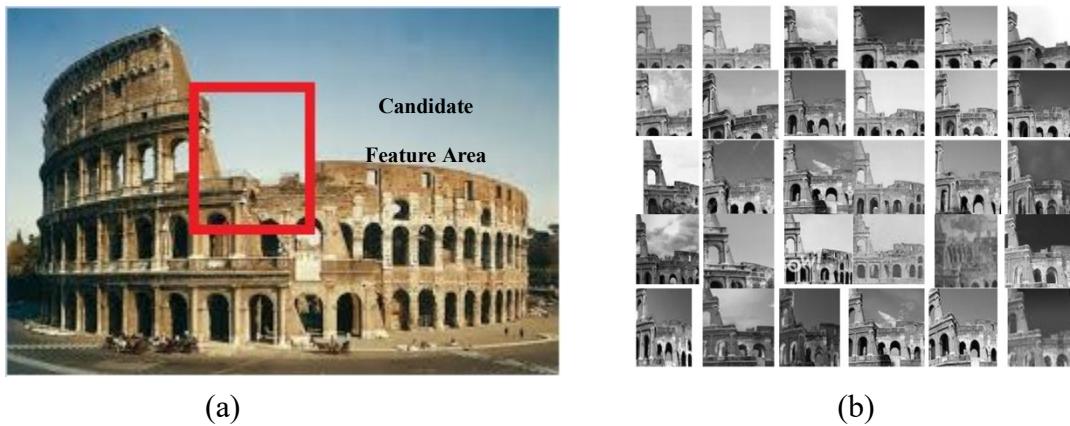


Figure 3.8: Coliseum of Rome image (a) The candidate feature area (b) Sample of training set for area 1

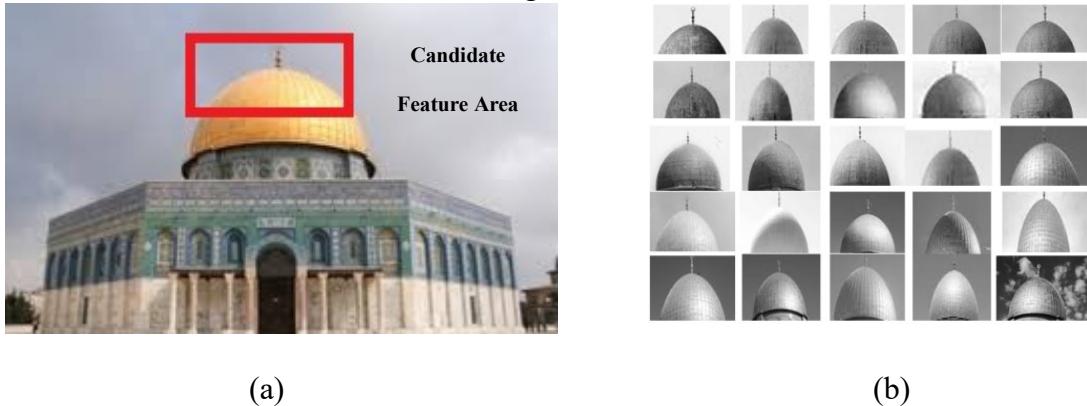


Figure 3.9: Dome of the Rock of Jerusalem image (a) The candidate feature area (b) Sample of training set for area 1

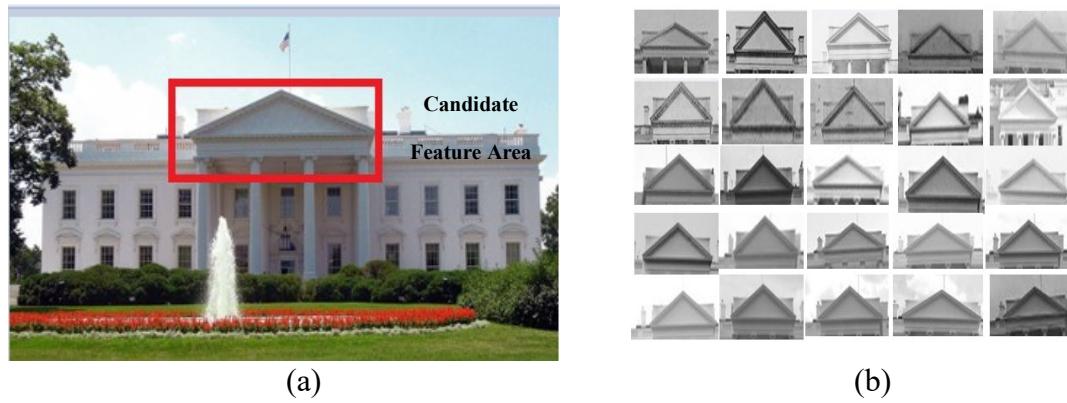


Figure 3.10: Whitehouse of Washington image (a) The candidate feature area (b) Sample of training set for area 1

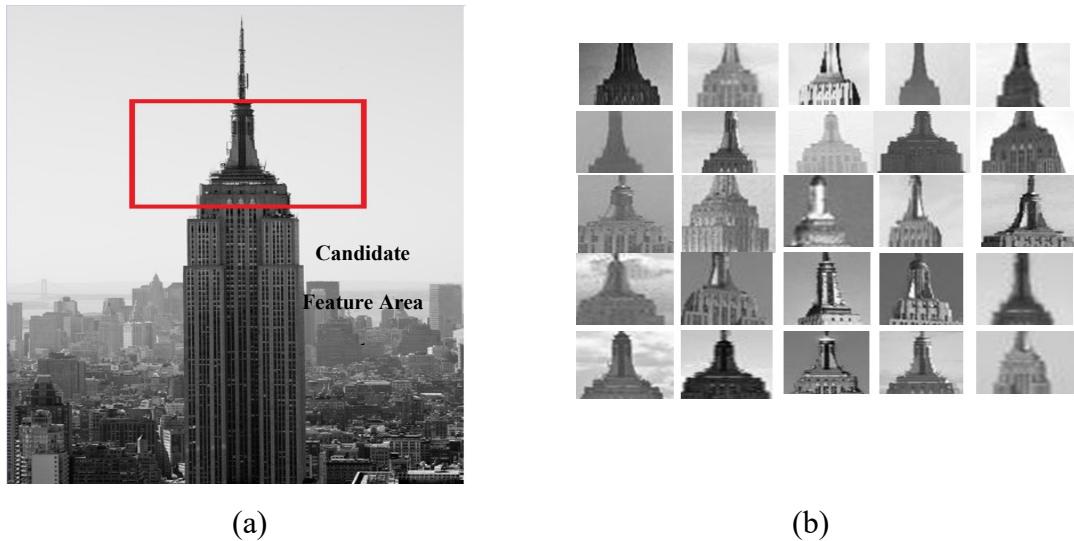


Figure 3.11: Empire State Building of New York image (a) The candidate feature area (b) Sample of training set for area 1

The extracted feature training set images of the two-features group, are shown in Figures 3.12 through 3.15.

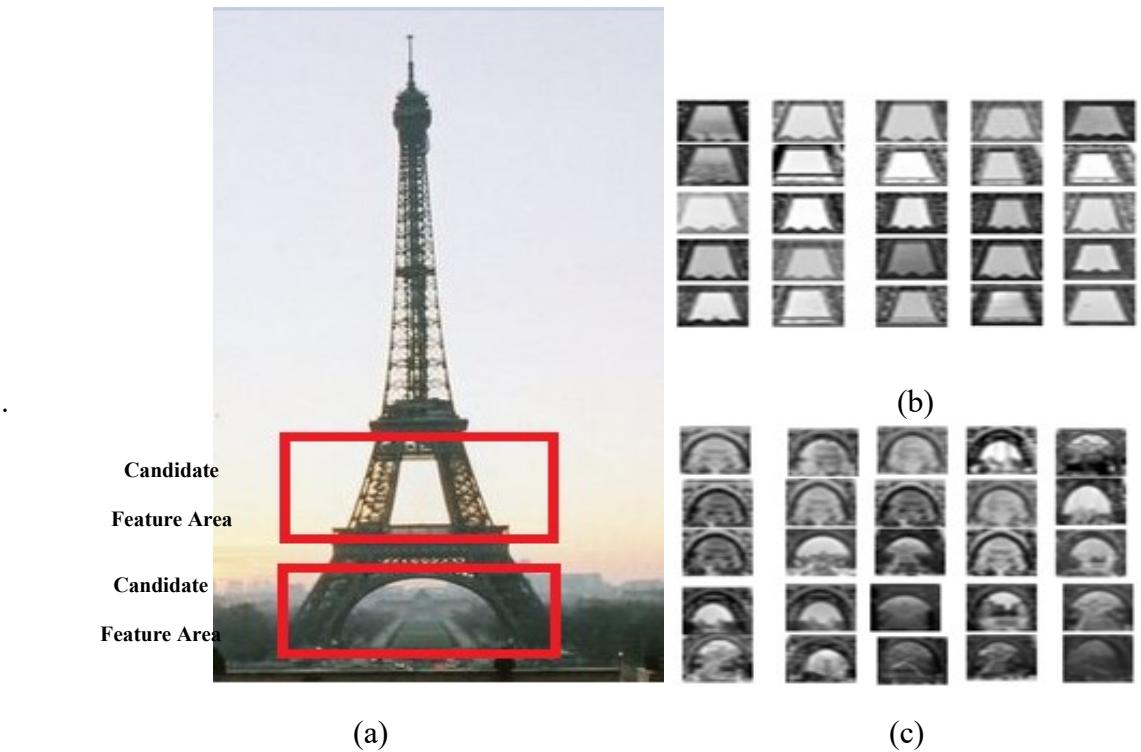


Figure 3.12: Eiffel Tower image (a) The two candidate feature areas (b) Sample of training set for area 1(c) Sample of training set for area 2

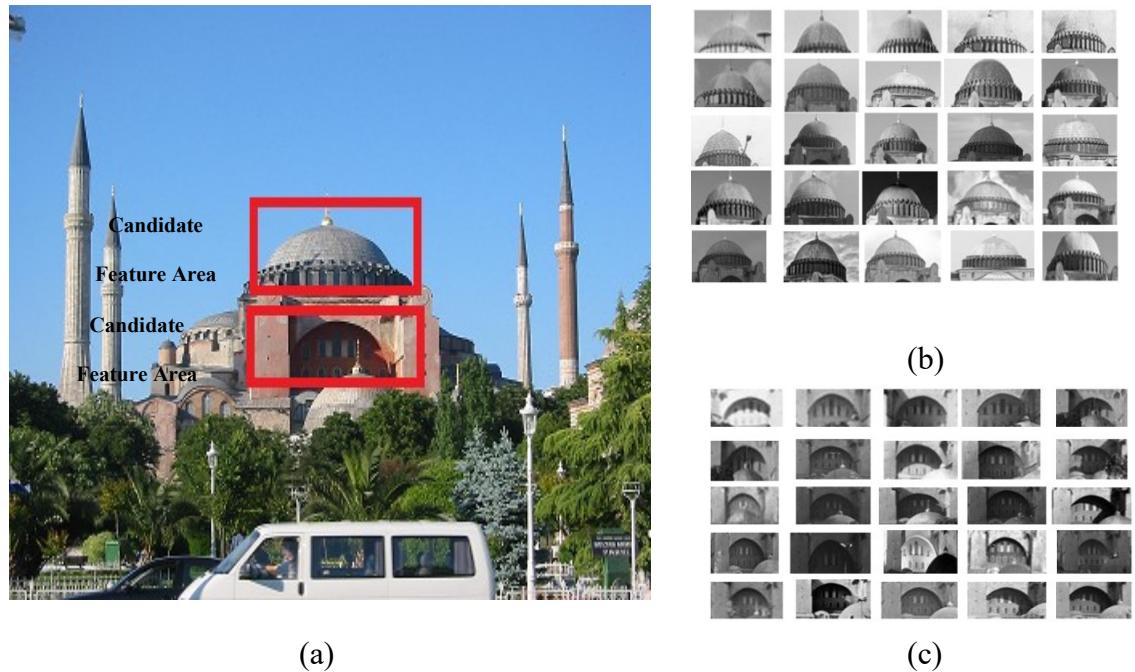
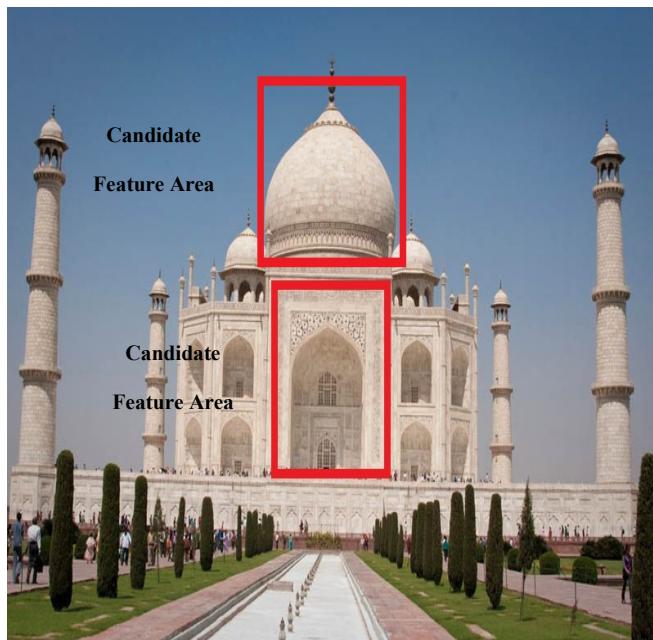
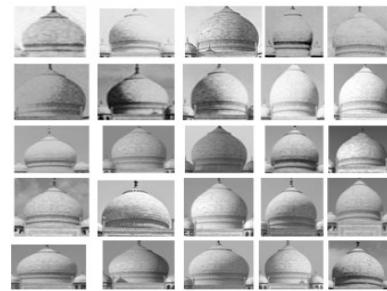


Figure 3.13: Hagia Sophia of Turkey image (a) The two candidate feature areas (b) Sample of training set for area 1(c) Sample of training set for area 2



(a)



(b)



(c)

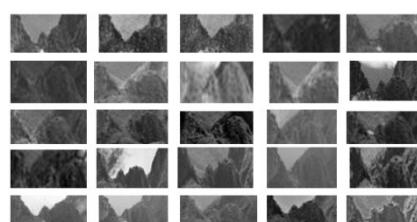
Figure 3.14: Taj-Mahal of India image (a) The two candidate feature areas (b) Sample of training set for area 1(c) Sample of training set for area 2



(a)



(b)



(c)

Figure 3.15: Machu-Picchu of Peru image (a) The two candidate feature areas (b) Sample of training set for area 1(c) Sample of training set for area 2

While the three-features landmarks candidate features are shown in Figures 3.16 through 3.18.

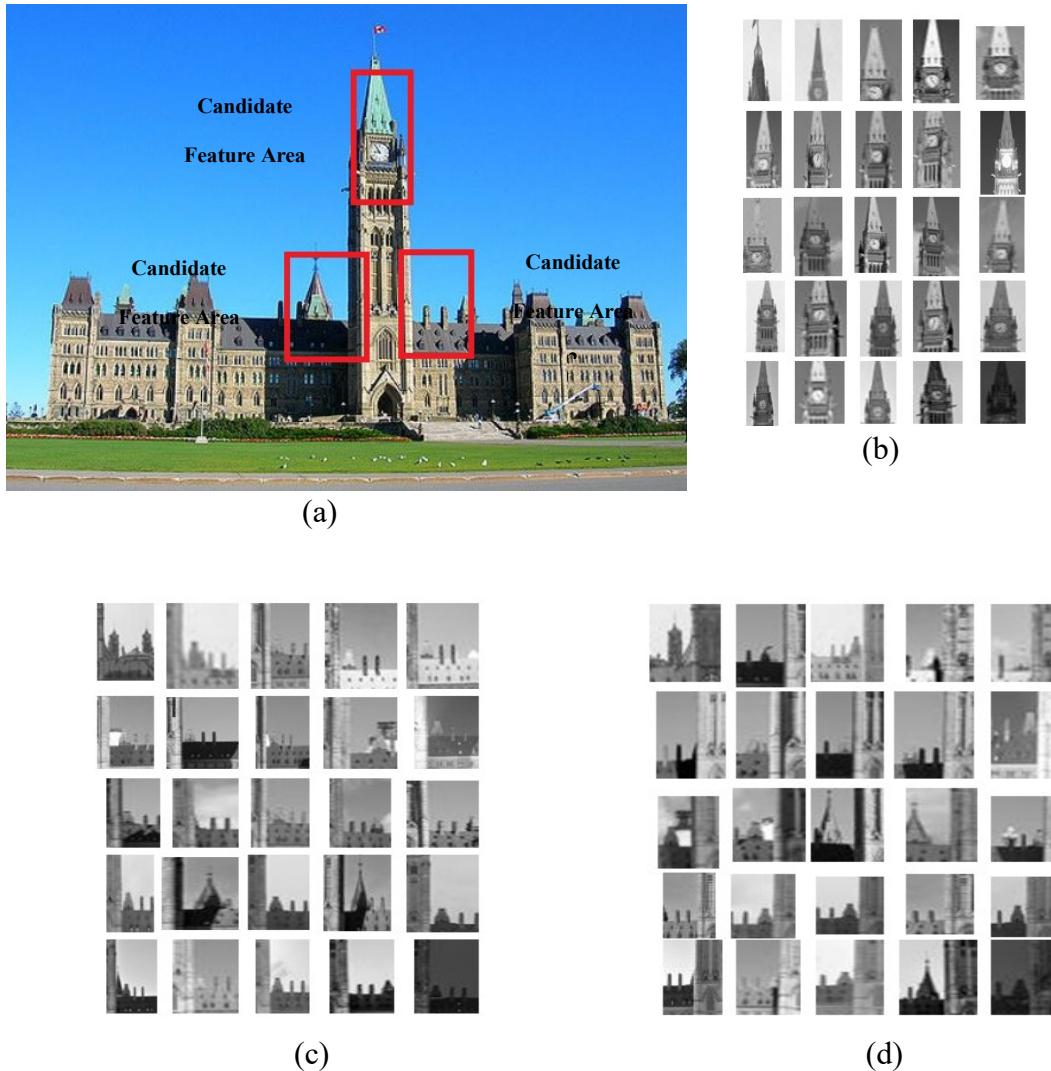


Figure 3.16: Parliament Hill of Canada image (a) The three candidate feature areas (b) Sample of training set for area 1 (c) Sample of training set for area 2 (d) Sample of training set for area 3

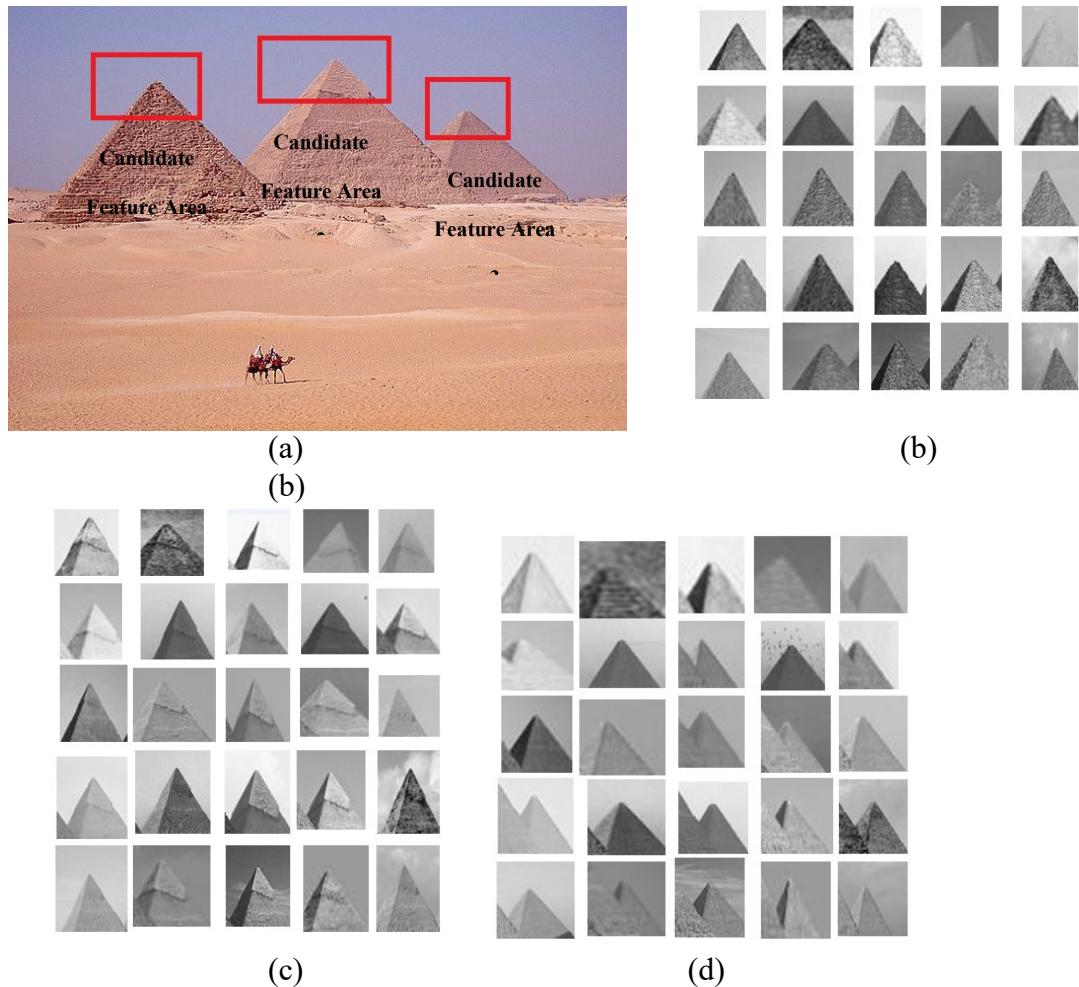
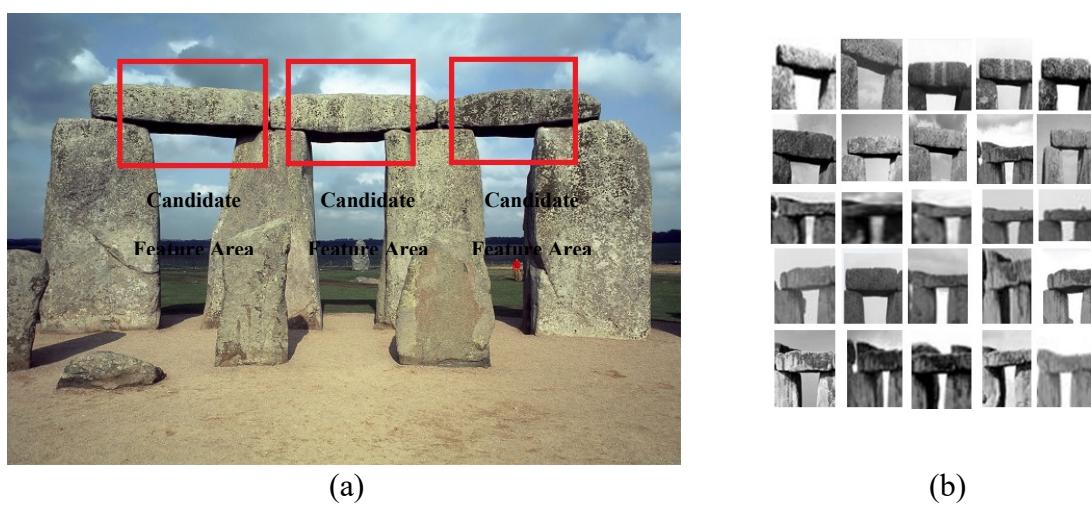


Figure 3.17: Pyramids of Giza image (a) The three candidate feature areas (b) Sample of training set for area 1(c) Sample of training set for area 2 (d) Sample of training set for area 3



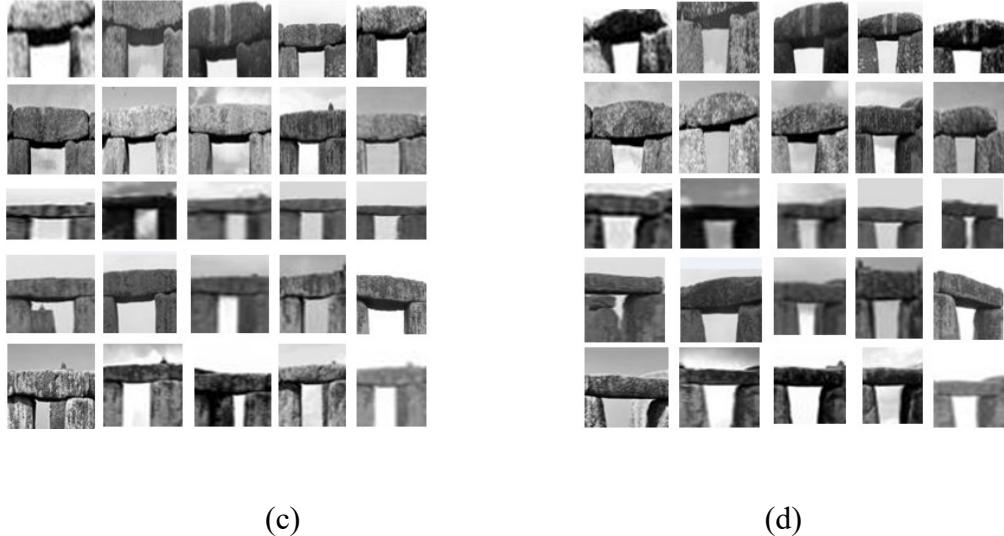
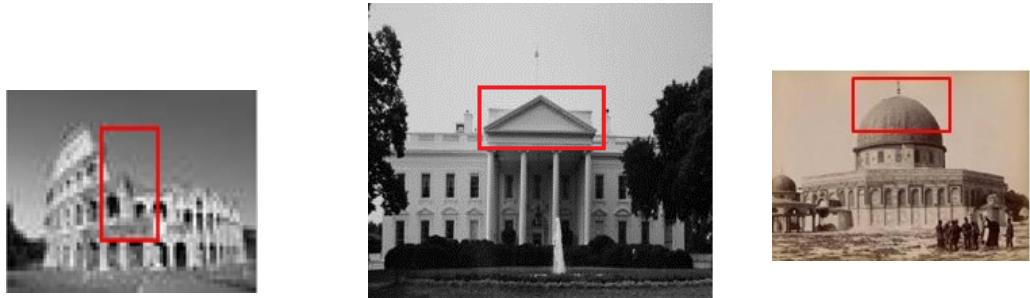


Figure 3.18: Stonehenge of the United Kingdom image (a) The three candidate feature areas (b) Sample of training set for area 1(c) Sample of training set for area 2 (d) Sample of training set for area 3

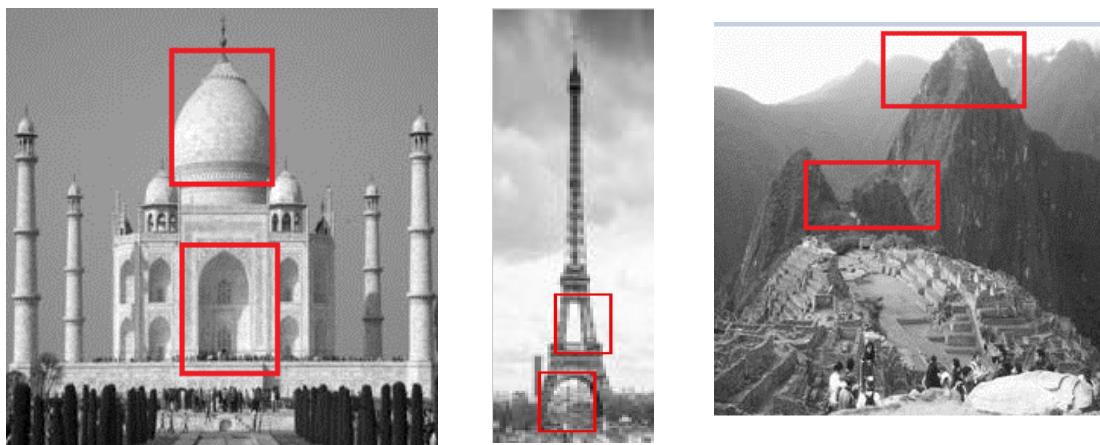
3.4 Experimental Results and Analysis

The collected raw data of eleven landmarks which consists of 50,000 to 350,000 images for each landmark was fed and Algorithm 3.1 was applied to each bunch of landmark images. The outcome of the classification system was two folders for each landmark. The first folder contains the landmark images with one, two, or three features as classified earlier, and the second folder contains the images failed to obey algorithm 3.1. The selected images were more processed to improve their contrast and light and then cropped to keep the most relevant parts.

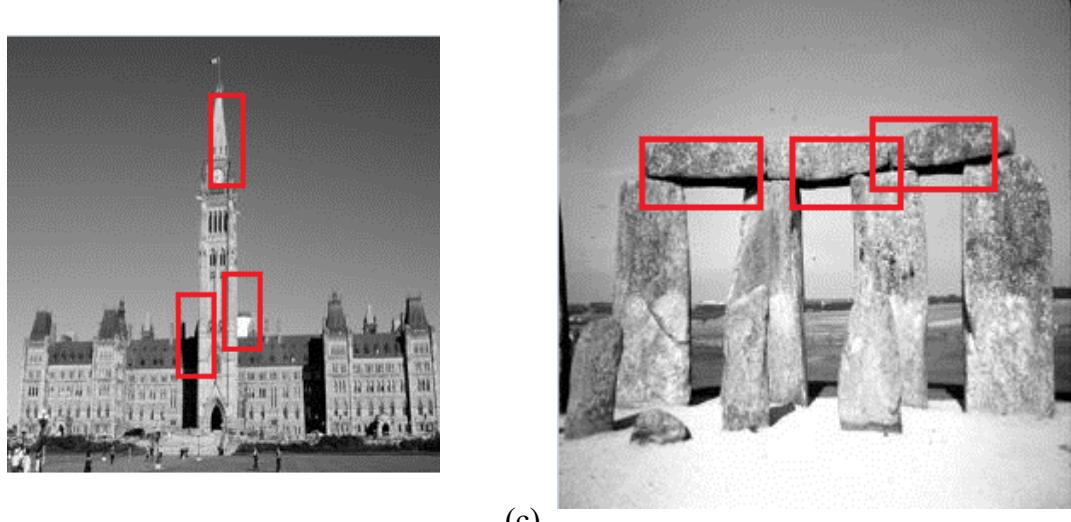
The features are most likely detected and Fig. 3.19 shows examples of correct extracted features.



(a)



(b)



(c)

Figure 3.19: Correctly detected features (a) One candidate feature images (b) Two candidate features images (c) Three candidate feature images

Along with the positive detected features, an exception of problematic false positives has been detected as shown in Fig. 3.20. The false positive problem was partially addressed by selecting the features with highest matching scores.

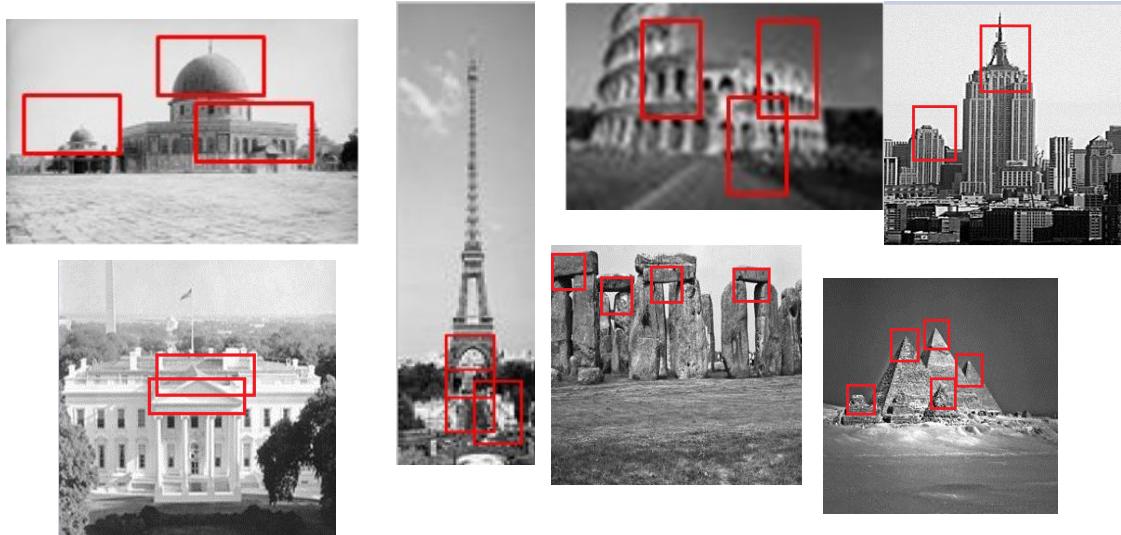


Figure 3.20: False positive detected features

Applying Algorithm 3.1 on the data set for landmarks with one-, two- and three-candidate feature reveals the results shown in Table 3.1:

Table 3.1: Applied algorithms results on landmark images collected from the image websites

No. of Candidate Features	Total Images	True Positive (tp)	False Negative (fn)	False Positive (fp)	True Negative (tn)
One-Feature Image Set	405,302	194,025	106,803	71,338	33,136
Two-Features Image Set	482,789	174,336	108,044	139,571	60,838
Three- Features Image Set	308,712	188,282	65,126	40,975	14,329
TOTAL	1,196,803	556,643	279,973	251,884	108,303

Table 3.1 reveals that the classification algorithm performed effectively on the raw data of each landmark which is comprised of both modern and historic images and non-

landmark images as well. This algorithm well merged Gabor filtering with the neural network to detect and extract the dominant landmark image features despite the diverse capturing environments such as lighting conditions, camera technology and geographical variations and hence, to classify images as belonging to the specific landmark or not.

The analysis of Table 3.1 results was done using the precision and recall metrics to calculate precision, accuracy, recall, and F-measure factor which is the harmonic mean of the measured precision and recall [30].

Table 3.2: Accuracy table of the algorithm applied on all landmarks image sets

No. of Candidate Features	Precision $tp/(tp+fp)$	Recall $tp/(tp+fn)$	Accuracy $(tp+tn)/(tp+tn+fp+fn)$	F1 Score $2*tp/(2*tp+fp+fn)$
One-Feature Image Set	0.7311	0.6449	0.5604	0.6853
Two-Features Image Set	0.5553	0.6173	0.4871	0.5847
Three- Features Image Set	0.8212	0.7429	0.6563	0.7801
TOTAL	0.7088	0.6769	0.5806	0.6925

The results shown in Table 3.2 declare that the precision and the accuracy of this mechanism increased as the number of the candidate features of landmark images increased, and the overall performance enhanced by increasing the number of the extracted features. The precision and accuracy results shown in Table 3.2 reveals that the overall accuracy test of the results is relatively high.

For example, the precision of the three-feature image set is 0.822, and the accuracy is 0.6536 which are good classification metrics in the sense that these metrics (precision, recall, accuracy, and F1) highest score is one while the lowest score is zero. These two measurements; which approach one, reflect good performance.

Image subset selection sequence of operations illustrated in Fig. 3.21. The operation sequence takes place as follows:

- The operation starts by an inquiry about a certain address, guide mark, or coordinates. This inquiry is matched with the UNESCO landmark database to decide while this inquiry is a landmark.
- If it is a landmark, the system searches in the Flickr and Google Images websites to detect the oldest available image of this landmark.
- The oldest image year is checked, if the year is greater than or equal 1800 and less than 1980, then images of every ten years inserted into a folder with a designation label of this period and the operation is repeated until the year 1980. Otherwise; if the year is greater than 1980, then images of every year are downloaded and kept in a separate folder for each year.
- All the images are tested to classify them as landmark images and non-landmark images; this test is performed in two stages:
 - The first stage to put images classified as roughly landmark images which may be true positive (TP) and false negative (TN), while the roughly non-landmark images consists of true negative (TN) and false negative (FN) images.
 - Second classification round classifies the landmark images as true positive images and sent to the landmark folder and false positive image sent to non-landmark images folder. Meanwhile, the non-landmark images classified as true negative images which sent to only non-landmark images, while the false negative images are sent to only landmark images.

A matching test then performed on all images in the landmark folder by simply subtracting each pair of images to check if an image is duplicated in this bin if a

duplicated image is detected, it is removed from this bin.

The image subset selection algorithm applied on all sets of landmarks images picked up the images with highly detectable features despite the considerable variances in images views, scales, and lighting conditions in an automated manner shows that this subset selection mechanism effectiveness is reasonably accepted and represents a promising image set categorization and selection [31].

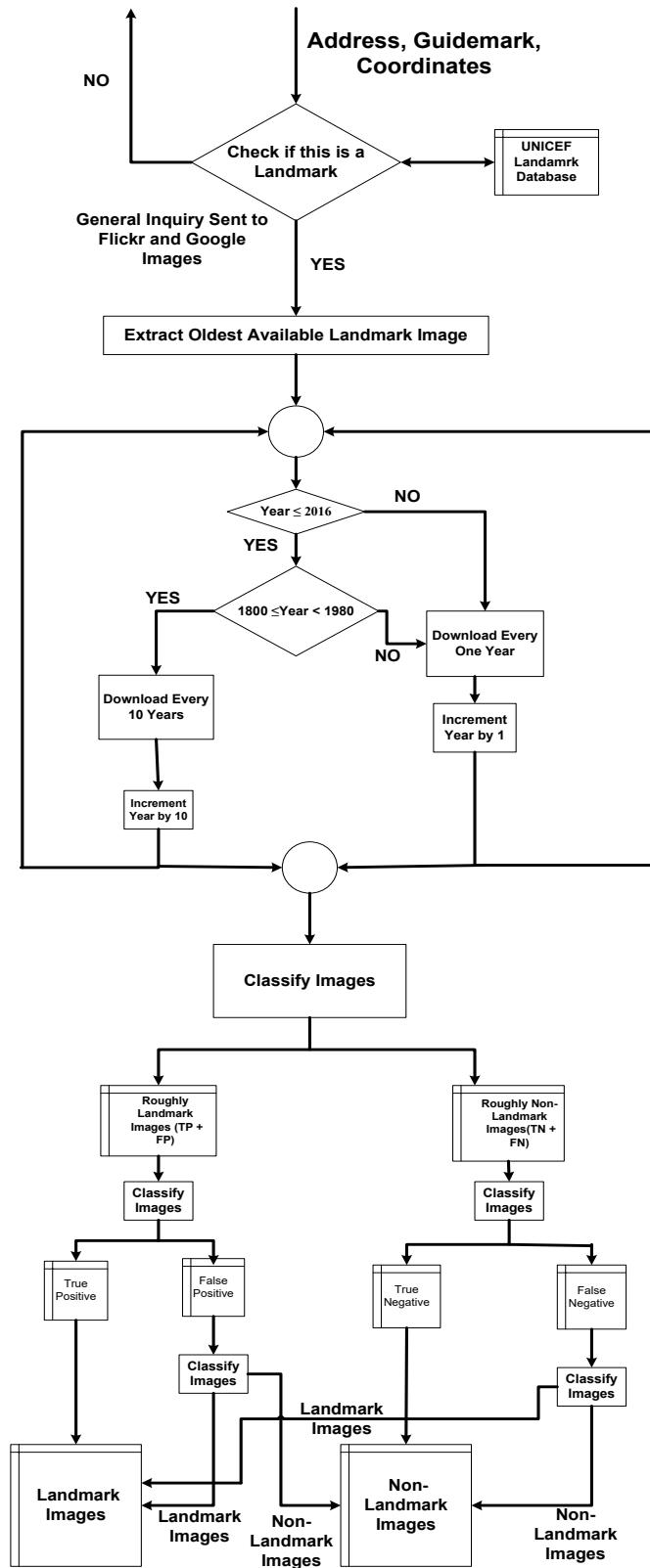


Figure 3.21: Flow chart of operation sequence in image subset selection subsystem

Chapter 4

Automatic Feature Extraction

In an attempt to perform an automated image registration of landmarks' historic and modern images, the feature extraction process should extract highly reliable features optimally serve in the image registration process. So, a careful study of the image set nature, the required registration objective, and time limits and computation complexities, supported by techniques, methods, and criteria of Sec. 2.4 in Chapter 2 will formulate the framework of the desired image feature extraction technique. This study should be constrained by:

- Selection of consistent features across images, such as pixels, edges, or objects.
- Decision of the most suitable technique either dense vs. sparse, parametric vs. non-parametric, single- or multi-modal methods ...etc.
- Obey the computation complexity constraints to find the required parameters within a planned time frame.

This will guide to the appropriate techniques and approaches.

In this chapter we'll demonstrate how to address the difficulties and challenges of automatic feature extraction in a mixed set of historic and modern images. In section 4.1, the feature extraction through the optical flow of SIFT features of modern and historic images is explained. The success of this technique in extracting features is explored and contrasted with standard SIFT-Flow techniques [22, 105, 120, 121] in the approach performance in terms of computation complexity and fields of application as shown in Table 5.1 Chapter 5 in this section as well. While Section 4.2 will present standard

feature extraction techniques application on the image set of landmarks to study and the performance of standard feature extraction techniques and methods on a mixed set of historic and modern images. The proposed ORB/SURF hybrid technique offers improvements, and we compare this technique to other standard techniques with an analysis of its performance, effectiveness, and computation complexity. In section 4.3 we will highlight a general performance evaluation of dense SIFT of optical flow technique on a mixed of historic and modern landmark image sets with the standard feature extraction techniques applied to the same data sets.

4.1 Feature extraction: Optical Flow of Extracted Features

Inspired by optical flow in which an image is aligned to its next image [105], the optical flow estimation [106] and SIFT descriptor [8] are used to inspect the feature flow which is used to precisely register the images over a long timeline.

The optical flow of a pair of images can be estimated as if these images were ordered regardless of the original capturing sequence. The optical flow approaches, which are based on the local Taylor series to compute the motion between two image frames which taken at times t and $t + \Delta t$ uses the partial derivatives with respect to temporal and spatial coordinates [107]. For a $2D+t$ dimensional case, a voxel at a location (x,y,t) with intensity $I(x,y,t)$ will have moved by Δx , Δy and Δt between two image frames, under the following brightness constancy constraint:

$$I(x,y,t) = I(x + \Delta x, y + \Delta y, t + \Delta t) \dots \quad (4.1)$$

And assuming a small movement, the image constraint at $I(x, y, t)$ with Taylor series can be developed to [107]:

$$I(x + \Delta x, y + \Delta y, t + \Delta t) = I(x,y,t) + \frac{\partial I}{\partial x} \Delta x + \frac{\partial I}{\partial y} \Delta y + \frac{\partial I}{\partial t} \Delta t \dots \quad (4.2)$$

These equations follow:

$$\left(\frac{\partial I}{\partial x} \right) \left(\frac{\Delta x}{\Delta t} \right) + \left(\frac{\partial I}{\partial y} \right) \left(\frac{\Delta y}{\Delta t} \right) + \left(\frac{\partial I}{\partial t} \right) \left(\frac{\Delta t}{\Delta t} \right) = 0 \quad \dots \quad (4.3)$$

Which results in:

$$\left(\frac{\partial I}{\partial x} \right) f_x + \left(\frac{\partial I}{\partial y} \right) f_y + \left(\frac{\partial I}{\partial t} \right) = 0 \quad \dots \quad (4.4)$$

where f_x and f_y are the x and y components of the optical flow of $I(x,y,t)$, and $\frac{\partial I}{\partial x}$, $\frac{\partial I}{\partial y}$, and $\frac{\partial I}{\partial t}$ are the derivatives of the image at (x,y,t) in the corresponding directions I_x , I_y , and I_t

and can be written in the derivative form:

$$I_x f_x + I_y f_y = -I_t \quad , \text{ or} \quad \dots \quad (4.5)$$

$$\nabla I^T \cdot \vec{f} = -I_t \quad \dots \quad (4.6)$$

Which results in an equation in two unknowns. The calculation of the optical flow requires another set of equations constrained with additional conditions. The common methods of optical flow determination introduce additional conditions for calculating the optical flow [108]:

- Phase correlation
- Block-based methods
- Methods based on partial derivatives of image signal and/or higher-order partial derivatives, such as:
 - Lucas-Kanade method
 - Horn-Schunck method
 - Buxton-Buxton method
 - Black-Jepson method
- Discrete optimization methods

The image set consists of modern and historic images as noted in Fig. 3.1. They are substantially different in pixel displacement between the objects may be larger than the typical regular motion fields in the optical flow. Hence, the optical flow methods [109, 110] which are based on the constraints of brightness constancy and coarseness level are not applicable. This can be handled through a discrete optimization method in which the image matching is performed at every pixel of the quantized search space [111] and the corresponding parameters minimize the distance between the reference and the sensed images. For each pixel of the images, the SIFT descriptor [8] is calculated to facilitate a descriptor matching scheme which produces a general correspondence in a preregistration step. To optimize the correspondence search, a discrete optimization method is formulated through an energy function.

4.1.1 Dense SIFT Features and Optical Flow

To facilitate the next steps of image registration, every pixel in the sensed image is labeled to represent the displacement relative to the reference image. The SIFT feature detection and extraction technique produces a sparse feature per image described by local gradient components [8]. Every image pixel has been processed to get a dense feature representation which supports a meaningful SIFT images. A meshgrid of a window of 16 x 16 pixels surrounding every image pixel have been selected and divided into sub meshgrids of 4 x 4 cells to calculate the orientation in 8 different angles to formulate every image pixel descriptor which is a vector of (4 x 4 x 8) dimension as shown in Fig. 4.1.

The SIFT images are produced by mapping the dense SIFT descriptors with the dominant components of the RGB original image. This mapping is performed by

selecting the highest three components of each descriptor with the dominant RGB color space.

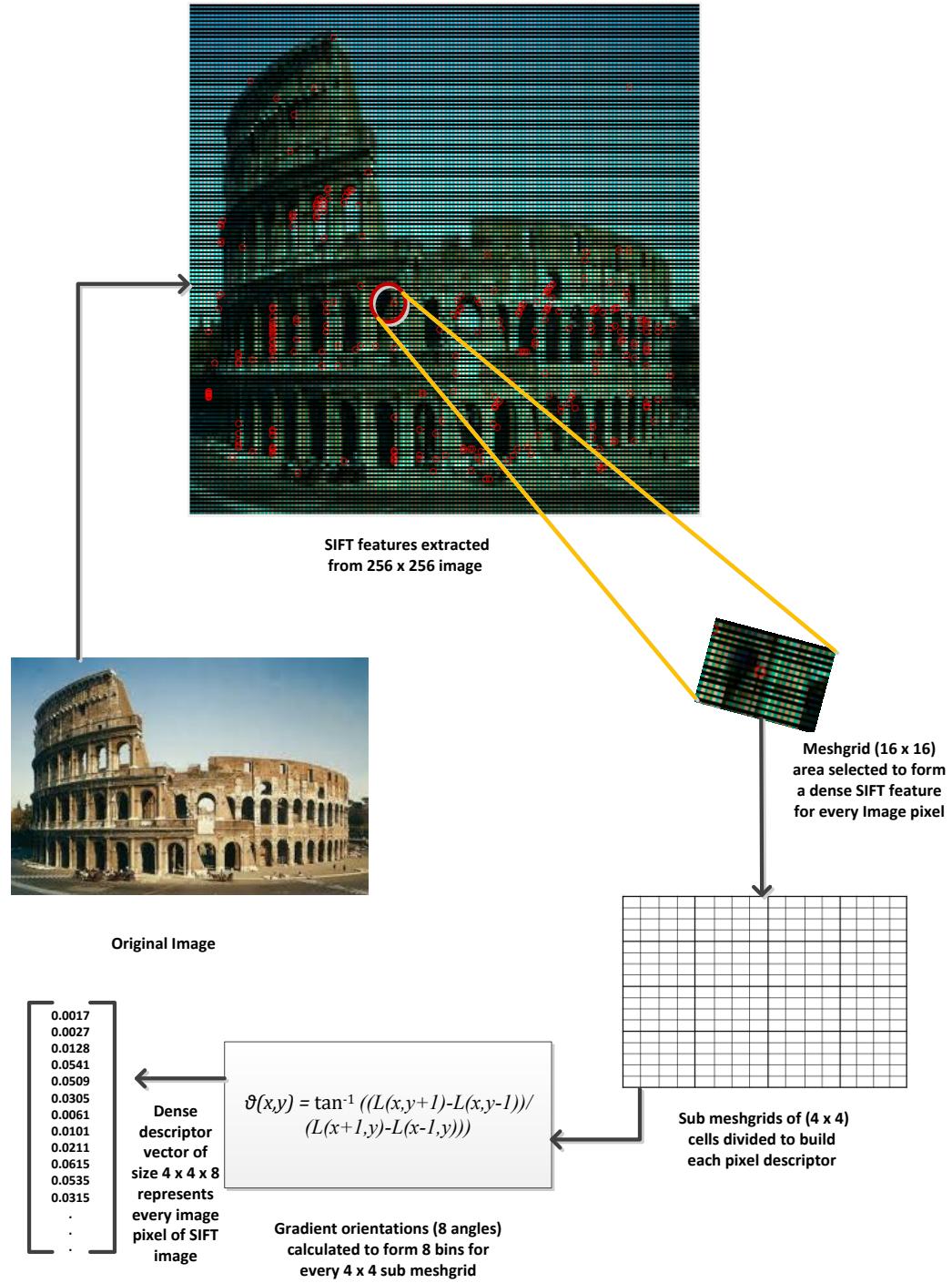


Figure 4.1: Dense SIFT descriptor calculation for every image pixel of the original landmark image

4.1.2 Results and Discussion

The dense descriptors for the detected SIFT features are extracted from the historic and modern images of the landmarks. To get the best match of images captured in different conditions, the window size surrounding each pixel changed for the values 8, 16, and 32. The obtained results are shown in Figures 4.2 through 4.4.

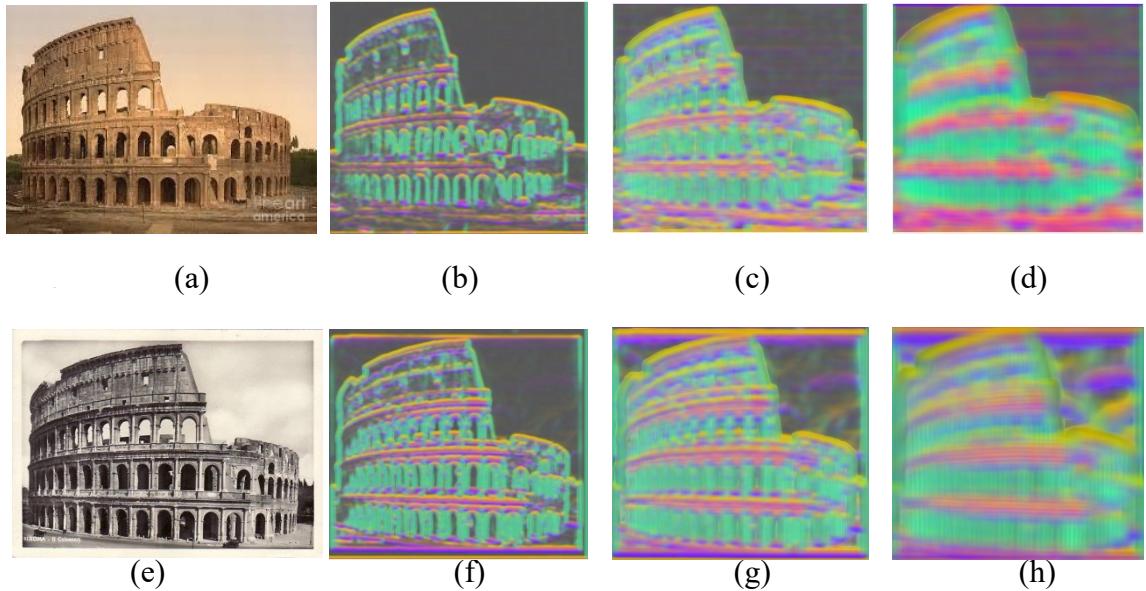


Figure 4.2: Dense SIFT descriptor for the Coliseum landmark images of (a) 1910 and (e) 1920. The dense SIFT descriptors for window size 8 surrounding the pixel shown in (b) and (f), window size 16 shown in (c) and (g), window size 32 shown in (d) and (h).

The dense SIFT image samples of all other landmarks are depicted in appendix A.

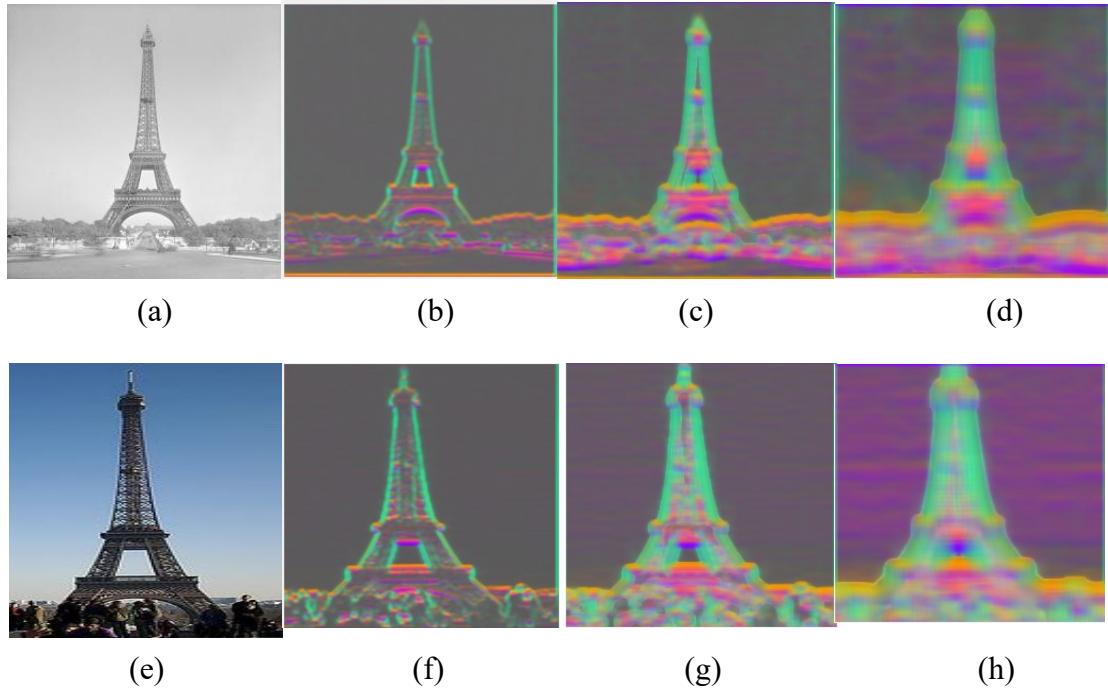


Figure 4.3: Dense SIFT descriptor for the Eiffel Tower landmark images of (a) 1898 and (e) 2008. The dense SIFT descriptors for window size 8 surrounding the pixel shown in (b) and (f), window size 16 shown in (c) and (g), window size 32 shown in (d) and (h).

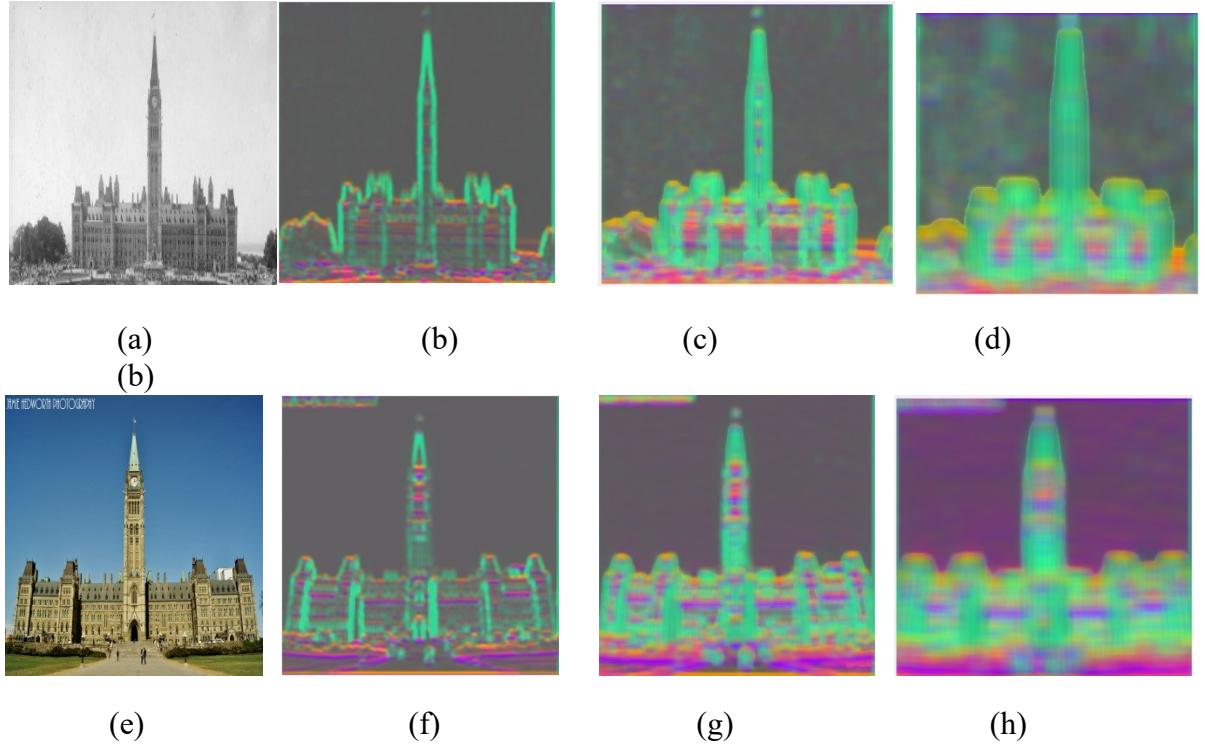


Figure 4.4: Dense SIFT descriptor for the Parliament Hill of Canada landmark images of (a) 1927 and (e) 2015. The dense SIFT descriptors for window size 8 surrounding the pixel shown in (b) and (f), window size 16 shown in (c) and (g), window size 32 shown in (d) and (h).

We notice that images (b) and (f) of the three figures 4.2 through 4.4 reveals more detailed features; especially the edges can be clearly recognized while images c through h of the three figures seem to be blurry. This is because of the window size; which is a meshgrid of 8x8 of images (b) and (f). This offers clear edges in the cost of computation complexity. Other images (c through h) of the three figures look blurry, with the regions that visually appear of the same color are almost similar regions or contain the similar features.

4.2 Feature extraction: Standard Techniques

4.2.1 Feature Extraction and Matching Techniques

Image matching can be achieved depending on different image components such as shape, texture, or color. However, as we aim to register images based on views from roughly the same view point, a number of point features have to be extracted to calculate the registration (or warp) matrix accurately. The accurate alignment can be achieved through calculating accurate point correspondence. Hence, we are studying and testing point feature matching techniques in this work.

To ensure that the feature extraction results do not vary according to chosen (or imposed) conditions, and hence, to achieve effective matching, we generally search invariant properties [43]. Crucial invariant means the immunity to change in scale, illumination level, rotation, viewpoint, or location. Feature detectors and descriptors have been widely presented in the literature.

Many of these techniques have been reviewed in Chapter 2 Sec. 2.5.2, with an explanation to their features and a comparison between their strengths and drawbacks.

Landmark image sets captured over long time periods and various different capturing conditions have been applied to the standard feature extraction techniques to evaluate the efficiency of these techniques upon mixed set of historic and modern images. The selected techniques are Harris, SIFT, SURF, BRIEF, ORB, and BRISK. The evaluation is performed by matching images of the eleven landmarks. Every image will be matched with all other images in the set. The analysis of matching results will show the strengths and weaknesses of these techniques and provide good decision basis for the suitability of these techniques to get an effective processing of our data set. Moreover, an extra effort

has been done to experience some hybrid detectors/descriptors of these techniques and led to a novel application of a hybrid ORB/SURF matching technique which shows more accurate results and performs more effectively as it was applied in matching our data set of mixed historic and modern images.

4.2.1.1 Matching Procedure

Generally, the point matching of any two images using any technique will show too many matching points of both inliers and outliers. Comparing two images of Eiffel Tower landmark using SURF technique as point detector and descriptor illustrated in Fig. 4.5 shows too many matching points. The keypoint matching was performed using the BruteForce matching technique by matching N feature descriptors of reference image (Fig. 4.5 a) with M feature descriptors of sensed image (Fig. 4.5 b). For the NxM features, each feature descriptor of the reference image will find the closest descriptor of the feature set of the sensed image [102]. This technique produces best results of a minimal number of outliers compared to other matching techniques [16].

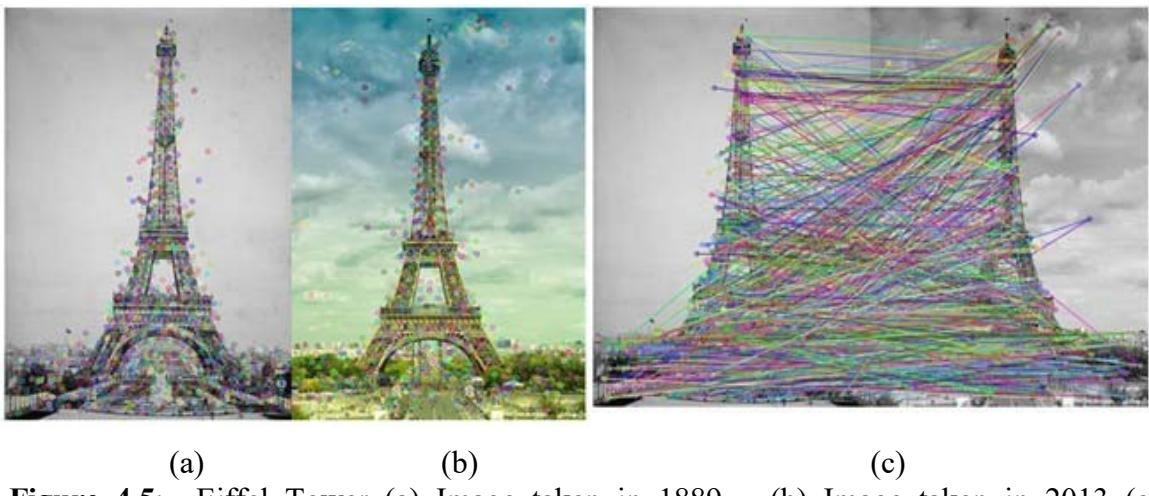


Figure 4.5: Eiffel Tower (a) Image taken in 1889 (b) Image taken in 2013 (c) SURF/SURF matching of left keypoints with those on right image using Brute Force algorithm

4.2.1.2 Disparity Gradient Filter

The disparity gradient filter was applied to get a better matching accuracy by keeping the inliers and discard or minimize the number of outliers. The outlier matches are removed by the disparity gradient filter based on the deviation threshold of the matching points motion vectors from the median of the whole set [103]. The disparity gradient filtering was originally developed to accurately determine how to merge images in stereo vision [103]:

$$d = \frac{|r_L - r_R|}{\frac{1}{2} |r_L + r_R|} \quad \dots \dots \dots \quad (4.11)$$

where d is the disparity gradient, vector r_L in the left image and the corresponding vector r_R is in the right image.

The disparity gradient filter (DGF) was applied to a pair of images of on the matching keypoints extracted using SURF/SURF, and the results are shown in Fig. 4.6 [16].

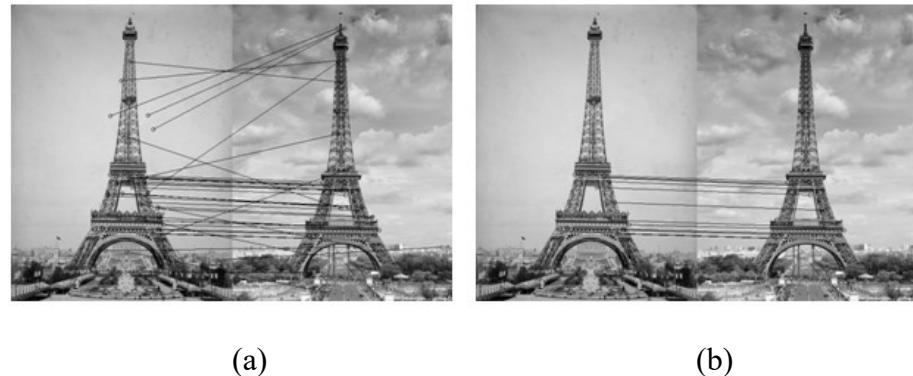


Figure 4.6: Applying disparity gradient filter (DGF) on features extracted using SURF/SURF (a) Before applying DGF (b) After applying DGF[104]

4.2.2 Harris Corner Detector

Harris corner detector which is based on the eigenvalues of the second-moment matrix is a well-known detector in feature detection literature [44]. This detector was

applied to the landmark data set of modern and historic images. The combination of Harris/Harris refers to Harris detector and Harris descriptor, and this combination will be used from now on in this thesis. An application example of Harris/Harris is shown in Fig. 4.7 of two images of Pyramids of Giza landmark; a historic image from 1880 and a modern image from 2006. Matching process of these two images shown in Fig 4.7 (a) while Fig 4.7 (b) shows the matching process of image from 1880 with itself.

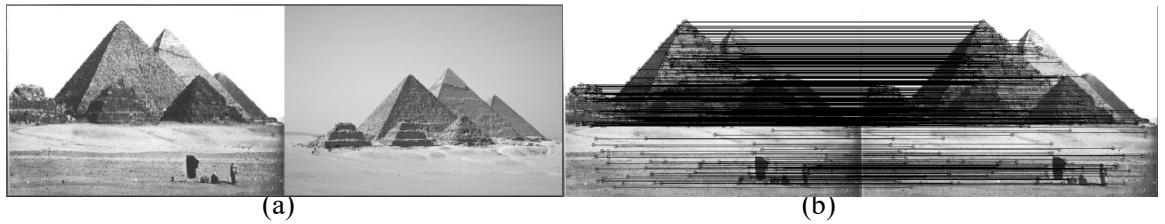


Figure 4.7: Applying HARRIS/HARRIS feature extraction on (a) images of 1880 with image of 2006 and (b) image of 1880 with itself for the Pyramids of Giza landmark

Table 4.1 shows the complete matching results of the images of the Eiffel tower landmark. Every image of the Eiffel Tower 10 images data set is matched with all other images in the set based on the Harris/Harris detected features. Green cells (or 1) refer to highly matched images, while the pink cells (or 0) refer to poorly match or unmatched images [16].

Table 4.1: Eiffel Tower landmark matching using HARRIS/HARRIS matcher

HARRIS/HARRIS	1889	1889b	1890	1937	1945	1967	1967b	1986	2005	2011
1889	1	0	1	1	1	0	0	0	0	0
1889b	0	1	1	0	0	0	0	0	0	0
1890	1	1	1	1	0	1	0	0	0	0
1937	0	0	0	1	0	1	0	1	0	1
1945	0	0	0	0	1	0	0	0	0	0
1967	0	0	1	1	1	0	0	0	1	1
1967b	0	0	0	0	0	0	1	0	0	0
1986	0	0	0	1	0	0	0	1	1	0
2005	0	0	1	0	0	0	0	0	1	1
2011	0	0	0	0	0	0	0	0	1	1

The complete image comparisons of all landmarks based on features extracted by Harris/Harris features detector/descriptor tables are shown in Appendix B. The tables (4.1 and tables B.1 through B.9 in Appendix B) show the overall poor performance of Harris corner as feature extractor on the mixed set of historic and modern images of landmarks.

4.2.3 SIFT/SIFT

Scale invariance of the SIFT feature descriptor allows a variety of applications where the invariant properties are essential due the nature of these applications. Scale space principle of the SIFT descriptor utilized the scale invariance by detecting stable keypoint locations using scale-space extrema in the difference-of-Gaussian (D_oG) function convolved with the image $I(x, y)$, which can be computed from the difference of two nearby scales separated by a constant multiplicative factor k as described by equations 2.24 through 2.26 in Chapter 2. SIFT/SIFT was applied to our landmark image data sets. The matching results of one of the Eiffel tower landmark is shown in Table 4.2 which shows good matching results while Table 4.3 shows the matching results of the Taj-Mahal landmark with a poor matching. Other SIFT/SIFT matching tables of the other landmarks are shown in Appendix C.

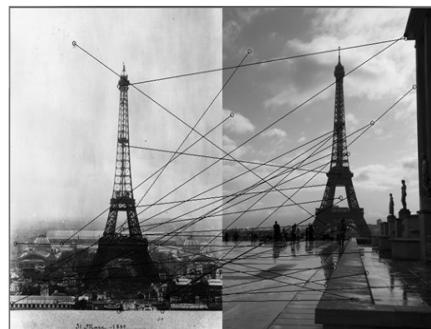
Table 4.2: Eiffel Tower landmark matching using SIFT/SIFT matcher

SIFT/SIFT	1889	1889b	1890	1937	1945	1967	1967b	1986	2005	2011
1889	1	0	0	1	0	1	1	1	0	0
1889b	0	1	1	1	0	1	1	0	0	0
1890	1	1	1	1	1	1	0	1	0	0
1937	1	1	1	1	1	1	1	1	1	1
1945	0	0	1	0	1	1	1	1	0	1
1967	1	1	1	1	1	1	1	1	0	1
1967b	1	1	1	1	1	1	1	1	0	1
1986	0	0	1	1	1	1	1	1	1	1
2005	0	0	1	1	0	0	1	1	1	1
2011	1	0	0	1	1	1	1	1	1	1

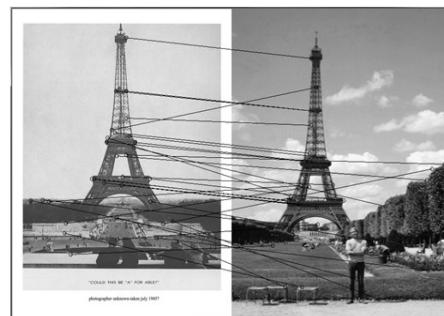
A careful study of SIFT/SIFT application on all landmarks data sets shows that this detector/descriptor combination performs well with some landmarks and images while shows poor performance on others. As an example, SIFT/SIFT matching of Eiffel Tower landmark is the highest accuracy while Taj-Mahal is the worst among other landmarks. Fig. 4.8 shows examples of SIFT/SIFT applications on mixed historic and modern images of Eiffel Tower, (a) shows a bad matching for the image pair 1889-2005 while (b) shows a successful matching for the pair 1945-1967 (relatively two old images). Of note, the corners representing matching efforts of the oldest (historic) to the newest (modern) image are where the failures are most common [16].

Table 4.3: Taj-Mahal landmark matching using SIFT/SIFT matcher

SIFT/SIFT	1874	1890	1910	1910b	1943	1966	1986	1993	2006	2011
1874	1	1	0	0	1	0	0	0	0	0
1890	0	1	0	0	0	0	0	0	1	0
1910	0	0	1	0	0	0	0	0	0	0
1910b	0	0	0	1	0	0	0	0	0	0
1943	0	0	0	0	1	0	0	0	1	0
1966	0	0	0	0	0	1	0	0	0	0
1986	0	0	0	0	0	0	1	1	0	0
1993	0	0	0	0	1	0	1	1	1	0
2006	0	1	0	0	0	0	0	1	1	1
2011	0	0	0	0	0	0	0	0	1	1



(a)



(b)

Figure 4.8: Keypoint matching of images of Eiffel Tower (a) Bad matching in the pair 1889-2005 and (b) Successful matching in the pair of 1945-1967b

4.2.4 SURF/SURF

The SURF local feature detector and descriptor are partly inspired by a SIFT descriptor which is several times faster than SIFT and more robust against different image transformations than SIFT [9]. Scale and rotation invariance properties of SURF detector and descriptor achieved through depending on integral images to reduce the computation times by the determination of Hessian matrix with three integer operations using precomputed integral image. The entry of the integral image $I\Sigma(x)$ at a location $a=(x, y)^T$ represents the sum of all pixels of the input image I within a rectangular region formed by the origin as described in equations 2.27 through 2.29 in Chapter 2.

The matching results using SURF/SURF matching for the Eiffel tower are shown in Table 4.4 [16] and the complete results of all the landmarks are shown in Appendix D.

Table 4.4: Eiffel Tower landmark matching using SURF/SURF matcher

SURF/SURF	1889	1889b	1890	1937	1945	1967	1967b	1986	2005	2011
1889	1	1	0	0	0	0	1	1	0	0
1889b	1	1	0	0	1	1	1	1	1	0
1890	1	1	1	1	0	1	0	1	0	1
1937	0	1	0	1	0	1	0	1	1	1
1945	1	1	1	1	1	1	1	1	0	0
1967	1	1	0	1	1	1	1	1	1	1
1967b	1	0	0	0	0	0	1	1	1	0
1986	0	0	0	1	0	1	1	1	1	1
2005	0	0	0	0	0	1	1	1	1	1
2011	0	0	1	1	0	1	1	1	1	1

4.2.5 BRIEF/BRIEF

Binary strings as an efficient feature point descriptor (BRIEF) is a highly discriminative even when using relatively few bits and computed using simple intensity different tests [46]. BRIEF/BRIEF was applied to the landmarks image data sets. Fig. 4.9

shows its application to the Whitehouse landmark on two pairs of images while the full matching results of applying BRIEF/BRIEF are summarized in Fig. 4.12 [16].



Figure 4.9: Applying BRIEF/BRIEF feature extraction on (a) images of 1993 and 2012 and (b) on images pair 1948 - 2009 for the Whitehouse landmark

4.2.6 BISK/BRISK

Inspired by the leaders' feature detection and matching techniques SIFT and SURF, the binary robust invariant scalable keypoints (BRISK) feature detection and matching has been presented with lower computation complexity than SURF [10]. BRISK descriptor is computed based on the brightness comparisons by configuring a circular sampling pattern and based on a prior sufficient knowledge of the scene and camera. BRISK/BRISK applications results on the landmark data sets are summarized in Fig. 4.12, while Fig. 4.10 shows the feature extraction results on a pair of images of Stonehenge landmark.



Figure 4.10: Applying BRISK/BRISK feature extraction on the pair image of 1955-2009 for the Stonehenge landmark

4.2.7 ORB/ORB

Oriented FAST Rotated BRIEF (ORB) feature detector based on FAST and BRIEF feature detectors represents an attempt to provide a fast and efficient alternative to SIFT. ORB feature detector calculation is based on taking the intensity threshold between the center pixel and those in a circular ring about the center as a parameter, and computes the corner orientation which is called the intensity orientation [47]. The centroid calculation and the descriptor formulation are described in equations 2.31 through 2.33 in Chapter 2.

The ORB/ORB was applied to the complete image set of the 11 landmarks; the matching results of these images for all landmarks are shown in Fig. 4.12.

4.2.8 Hybrid ORB Detector /SURF Descriptor

In attempt to get a robust matching technique based on the standard techniques, many detector/descriptor hybrid techniques have been tested on the landmark images sets. The hybrid ORB/SURF is based on the ORB detector by relying on machine learning to address the ordering distribution problem of corner appearance, and the SURF descriptor, a vector-based descriptor generated through defining subregions around the identified feature with array elements [16]. This hybrid technique outperforms the performance of other standard techniques including the leading techniques such as SIFT/SIFT and SURF/SURF.

Hybrid ORB/SURF matching results as compared to SIFT/SIFT and SURF/SURF matching of modern to historic images shows a high margin of matching improvements. Table 4.5 shows the matching results of ORB/SURF compared to SURF/SURF results for the empire state building landmark. The blue cells refer to improvements in matching process compared to table C.4 matching of SURF/SURF for the same landmark.

Table 4.5: Empire State Building landmark matching using ORB/SURF matcher

ORB/SURF	1936	1940	1942	1955	1957	1959	1996	2007	2007b	2011
1936	1	1	2	0	2	0	0	1	0	1
1940	1	1	1	1	1	2	0	0	2	0
1942	0	2	1	1	0	2	1	2	2	2
1955	0	0	0	1	2	2	1	1	1	1
1957	1	0	1	1	1	1	2	1	0	2
1959	0	1	2	1	1	1	0	1	1	1
1996	0	0	0	0	0	1	1	1	1	1
2007	2	0	2	2	0	1	1	1	0	1
2007b	0	1	1	0	0	1	1	0	1	0
2011	0	2	2	2	2	2	2	1	2	1

4.2.9 Results and Analysis

We will classify the results into three parts. First one shows the success feature matching techniques in terms of accuracy tables, while the second part will show the runtime of each technique. Finally, we will focus on the effectiveness of our approach compared to other techniques [16].

4.2.9.1 Technique Accuracy

For each landmark, there were 100 tests performed and the successful ones divided by the total tests represent the accuracy. The accuracy is computed as the total matches divided by the total tests for each landmark.

The accuracy of the applied feature extraction techniques on our complete set of image can be classified into two ranks. The First rank is the high performance techniques which the accuracy rate of them is summarized in Table 4.6. Beside our proposed hybrid ORB/SURF technique, two other techniques which are SURF/SURF and SIFT/SIFT showed better matching results on the modern/historic image sets. Fig. 4.11 shows the accuracy comparison of each of the standard techniques SIFT/SIFT and SURF/SURF as well as the hybrid ORB/SURF technique [16].

Table 4.6: Accuracy of high performance image matching techniques for each landmark, with highest accuracy rates in bold [16].

Landmarks	SIFT/SIFT Accuracy	SURF/SURF Accuracy	ORB/SURF Accuracy
Coliseum	54%	40%	69%
Dome	27%	37%	39%
Eiffel	73%	63%	77%
Empire	30%	46%	70%
Hagia	23%	25%	28%
Machu	27%	33%	40%
Pyramid	16%	14%	26%
Stonehenge	28%	22%	33%
Taj	22%	29%	32%
Whitehouse	34%	33%	46%
Overall	33.40%	34.10%	46.00%

Table 4.6 shows that the hybrid ORB/SURF combination is more effective than the leading standard technique in the sense of technique accuracy. The accuracy benchmark based on the overall amount of accurate matched keypoints. We hypothesize that the ORB binary detectors accuracy in detecting both historic and modern features particularly due to the advantage of its dependency on the intensity threshold of cornerness measure. This measure is enhanced by a scale-pyramid of the image. The detected ORB features are described by defining subregions around the features and each consists of 64 floating-point number will increase the distinctiveness of the descriptor. Accuracy comparison of ORB/SURF with other leading two techniques SIFT and SURF are illustrated in Fig. 4.11 [16].

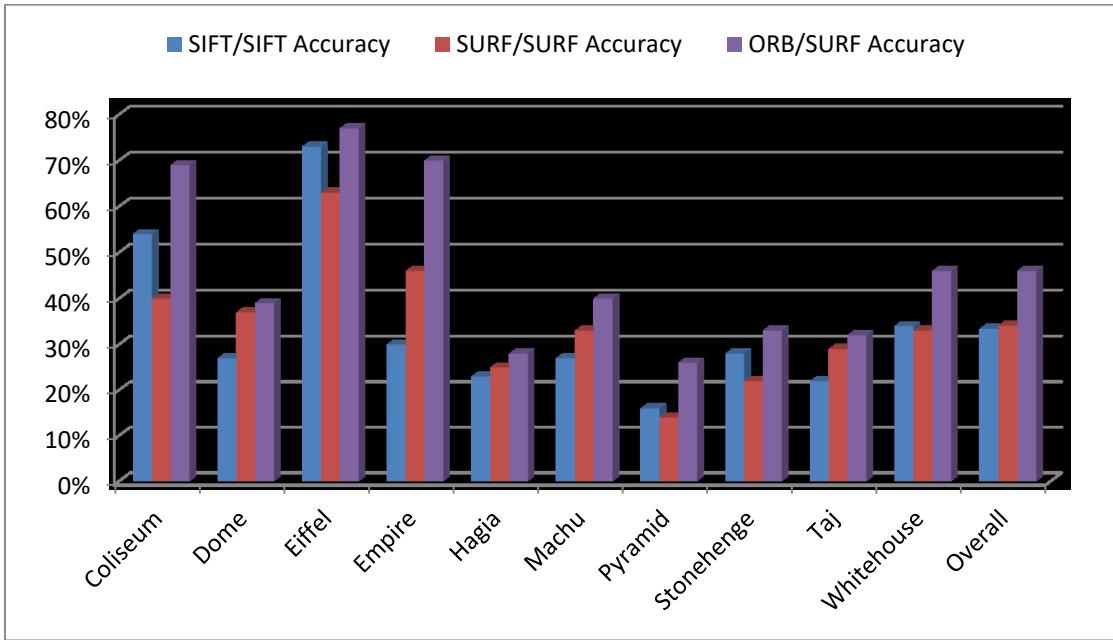


Figure 4.11: High performance feature matching techniques accuracy comparison

Fig. 4.12 summarizes the performance of the standard techniques in which we can note the poor performance of matching historic to modern images of the landmarks image data sets [16].

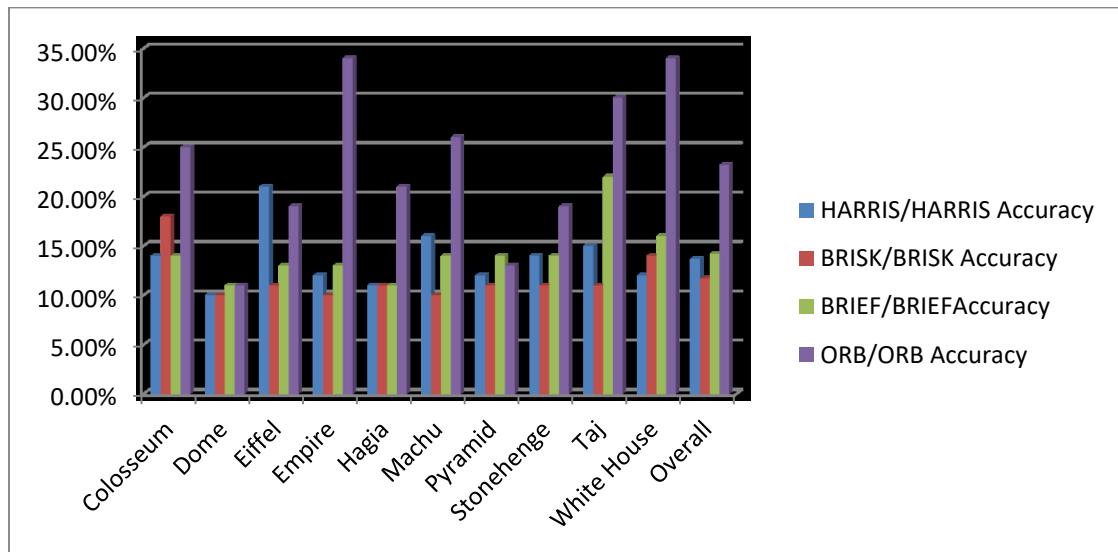


Figure 4.12: Low performance feature matching techniques accuracy comparison

4.2.9.2 Runtime

To compare the computation efforts of the three high accuracy techniques in dealing with a mixed set of historic and modern landmark images, the runtime of all the landmarks image sets processed by these techniques have been recorded. An extensive monitoring of test times on an Intel Core i7-3770 3.40 GHz CPU was recorded and are shown in Table 4.7. This table shows the average required runtime of each image pair processed by these three techniques.

In the runtime performance point of view, we can note that SURF/SURF performs better than the other two techniques followed by ORB/SURF. Taking into account that the matching environment is an off-line, we can conclude that the ORB/SURF performed in an acceptable time [16].

Table 4.7: Runtime of high performance matching techniques

Techniques	SIFT/SIFT	SURF/SURF (Threshold 400)	ORB/SURF ORB(1000 Points)
Runtime in ms	7.534	5.575	6.734

4.2.9.3 ORB/SURF Effectiveness

The evaluation of ORB/SURF technique compared to other leading standard and other standard techniques showed that this technique outperforms other techniques in modern to historic image matching. Tables 4.4 and C.4 show that this technique improves the behavior of SURF/SURF and SIFT/SIFT of dealing with associated problems of capturing environments and camera technology. This hybrid technique performs in an acceptable runtime of ORB/SURF compared to other techniques [16].

4.3 Dense SIFT of Optical Flow and Standard Techniques Evaluation

As a general evaluation of the results obtained by the standard feature detection and extraction techniques and the results of the dense SIFT descriptors of the optical flow, it is noticed that the standard matching techniques work are generally work well on either the modern images or images of flat and simple texture images as can be noticed in the lower right corners of matching tables (which show the matching process of probably all modern landmark images), or images of flat, simple texture, or gradual transition texture images like Eiffel Tower, Coliseum of Rome, and Stonehenge landmarks. While these techniques are poorly performed on matching a mixed of historic and modern images as can be noticed in the center part of the matching tables.

In contrast, we can note that the pixel dense descriptors of Figs. 4.2 through 4.5 represent a good matching basis by detecting the image features of both historic and modern images and offer more controllable factors like a window size.

Chapter 5

Aligning Images

The image registration process aims to overlay two or more images of the same scene taken from different viewpoints, at different times, and/or by different sensors in a geometrically optimum alignment overcoming the different imaging conditions [18]. Regardless of the image registration methods and techniques as pixel-based or feature-based, or local or global registration, it is required to establish the mathematical relationships that map pixel, feature, or object coordinates from one image to another [24].

Image registration methodology comprised different stages may be done manually or automatically. Manual object detection in which the salient and distinctive closed-boundary regions, edges, contours, corners, or line intersections are detected with an appropriate mathematical description represents a major step in an aim of image registration. Complementary steps of feature matching, transform model estimation, and image resampling and transformation [18] can also characterized as manual or automatic.

Our aim is an automatic image registration of landmark image sets of historic and modern images; which is a challenging task due to geometric deformations, radiometric discrepancies, and the required pre-knowledge of the models or detailed scene nature, has been realized and will be described in section 5.1 in this chapter. Section 5.2 will demonstrate the image warping process performed on the image with features extracted by the standard feature detection and extraction techniques.

5.1 Automatic Image Registration (Optimized Optical Flow of Detected Features)

Automatic image registration of modern to historic images is a challenging problem due to difficulties of capture environments and the various technologies of cameras used to capture these images. Optical flow provides information of intensity patterns of motion velocities in the sequence of images. Analogous to the tracking process of images which is computed through the optical flow based on dense sampling of the time domain, the flow of SIFT features assumes dense sampling of timelined images to implement the scene alignment [105].

Computation of optical flow is an important step in global image registration. The extracted dense feature descriptors have to be matched depending on the flow vectors and the flow field has to be smooth. To construct the energy function of the SIFT descriptor flow three parts have to be considered. The data part of the function which describes the SIFT descriptor to be matched with the flow vector [19]:

$$E_{data}(f) = \sum_g \| sd_1(g) - sd_2(g + f(g)) \|_1 \dots \dots \dots \dots \dots \dots \dots \quad (5.1)$$

where $g=(x,y)$ denotes the coordinates of reference and sensed images, sd_1 and sd_2 are the descriptors of SIFT features for reference and sensed images respectively.

The second part is the displacement data for the sensed image in order to find the optimal label for each pixel,

$$E_{displacement}(f) = \sum_g \eta (|u(g)| + |v(g)| + \sum_{(g,q) \in \epsilon} \min(\alpha|u(g) - u(q)|, d)) \dots \dots \dots \dots \dots \dots \dots \quad (5.2)$$

where $f(g) = (u(g), v(g))$ is the flow vector at g , the set ε contains the spatial vicinities, η is the weighting factor which measures how the point g devotes in the displacement field d [105].

The smoothness part controls the adjacent pixels to be similar:

$$E_{smoothness}(f) = \min(\alpha |v(g) - v(q)|, d) \quad \dots \quad (5.3)$$

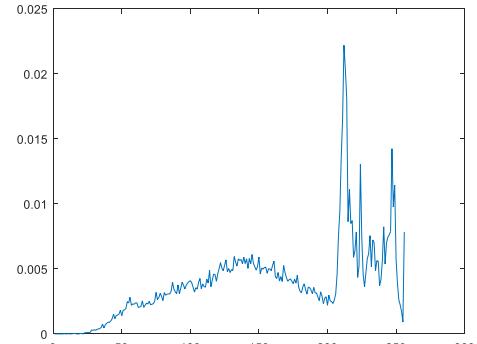
The three parts will comprise the energy function,

$$E(g) = E_{data} + E_{displacement} + E_{smoothness} \quad \dots \quad (5.4)$$

As the historic and modern landmark images features has been detected with sparse SIFT features in section 4.1 of chapter 4, the image histograms of landmark images show that the dense SIFT features provide an enhanced matching platform for the matching process and hence for the automatic image registration. Fig. 5.1 shows the mage histograms of the original images and the dense SIFT images.



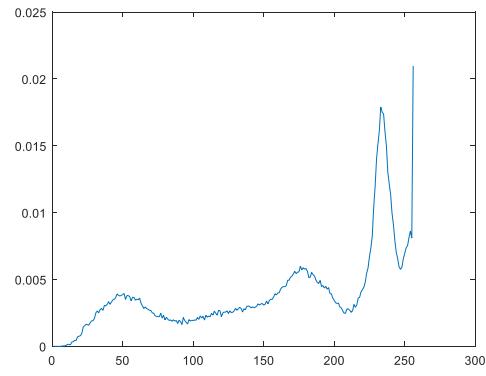
(a)



(b)



(c)



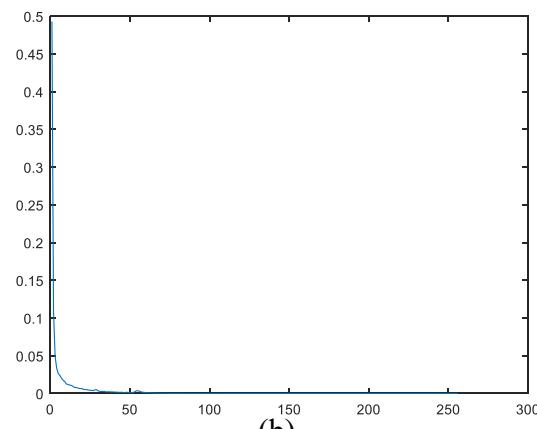
(d)

Figure 5.1: Image histograms for Coliseum landmark images (a) Image of 1910 (b) Image histogram (c) Image of 1913 (d) Image histogram

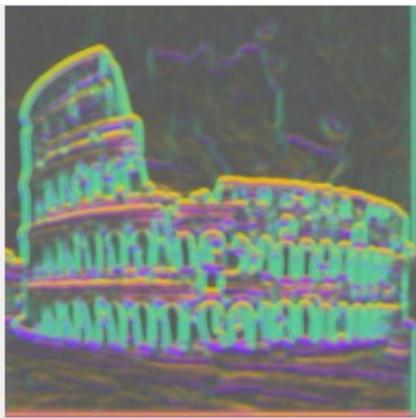
The histogram images of two historic images of Fig. 5.1 show obvious differences between the histograms as shown in Fig. 5.1 (b) and (d). While the histogram of the dense SIFT features of Fig. 5.2 shows that the differences have been almost diminished and the dense SIFT features clarifies the feature similarities illustrated in the same color in Fig. 5.2 (a) and (c) while the histograms are almost similar in (b) and (d).



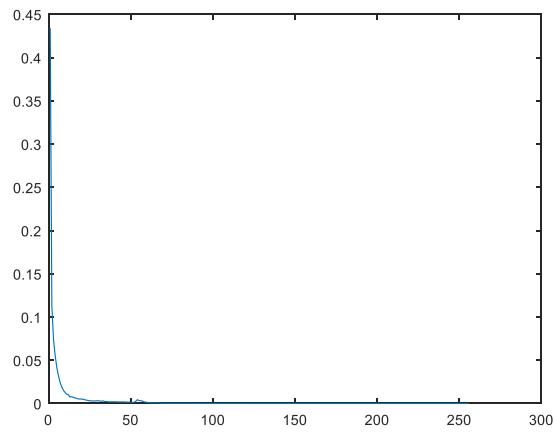
(a)



(b)



(c)



(d)

Figure 5.2: Dense SIFT images for images of Fig. 5.1 of Coliseum landmark images
 (a) Dense SIFT of image of 1910 (b) Image histogram (c) Dense SIFT of image of 1913 (d) Image histogram

It can be noticed from Figures 5.1 and 5.2 that the dense SIFT of two historic images coming from multi-sensors with cameras of unknown parameters give a homogeneous image base for further steps in the automated image registration by the optical flow of extracted features. More results of different landmarks mixed historic and modern images are shown in Fig. 5.3.

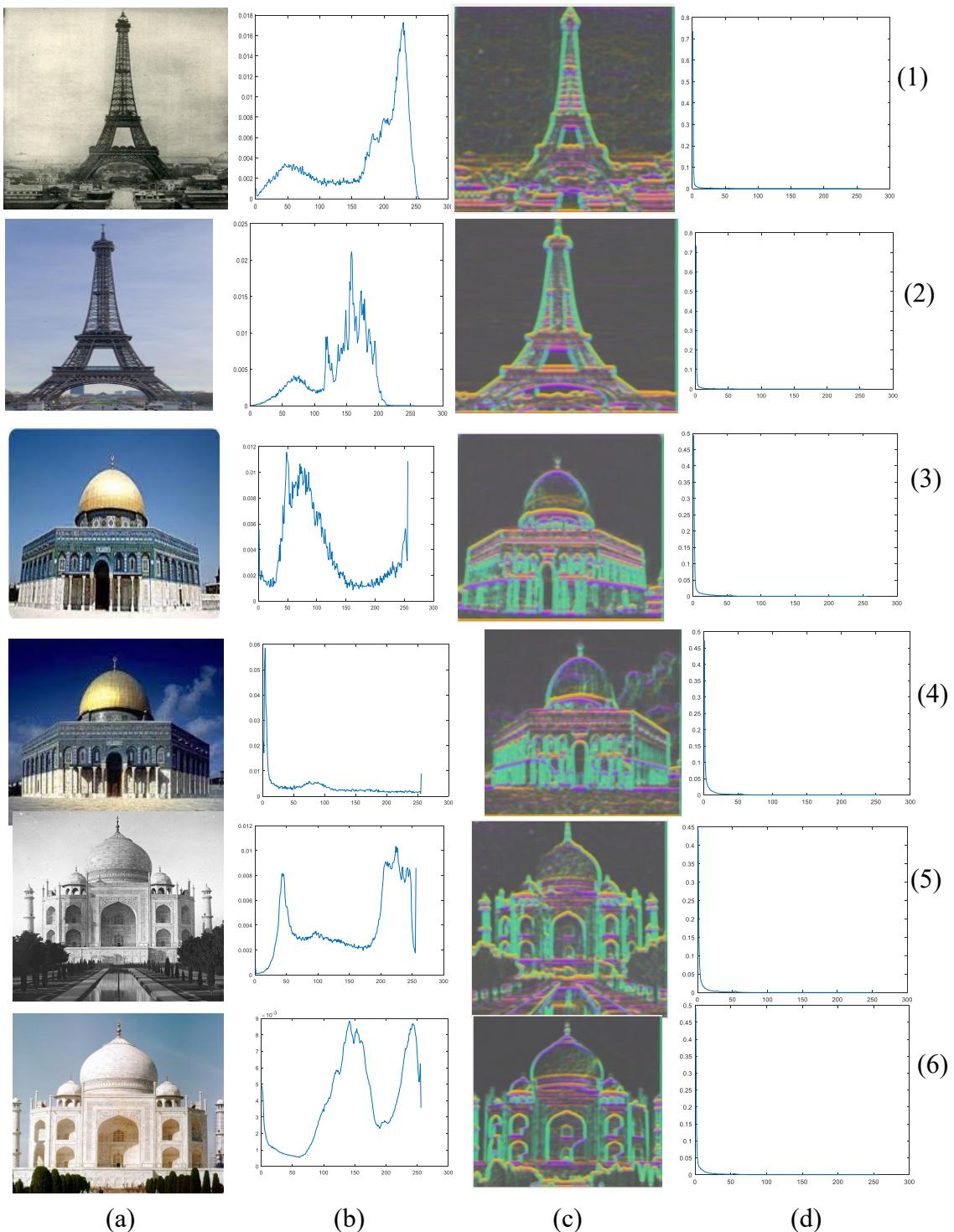


Figure 5.3: Image histograms for: column (a) rows (1:1890 and 2:2007) Eiffel Tower (3:2011 and 4:2015) Dome Rock (5:1910 and 6:1959) Taj-Mahal landmarks with their histograms in column (b), their dense SIFT features in column (c), and histograms of dense SIFT in column (d)

5.1.1 Loopy Priority Belief Propagation for Optical Flow

The correspondence process can be formulated as a discrete optimization problem (Equations 5.1 through 5.4). The optimization process is computed using a loopy priority belief propagation algorithm [21]. Given reference image R and the sensed image S . The target region T of the reference image has to be matched and aligned with an object in the sensed image S in an optimized way.

The discrete Markov Random Field (MRF) is proposed as an optimization basis for the matching process. The patches represented by their centers p consist of $w \times h$ regions are labeled by L of MRF, while the nodes of the MRF will be denoted with vertical and horizontal spacing of gap_x and gap_y pixels on an image lattice. The MRF nodes U are all lattice points whose $w \times h$ neighborhood the target region with edges \mathcal{E} will formulate a 4-neighborhood on the lattice [21] as shown in Fig. 5.4.

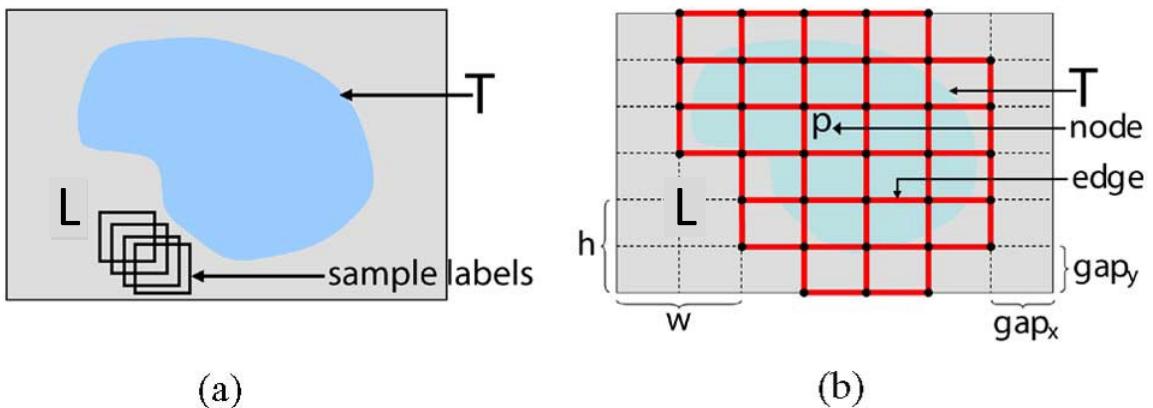


Figure 5.4: Image completion structure of MRF (a) Labels represented by $w \times h$ patches (b) Nodes and edges used during image completion [21].

To assign a patch or label from a reference image R with node in sensed image S and energy of the MRF must be defined so that only consistent patches are copied with each other. To place a patch x_p over a node p , the node potential (or label cost) $V_p(x_p)$ will

decide the copying process consistency, and expressed by the following energy equation:

$$V_p(x_p) = \sum_{dp \in \left[-\frac{w}{2}, \frac{w}{2}\right] \times \left[-\frac{h}{2}, \frac{h}{2}\right]} A(p + dp)(R(p + dp) - R(x_p + dp))^2 \dots \dots \dots \quad (5.5)$$

Where $A(\cdot)$ represents the consistency area. The final matching of matching area of R with an area of S can be expressed as:

$$F(x) = \sum_{p \in U} V_p(x_p) + \sum_{(p,q) \in \varepsilon} V_{pq}(x_p x_q) \dots \dots \dots \dots \dots \dots \dots \quad (5.6)$$

Applying Belief Propagation (BP) algorithm to MRF energy function is complicated by the costly computational complexity because of the large number of labels. To overcome this complexity, the loopy priority-BP optimization scheme in which significantly reduces the label occurrence with a few useful labels being kept for each node [21]. Pearl [20] has introduced BP as an iterative algorithm by solving a finite set of equations based on message passing concepts. BP repeatedly exchanges (propagates) messages between the nodes by sending messages from a node to all its neighboring nodes and accepting messages from these nodes as well, as shown in Fig. 5.5.

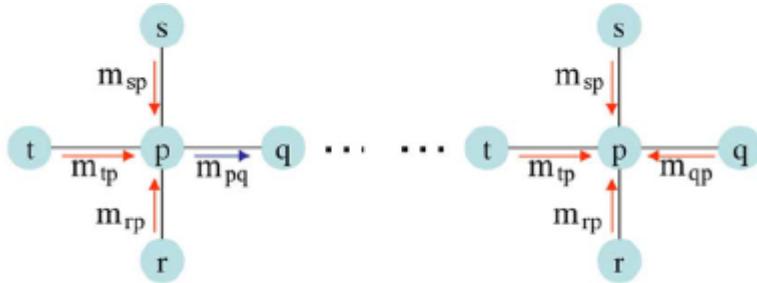


Figure 5.5: Messages propagation between nodes p and q (a) Node p makes use of all its neighboring nodes r,s,t to propagate a message to node q (b) Node p calculates its belief by collecting messages coming from its neighboring nodes [21].

The messages sent from node p to its neighboring node q denoted by $\{mpq(x_p)\}_{x_q \in L}$ represent the opinion of assigning label x_p to node p . This process is repeated until all

nodes agree with each other and the beliefs computed as:

In BP, the total number of labels to be processed is $|L|$ with a complexity of $O(|L|^2)$.

To reduce the total number of labels to be processed, a message scheduling scheme is suggested and to facilitate the circulation of valuable messages and speed up BP's convergence [21]. Message scheduling scheme calculates the priority of each node which is proportional to the available labels for that node as shown in Fig. 5.6.

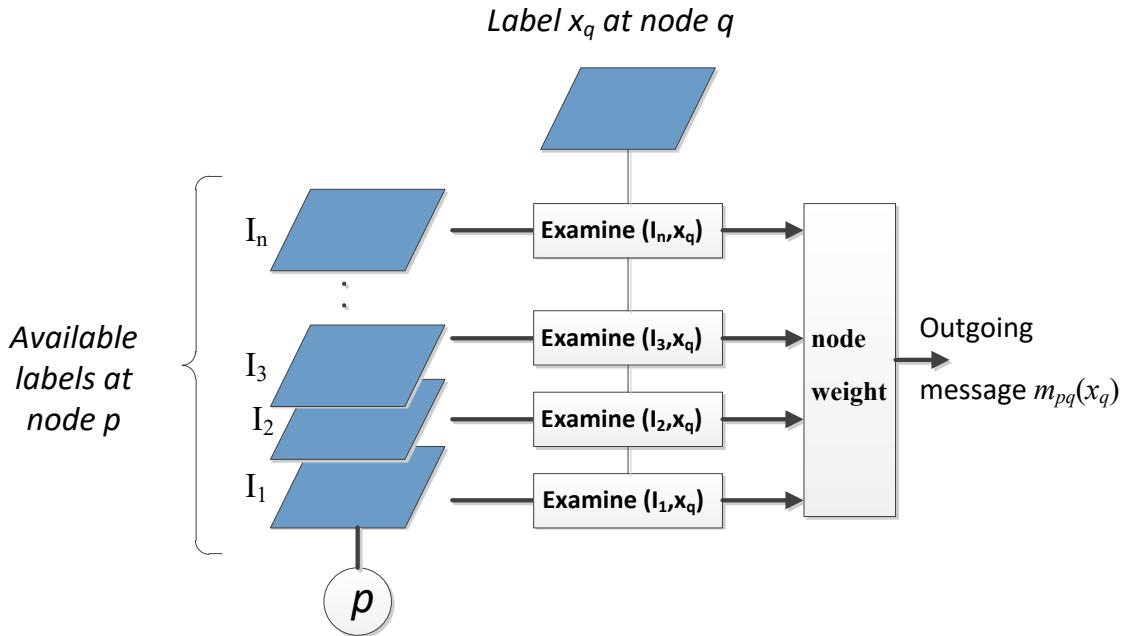


Figure 5.6: Node weight calculation to decide message priority

The belief estimation of a certain node depends on the likely labels of a node by counting the belief $b_p(x_p)$ exceeding a certain threshold. The greater belief results from the more labels, with high probability for that node.

5.1.2 Results and Discussion

The proposed modern to historic image registration process was applied on our

landmark image sets. The pixel displacement field which represents the dense feature correspondence was estimated. The best matches and hence correspondence based on the minimum energy of equation 5.3. Pixel inverse warping was applied on a pixel-by-pixel method in which every pixel was warped independently according to its horizontal and vertical displacements DX and DY [113]. The inverse warping technique has two parts, the warper constructs a continuous representation of the reference image and samples the continuous image with rays sent out from the desired image [113].

The registration results show a high quality registration process and this represents one of the first successful attempts to overcome the technical and environmental difficulties present in image capture process. Fig. 5.7 shows the registration process of two images of the Eiffel tower set, the historic image was taken in 1889 and the digitization parameters are unknown. The modern image was taken in 2010 with modern digital camera with alignment energy of 14919.2 based on equation 4.10.

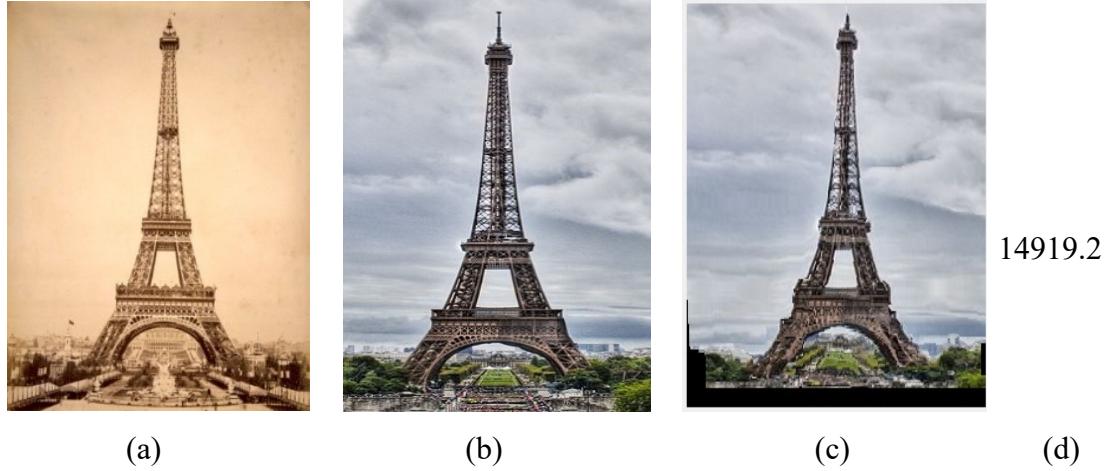


Figure 5.7: Modern to historic image registration (a) Historic image from 1889 (b) Modern image from 2010 (c) Registered image. (d) Alignment energy

The registered images of the Eiffel Tower full set landmark images are shown in Fig. 5.8.



Figure 5.8: Registered images of the Eiffel Tower landmark image set

The registered images for the Coliseum of Rome and Taj-Mahal landmarks are shown in Figs. 5.9 and 5.10.

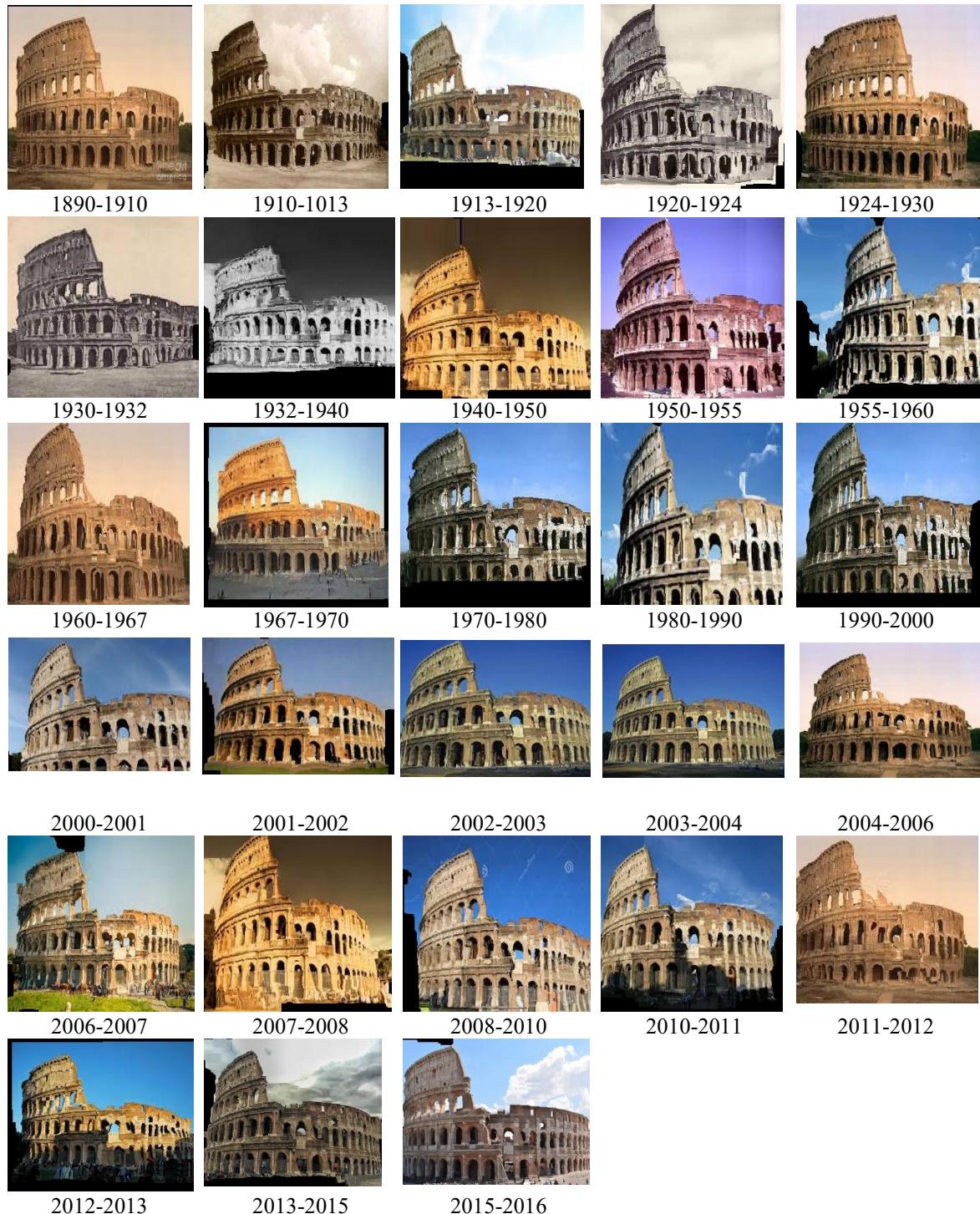


Figure 5.9: Registered images of the Coliseum of Rome landmark image set



Figure 5.10: Registered images of the Taj-Mahal landmark image set

The complete registration image set of the other eight landmarks are shown in

Appendix E.

The image sets of the landmarks we worked on are comprised of images contain a diverse combination of scenes, with different dense of textures, and objects non-regular symmetry natures. This deeply affects the registration quality of the landmark images as will be discussed later.

Image texture is a set of metrics calculated in image processing designated to quantify the perceived texture of an image. It gives information about the spatial arrangement of color or intensities in an image [114]. Texture analysis the image set of the landmarks, region-based image segmentation will be applied to the landmark images to show the texture nature of the different landmark images, and hence, to show the image texture distribution on the automated image registration approach. Textures might be distributed as no, simple, and dense textures through the overall image area. Different texture distributions are shown in Fig. 5.11.

A texture image was created using the entropy filtering, and next we binarize the images to show boundaries between the textures, and exhibit the rough boundary between textures. This will emphasize the diverse texture contents of the different image sets of the landmarks.

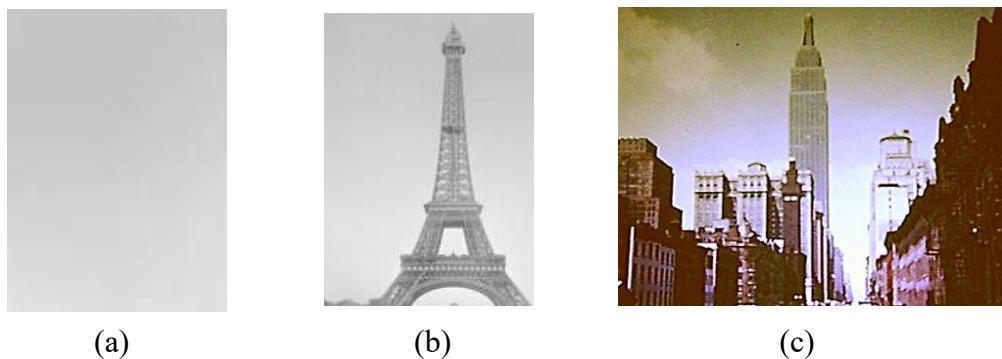


Figure 5.11: Texture Distribution (a) No texture image (b) Simple and easy transition texture image (c) Dense and overlapped texture image

For an image of the Eiffel Tower landmark, the texture analysis is shown in Fig. 5.12, (a) shows the original image, (b) is the entropy filtering texture image, (c) is the binarized texture image, and (d) is the holes and small areas removed. It can be noticed that main object properties of the tower lie in the white area of Fig 5.12(d) with no object overlapping inside this area.

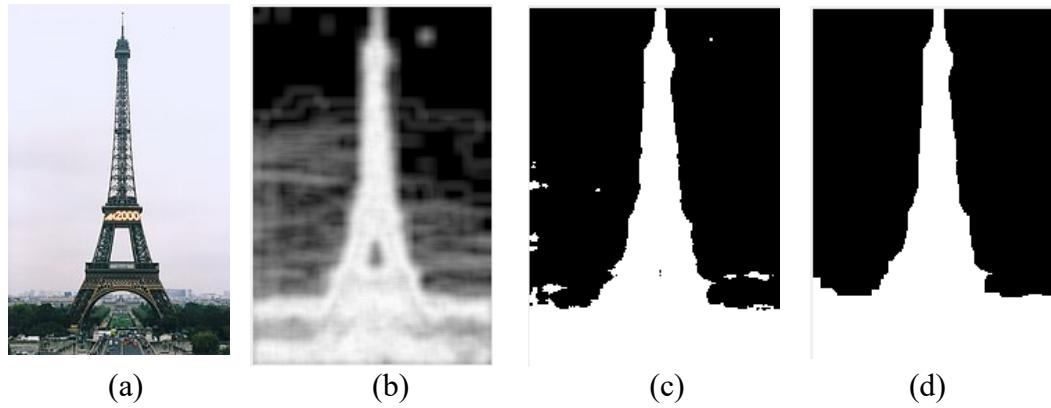


Figure 5.12: Texture segmentation for an Eiffel Tower landmark image

Also, it easily noticed that the black area of the same figure reveals that the tower surrounding area is almost textureless, this facilitates the feature detection, matching, and hence image registration process. More examples of the texture segmentation of landmarks images are shown in Fig. 5.13.

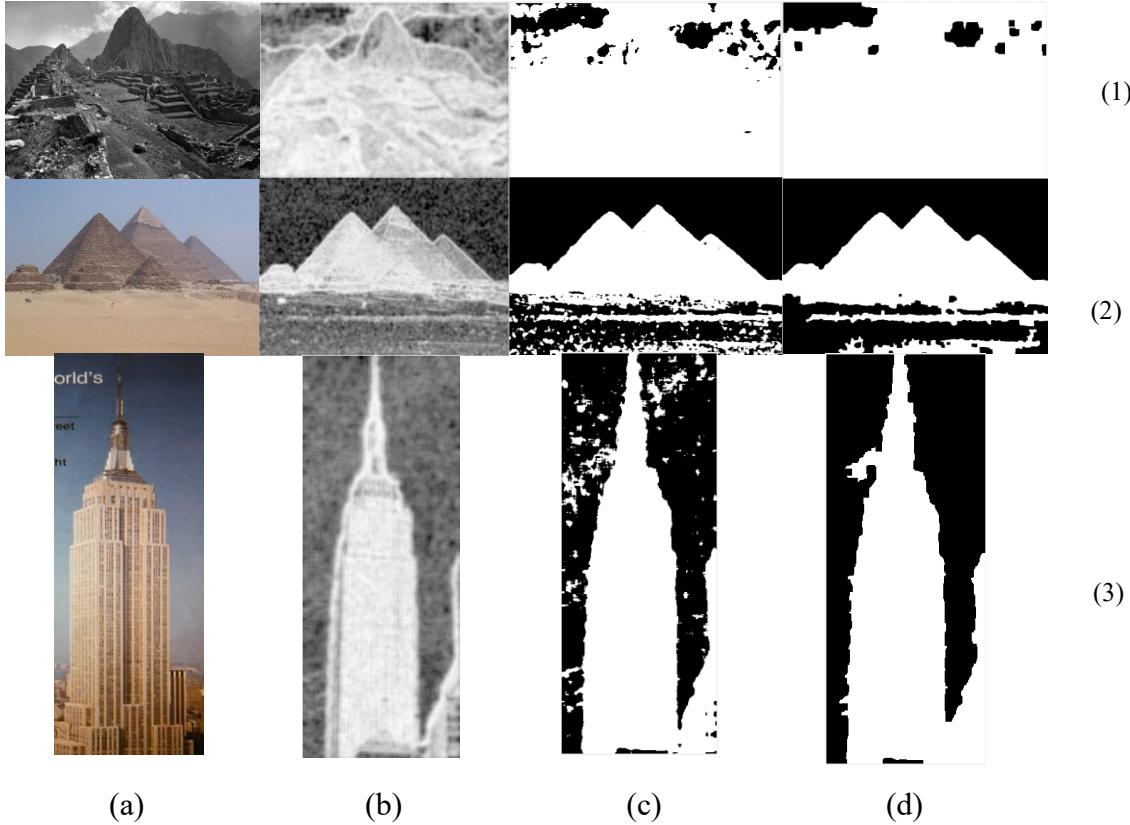


Figure 5.13: Texture segmentation Row (1) Stonehenge landmark image, Row (2) Pyramids of Giza landmark image, Row(3) Empire State Building landmark image with Columns (a) Original image (b) Entropy texture image (c) Binarized texture image and (d) Holes and small areas removing image

Row (1) of Fig. 5.13 shows a full of terrain image in which hills, grass, and grooves are overlapped, and consequently, the texture segmentation failed to show separate, independent textures in column (d). Row (2) shows almost separate, independent, and well segmented textures in column (d). While Row (3) of the Empire State Building shows a well separated textures in column(d) for the first glance, but a careful exam of column(a) image shows a lot of details including edges, corners, and shadows overlapped inside the empire tower.

Our image alignment based on scene alignment by matching salient object such as buildings, windows, bars, domes, and doors with different areas of plain surfaces. This

classified the image registration results into three categories based on the contents and nature of the available landmark views:

- Good registration: Eiffel tower of Fig. 5.8, Coliseum of Rome Fig. 5.9 and Whitehouse of Fig. E.8 (Appendix E). Analysis of the registered image in these figures reveals that the registration process is relatively accurate despite the clear differences either from the capturing environments or the capturing technology. There are clear illumination issues and significant distortion issues between the two images. The tower of the Eiffel Tower landmark edges and curvatures, curves and lines of Coliseum landmark, and edges and windows of Whitehouse landmark are highly aligned in the correspondence process, however there still exists numerous registration errors and visual artefacts.
- Acceptable registration: Taj-Mahal of Fig. 5.10, Dome Rock of Fig. E.1 (Appendix E), Hagia-Sophia of Fig. E.5 (Appendix E), Pyramids of Giza Fig. E.2 (Appendix E), and Stonehenge Fig E.7. In Taj-Mahal, Dome Rock and Haig-Sophia landmarks the dome curves and edges represent a good matching feature, while the other overlapped objects may negatively affect the overall matching and then registration process.
- Poor registration: Machu Picchu Fig. E.3, Empire State Building Fig. E.4, Parliament Hill of Canada Fig. E.6, (All figures in Appendix E). The overlapped objects and the type of the available images of these landmarks negatively affect the registration process. Despite that the computed minimum energy of the pixel displacement is used to warp the images, the registration process failed to show best alignment because the available images views and the resulted SIFT flow

features failed to simulate an image pair as of two consecutive images.

This mainly happened because the displacement mapping in which the actual geometric positions of points over the textures surfaces are displaced, often through the local surface normal; i.e. depending on the value the pixel displacement field evaluates at each point on the surface. Displacement mapping, unlike bump mapping, adds a small-scale detail to every point of the surface, and hence, adjusts the positions of surface elements. This pushes vertices of the base surface along the normal of the base surface and continuing until the polygons generated are close the size of a single pixel [115]. In this approach, the warping of two surfaces (or images) started be rendering the top plane of the sensed image, the finding points on the displaced surface that better match the reference surface or image.

5.1.3 Computation Complexity

In our approach of loopy priority belief propagation for SIFT features of optical flow, a considerable improvement on the computation complexity has been achieved over other approaches for automatic image registration. The computation complexity has been reduced to $O|L_{max}|$ as the L_{max} denotes the maximum number of labels after a significant label account minimization by assigning zero as initial label cost which refers that not all labels have significant beliefs through all the calculation rounds.

To compare our approach to other approaches based on optical flow of the features and the belief propagation techniques, Cheng et.al. [120] divided the optimization problem into number of local problems, in each iteration; each local problem attempts to estimate a possible solution to be optimal and performs the registration process based on

the least squares calculation which results in a more calculations as the number of matched points increases. Zexu et.al. [105] worked to find the canny edges and then to extract feature points based on information fusion method. This will be an exhaustive computation by using all the edges and points with ability chance of missing important features not located on the detected edges.

Liu et.al. [22] based their work on the coarse-to-fine matching scheme for the optical flow of SIFT features to perform image registration in a computation complexity burden of $O|L^2|$. Sharon [121] who used loopy belief propagation in image-based rendering worked on a small number of neighboring states at each node; which is only four states, in an exponential computation calculation. This algorithm works well with simple and natural images but fails on images with complex textures. Table 5.1 summarizes the computation complexity comparison of the related approaches.

Table 5.1: Comparison of the related registration approaches

Comparison Factor	Improved SIFT-Based Using BP [120]	Image Registration Based on Optical Flow and Feature Extraction [105]	SIFT-Flow: Dense Correspondence [22]	Loopy BP in Image-Based Rendering [121]	Our Approach: Priority Loopy BP
Computation Complexity	Computation increased with the number of matched points	$O L^2 $	$O L^2 $	$O L^2 $	$O L_{\max} $
No. of Neighboring Nodes	Max of 5	Restricted to points of edge i.e. single direction	Variable	Max of 4	Variable
Applications	Affected by matching threshold	Worked well with images of strong edges	Face recognition and same scene image alignment	Well with natural and simple images. Fail: Images with complex textures	Good quality for historic to modern image registration

5.2 Image Warping (Standard Techniques)

For the standard feature detection and matching of Section 4.2 in Chapter 4, many of the applied feature matching techniques provide many trusted, but sparse, matches that can be used to derive a planar homography matrix for each successive pair of image to provide a single view opportunity to the user. Image warping was applied using these matrices to get timeline-based images that are transformed to a single viewpoint for each successive pair of images. These images are shown in Fig. 5.14.

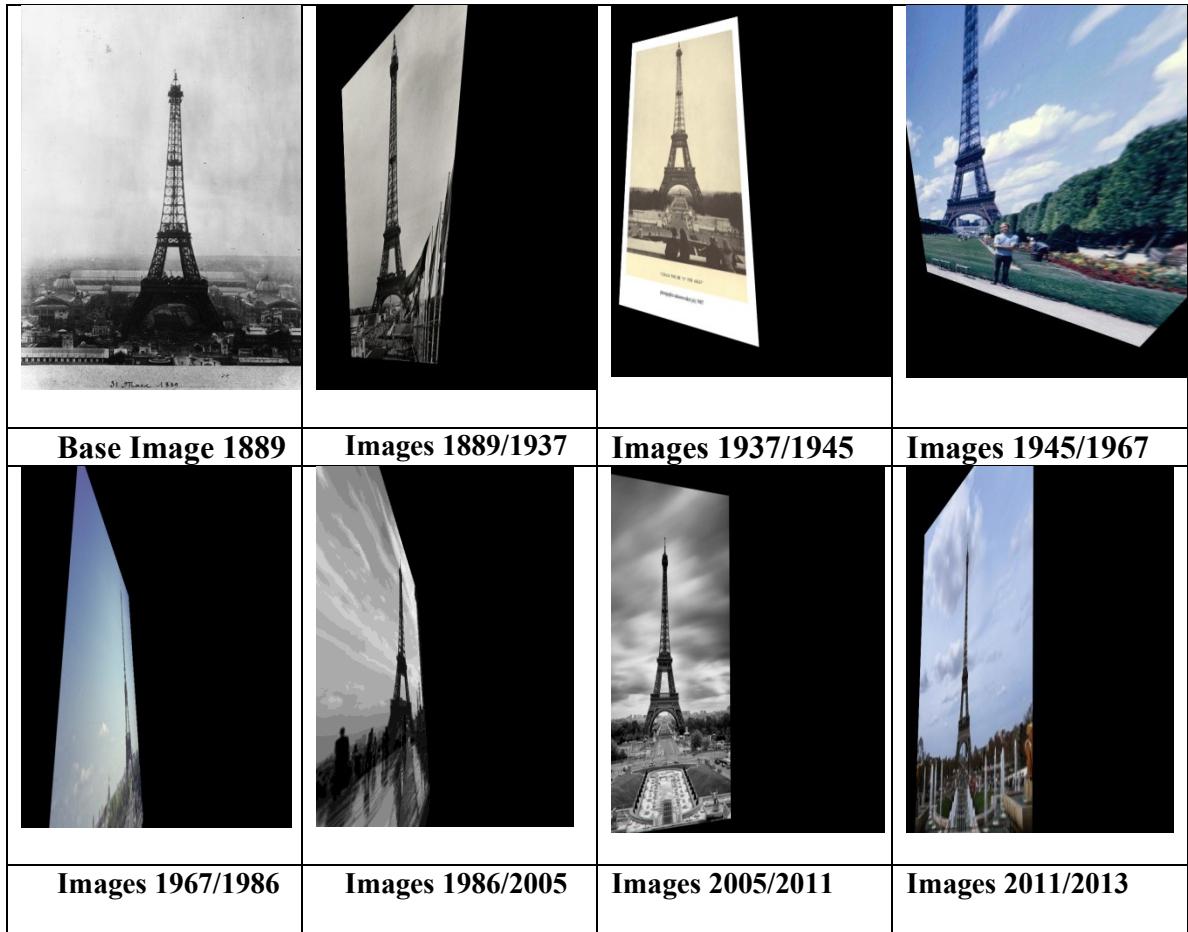


Figure 5.14: Warping images applied on the timeline of Eiffel Tower images from 1889 to 2013

We can note from Fig. 5.14 that the tower not lined up perfectly which means that more work to be done to get better matching results used in formulating the homography matrix and hence to get better image warping results when the images transferred into single viewpoint.

Chapter 6

Timelined Panorama Display

In this chapter, we'll demonstrate the timescape panorama system implementation which integrates all the system modules and subsystems. This implies all the process steps and procedures. Also, the system interface design of the timelined panorama depending on the state-of-art panoramic and comprehensive systems display tools and techniques [117]. Results and samples of the system are demonstrated and discussed. Finally, an evaluation of the current registered image view of the landmarks which shown in an timelined view will be presented. A comparison of this display is done with various panorama types.

6.1 Interface Design

The system functionality depending on the main tasks of Fig. 1.2 can be divided into two main parts: automatic operation and user intervention parts. The automatic image registration process is an automatic operation. Some activities of the timescape panorama building subsystem require the system designer and/or user intervention. A web-based geographic and informatics subsystem was generally built but requires many improvements which will be suggested for a future work. These were however implemented quickly to show the overall proof of concept.

The system software is divided into many modules which utilize the basic system functions. These modules are:

- Web-Based Geographic and Informatics (WBI) module
- Image Subset Selection (ISS) module

- Automatic Image Registration (AIR) module
- Timescapes Panorama Display (TPD) module

6.1.1 Web-Based Geographic and Informatics (WBGI) Module

This is an interactive module accepts the user searching about certain location or an address around the world, the module will search the geographical websites like Google maps, United States Geological Survey (<https://www.usgs.gov/>), or Mapping toolbox of MATLAB to get the exact address and coordinates of the requested location. Then, search a UNESCO landmark database like World Heritage List, to check if the given location is a registered landmark, and finally, to search the photo websites like Flickr or Google Images to collect photos of the designated landmark. Or otherwise, respond that the given location is not a landmark. A screenshot of the WBGI module to request a landmark is shown in Fig. 6.1.



Figure 6.1: Screenshot of a WBGI module

6.1.2 Image Subset Selection (ISS) Module

In the Image Subset Selection (ISS) module, the landmarks image sets are examined by a neural network system to select the most appropriate ones for each landmark that roughly characterize a single view for each landmark, to keep the images most compliant to the given criteria, and to discard the others.

In this module, a specific landmark is selected first to feed its training images into module to be trained in the network. The required training method and specific parameters are selected with the performance function. The images that are highly compatible with the applied criteria and selected parameters are saved in a specified folder. Fig. 6.2 shows screenshots of SSI module.

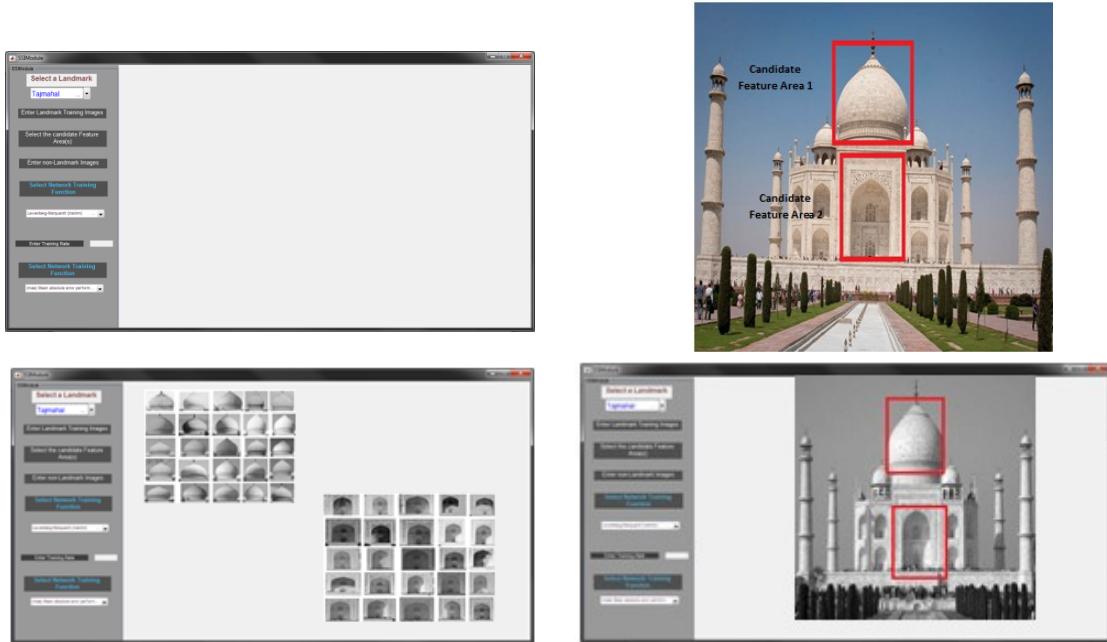


Figure 6.2: Screenshots of an ISS module

6.1.3 Automatic Image Registration (AIR) Module

The main task of this module is the automatic registration of pairwise images of the landmark image sets. The registration may be done chronologically from the oldest historic photo up to the modern image. Within this process, the similarity index and the error factor are calculated to select the reference image for each landmark image set. Consequently, the registration process is performed on each successive image pair up to the most recent image in the set or down to the oldest historic image. Screenshot of the AIR module is illustrated in Fig. 6.3.



Figure 6.3: Screenshot of an AIR module

6.1.4 Timescapes Panorama Display (TPD) Module

In this module, the Timescapes panorama will be displayed after a landmark to be displayed is selected by the user. The system will respond to the user request by showing the landmark reference image accompanied with geographical, historical, and coordinates information as shown in Fig. 6.4 (a). The system enables the user to navigate to the past

of a selected landmark, or to the recent registered images through the downward or upward scrolling respectively of the scroll wheel. The system will navigate to the past or the present showing the original landmark images in the lower right corner in the Fig. 6.4 (b), while showing the registered image of each of a pair of landmark images in the upper image accompanied with geographical, historical, and coordinates information.

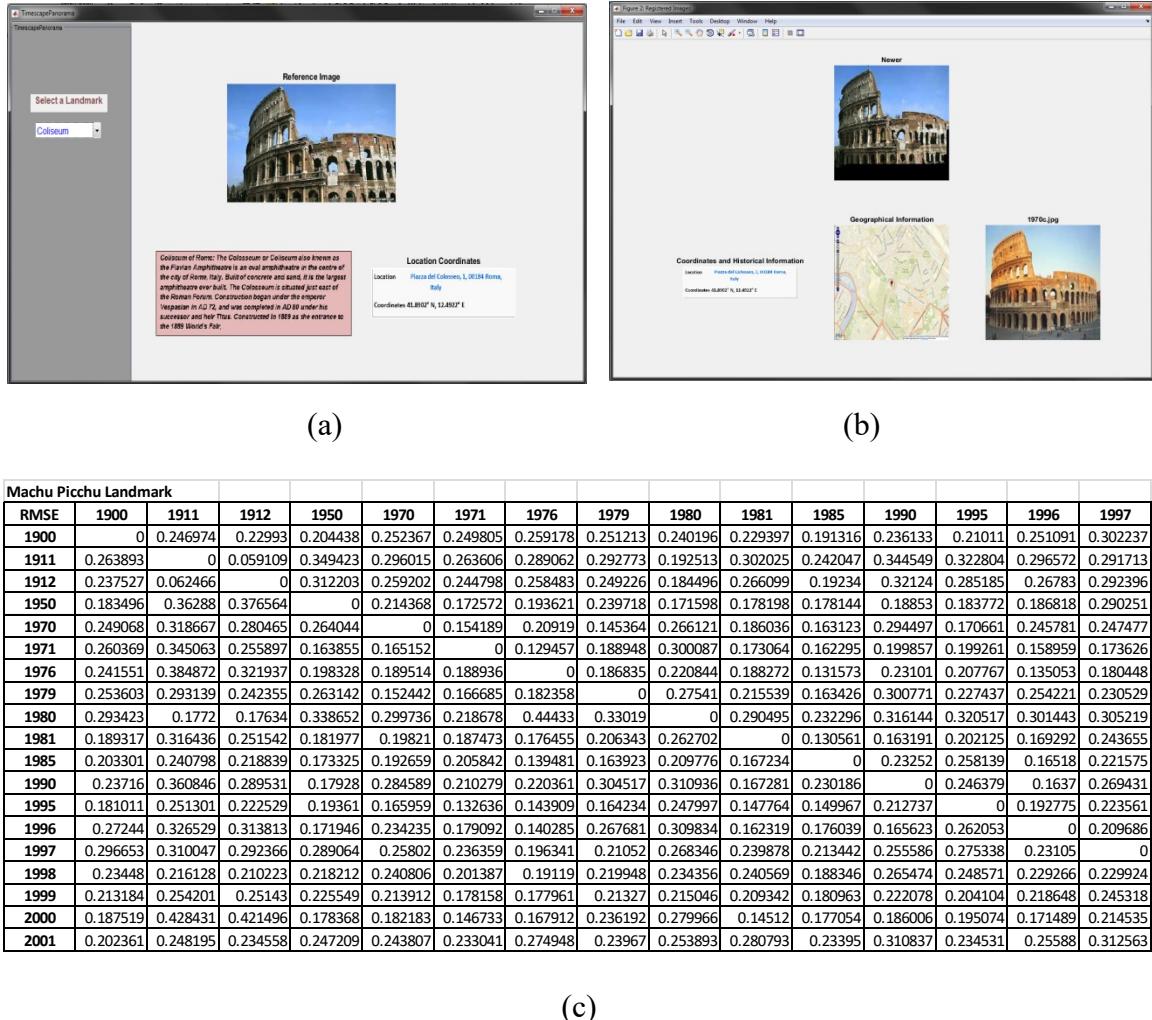


Figure 6.4: Screenshots of a TPD module (a) System response as a landmark selected (b) Timescape panorama display, lower right image is the original image, upper image is the reregistered image of each pair of images (c) Table shows the RMSE computed for each successive pair of images in the landmark image set

The table shown in Fig. 6.4 (c) shows part of a table of bunch of tables produced by TPD module to compute and select the reference image for each landmark image set based on the criterion explained in the next section.

6.2 Panoramic Display

Many techniques and commercial products for browsing sets of photos emphasize the subject of how people desire to organize and display photos or images. Many of these techniques use metadata such as photographer, capturing location, or time as a basis of photo organization [116]. Image browsing using geo-location information has been growing interest to arrange images in an interactive 2D map, divide images into clusters based on time and location, or presenting an interface for retrieving images using vicinity to virtual camera by obtaining the location from GPS system [117].

Our system aims to display a timescape panorama of landmarks around the world based on both spatial-temporal aspects and geographic information. It is initially defining the geographic spot with its coordinates and display the registered images in chronological order based on the reference image of every landmark.

The modern and historic images are put into a timescape presentation by first building a registration table for each image pair showing the precision of the registration process. Three widely used measures are proposed to measure the similarity of each successive pair of registered images include Structural Similarity Index (SSIM) which measures the similarity of two images depending on the degradation of structural information [118], the SSIM measure between two windows x and y each of NxN size is:

$$SSIM = \frac{(2\mu_x \mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad \dots \quad (6.1)$$

Where μ_x and μ_y are the average of x and y respectively.

σ_x^2 and σ_y^2 are the variance of x and y respectively.

c_1 and c_2 are two constants to stabilize the divisions with weak denominator.

The Root Mean Square Error (RMSE) which calculates the differences between the predicted values (reference image) and the observed values (sensed image) and it is widely used as a measure of accuracy [119], the RMSE of two images x and y :

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (x_t - y_t)^2}{n}} \quad \dots \dots \dots \quad (6.2)$$

The third measure is the Histogram Difference measure. Despite the fact that this measure is sometimes applied to measure images similarity, it is not applicable in our work since two different images may yield the same histogram.

The SSIM and RMSE were both applied to our image sets on the registered image pairs in order to judge the similarity of each successive pair of registered images. For each landmark the creation of the set of a timescape images is made up of every two consecutive images (in time) are registered. SSIM also applicable and shows acceptable similarity results and seems to work better on the registered images of overlapped structures or objects.

To determine the best reference image for the timescape, an NxN table of registration errors is generated by registering each image to all the remaining images in the dataset. To select the reference image for timescape, the following three equally weighted selection criterion was applied:

- the least registration error in all rows of lowest RMSE table

- highest similarity in all columns of the table
- the minimum sum of the RMSEs of each column in the table.

The image that has the highest score using the above criterion will be selected as the reference image for the timescape presentation. The timescape continues to show a timeline of registered images of the landmark proceeding downward to old history or upward to modern years depending on the mouse scroll movement.

6.3 Results and Discussion

To examine the performance of the available commercial panorama software, three existing panorama software were applied on images of the landmarks dataset as depicted in Fig. 6.5. The figure shows the poor images resulted from using various types of panorama software on images of landmarks dataset.

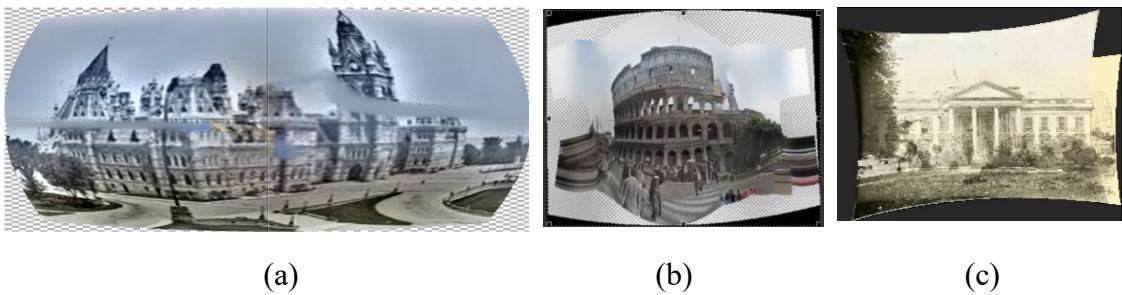


Figure 6.5: Panorama software used on modern and historic images (a) PTGui Panorama Software used with images of Parliament Hill of Canada (b) ArcSoft Panorama maker used with images of Coliseum of Rome (c) Autopano Giga Panorama software on the Whitehouse images

The three commercial panorama software packages failed to produce an acceptable timelined panorama because they are designed to build panoramas based on images collected by a same digital camera with all capturing factors available to the software to accomplish spatial panoramas. Also, this type of software failed to successfully deal with

historic images captured by old analog cameras with missing capturing factors and in different various capturing conditions.

Fig. 6.6 shows the Eiffel Tower landmark data set with the 1990 image selected as the reference image. A timescape panoramic display in chronological order displays the registered images by moving deeper into history or forward towards the present through mouse scrolling. This interface provides historical information (and change over time) about the landmark from the same viewpoint and important kind of contextual historical information.

A detailed user experience explanation accompanied with many screenshots showing the interaction details between the user and the system given in Appendix F. This appendix will show the system capability for the user navigation from the past up to the present of the selected landmark depicted by highly registered images of the landmark images which is a combination of both historic and modern images. This presentation showed the achieved success by this system to address the technical and environmental capturing difficulties between these images.

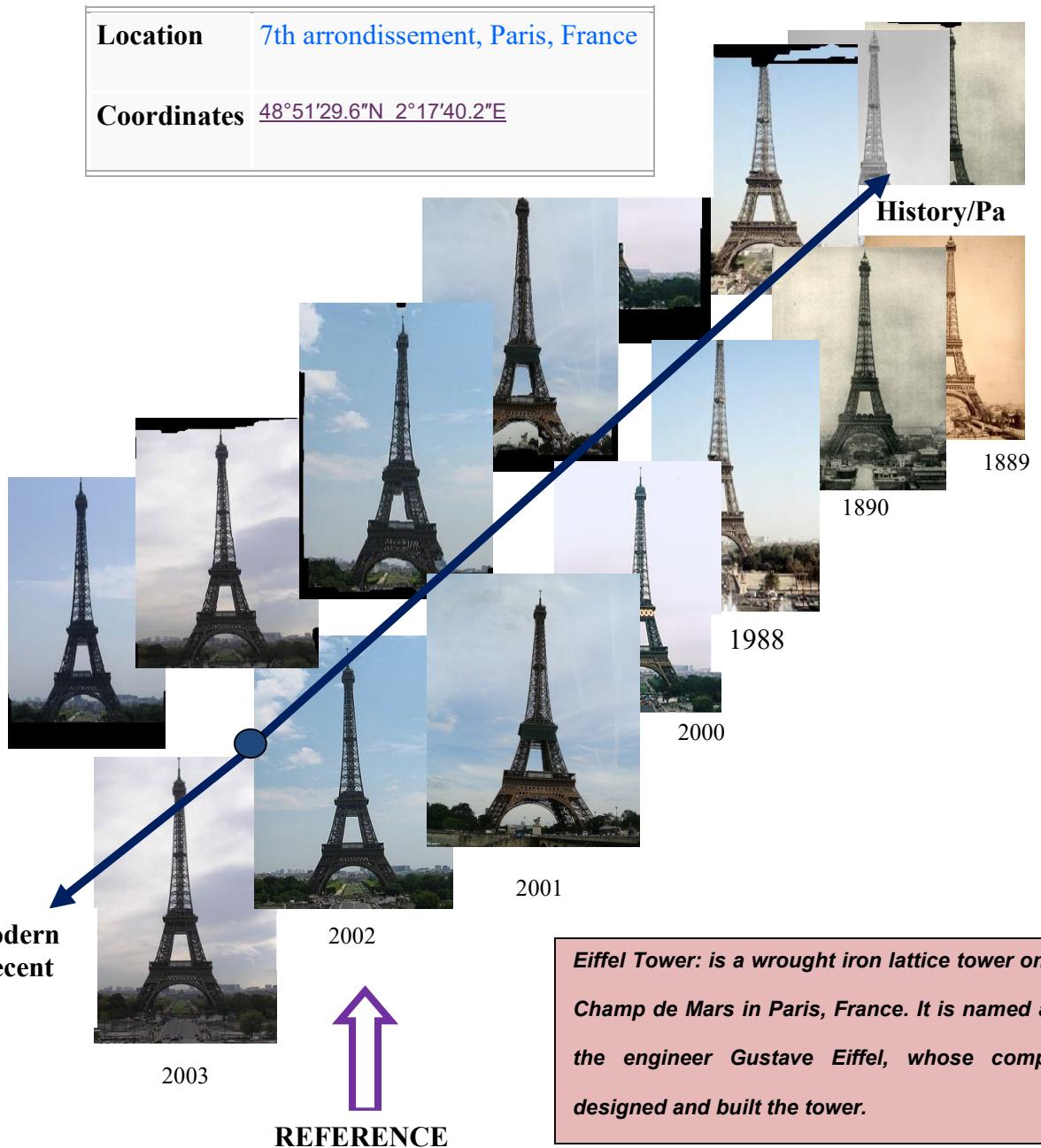


Figure 6.6: Timescape panorama for Eiffel Tower landmark exploring the registered historic and modern images starting with reference image of 2002

As shown in the chart of Fig. 6.7 that the highest similarity percentage of high SSIM record (Registered Images with $\geq 0.65\%$ Similarity) and the highest percentage of low RMSE records (Registered Images with $\leq 0.25\%$ Error) characterize the registration images of the landmarks image sets with views of considerable flat areas with less textures and objects with no overlapping or minimum overlapping.

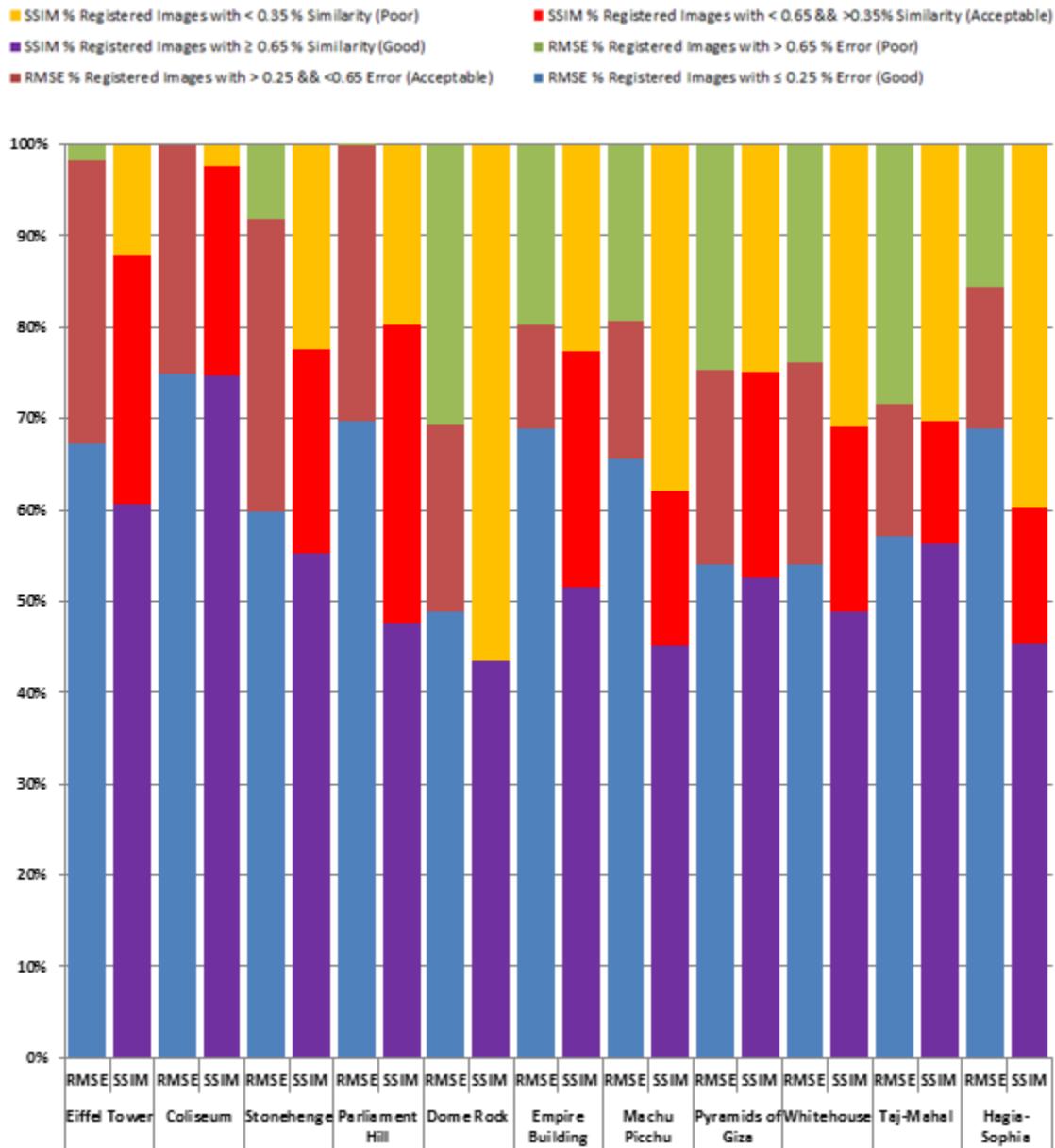


Figure 6.7: Percentage ratio of SSIM and RMSE for the registered landmark images

6.4 Timescape Panorama and Existing Spatial-Temporal Panoramas Evaluation

This work presents a timelined panorama for many landmarks around the world in a chronological way. There are many existing panorama types as classified and declared in Section 2.7 in Chapter 2. The spatial panoramas of 2-D and 3-D were built based on the stitching of images with overlapped views [24].

The temporal panoramas of Section 2.7.2 are classified as time-lapse videos or historic. Time-lapse videos [26] relies on images captured by same camera (single modal) in certain time frame, while historic perspective panorama presents series of events accompanied by multimedia (images mainly) with no interaction or processing between the image set. Rideau timescape [29] presented as navigation through some of available historical data sources. Table 5.2 summarized the evaluation of different panorama types.

Table 6.1: Comparison of panorama types

Evaluation Factor	Panorama					
	Spatial		Temporal		Timelined Panorama	
	2-D	3-D	Time-Lapse Videos	HPP	Rideau Timescapes	Timescape Panorama
Image Capturing	Single Sensor	Single Sensor	Single Sensor	Multi-Sensor	Multi-Sensor	Multi-Sensor
Image Set Period	Very Small Time Period	Very Small Time Period	Long Time Period	Long Time Period	Long Time Period	Very Long Time Period
Display Technique	2D Stitched Images	3D Stitched Images	Sequential Frames in Normal Rate	Separated Images and Events Arranged in Chronological Way	Wrapped Images in Chronological Way	Registered Images in Chronological Way
Camera Parameters Availability	Available	Available	Available	NA	Not Available	Not Available

Chapter 7

Conclusions and Future Work

7.1 Conclusions

This thesis presents a new timelined panorama concept in which it emphasizes the timescapes panoramas by merging the timelined image sequence and the geographic information of the famous landmarks around the world. To achieve this objective, the availability of the digital media for nearly everyone around the world including the historic images of the landmarks provided us with a coherent material for a historic review and display of these landmarks. Meanwhile, processing and presenting high quality display composed of both historic images captured by old obsolete capturing technology with modern images captured by modern digital technology represent a big challenge that our work focused on to address this challenge and to integrate a system for timescapes panorama display.

- The available standard techniques for feature detecting, description, and matching have been studied and applied to the image sets of both historic and modern landmarks images. Their performance has been studied and evaluated [104]. The matching was done using the standard techniques like SIFT, SURF and ORB with vector-based or binary-based descriptors are shown to be unsuitable to resolve this problem effectively [104].

In an attempt to achieve an effective matching process on modern and historic image sets, a novel hybrid method from these standard techniques have been tested which led to a novel application of the ORB and SURF techniques, in a

technique designated as ORB/SURF. Although this hybrid was found to outperform the standard techniques, it was still not sufficient to obtain an optimal matching for image registration, and more effort must be done to deal with the problem of historic photos [16].

- To select the most appropriate images which serve to achieve the thesis goal, and to classify the landmarks image sets, a mechanism for the selection of image subset of modern and historic images out of a larger landmark image set collected from the internet is designed. This selection depends on the extraction of dominant features using Gabor filtering and some manual assistance in selecting these features. These features are selected carefully from a preliminary image set and fed into a neural network as training set [98]. We show that Gabor filtering for such a task is a promising methodology when combined with a Neural Network for categorization. Dominant features from the images of many landmarks were selected as the training image set to be fed to the neural network. In the categorization stage, the raw images were applied as input to the neural network. The categorization results have shown the proposed method to be a promising first step in creating a good historical to modern timeline image sets [31].
- Registering both historic and modern photos in an accurate and precise manner allows an opportunity to support the cultural heritage preservations of such landmarks. This presents a timescape display through temporally registered images of landmarks in a single attractive temporal and spatial panoramic view [76].

This work, to our knowledge, is the first automated system which offers an opportunity for the historians and researchers to merge the historic view with the geographic information. A high quality image registration was performed using an optimized optical flow of SIFT features which resulted in better matching and registration that overcomes the difficulties of the data set which includes the traditional registration complexities such as illumination and geometric issues, but also the technological differences in the capture processes. This work contributes in the efforts of maintaining the cultural heritage by successfully generating timescape presentations of the media of historic landmarks and places from around the world [76].

- Providing a panoramic view of famous landmarks around the world offers artistic and historic value for historians, tourists, and researchers. This was integrated into a comprehensive work which presents a new concept of temporal panorama through a timeline display of aligned historic and modern images for many famous landmarks. Processing of historic images captured using older analog technology under various different capturing conditions represents a big challenge when they have to be used with modern digital images. Successful processing of historic images to prepare them for next steps of temporal panorama creation represents an active contribution in cultural heritage preservation through the fulfillment of one of UNESCO goals in the preservation and displaying famous worldwide landmarks [26].
- The concept studied and analyzed by this thesis which is timelined panorama has been realized as a novel application in the serving in the heritage cultural

preservation and introduces a timelined display for the researchers which might be accommodated in various different topics in computer vision, medical applications, and aerial imaging.

7.2 Future Work

- This system can be augmented by adding more capabilities to deal with almost all landmarks to construct an integrated panorama for most of historical, cultural, and other landmarks around the world. The Web-based Geographic and Informatics subsystem of our thesis structure shown in Fig. 7.1 can serve a core for the expansion process. Some of the required capabilities currently exist in our system, while others like comprehensive worldwide maps, an updated global landmark database, and others must be added to integrate the system.
- This system capability of automatic image registration of multimodal images might be applied in medical, aerial, and geographical application images captured in long timelined periods or captured by different capturing and recording devices. In medical applications, a timelined series of images might be registered, as an example different MRI images can be registered for diagnosis and treatment analysis, or a multi-modal images of different types like MRI, CT scan, and x-ray images can be registered.

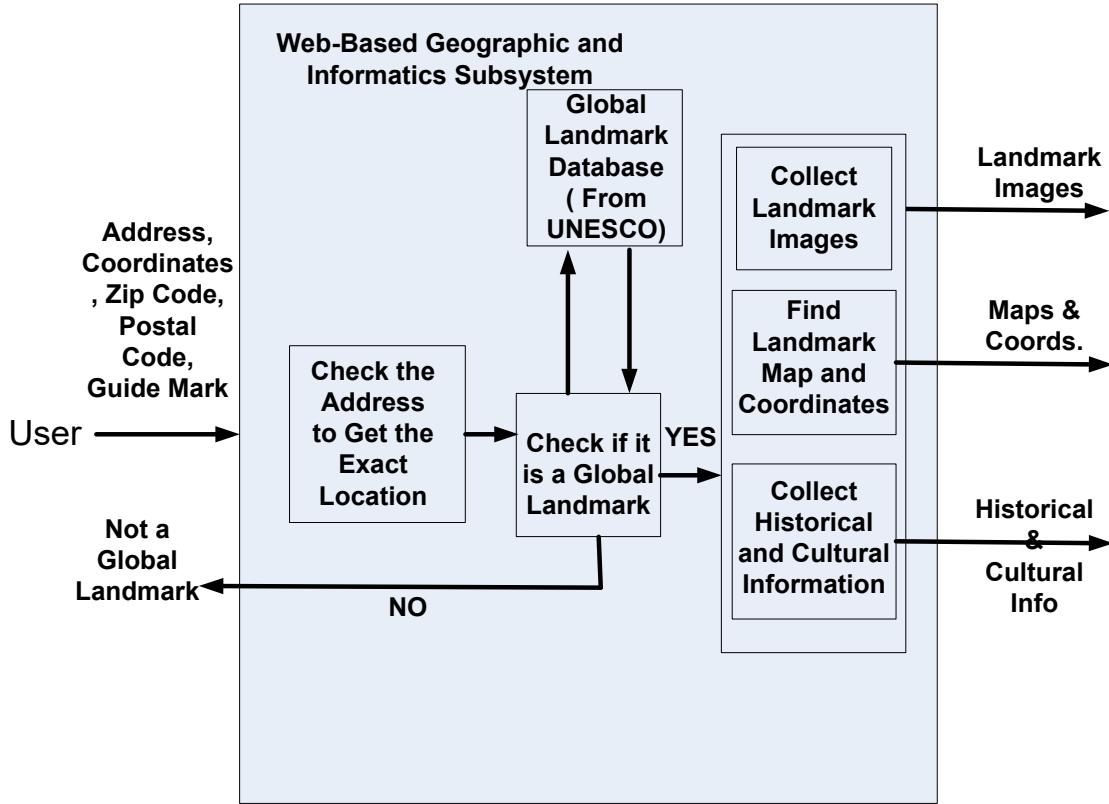


Figure 7.1: The Web-Based Geographic and Informatics Subsystem can be an expansion core for a general landmark timescape panorama

- For those landmarks of a diverse collection of historic and modern images captured from almost 3D directions, this display tool can be evolved to construct a 3D timescapes panorama rather than a timelined panorama display based single view.
- More work has to be done to address problems of overlapped object images and those image captured from far distances, like Machu Picchu, Parliament Hill of Canada, and Empire State building landmarks.
- To convert the system into totally automated one, the system parts which are now manually controlled like the image subset selection can be automated through the automation of the landmark dominant feature selection. These features are used in

the image data sets classification process. The automation process may be realized depending on the similar work in this field like Materka et. al. [112] in which it relies on the feature-selection techniques based on the value of Fisher Coefficient F and combined probability of classification error and average correlation between features to develop automatic techniques for the selection of image features for texture analysis of magnetic resonance test objects (phantoms). Puig et al. [35] worked on pixel-based texture classifiers and segmenters are typically based on the combination of texture feature extraction methods that belong to a single family (e.g., Gabor filters).

Appendices

Appendix A: Dense SIFT descriptors for landmark images

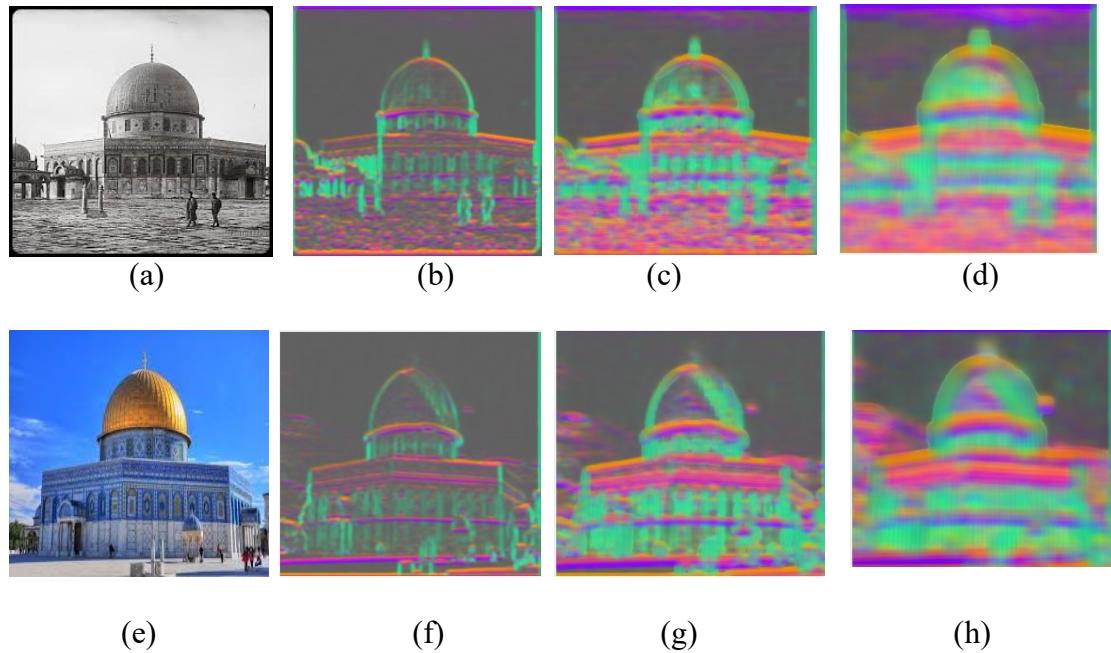


Figure A.1: Dense SIFT descriptor for the Dome Rock landmark images of (a) 1870 and (e) 1980. The dense SIFT descriptors for window size 8 surrounding the pixel shown in (b) and (f), window size 16 shown in (c) and (g), window size 32 shown in (d) and (h).

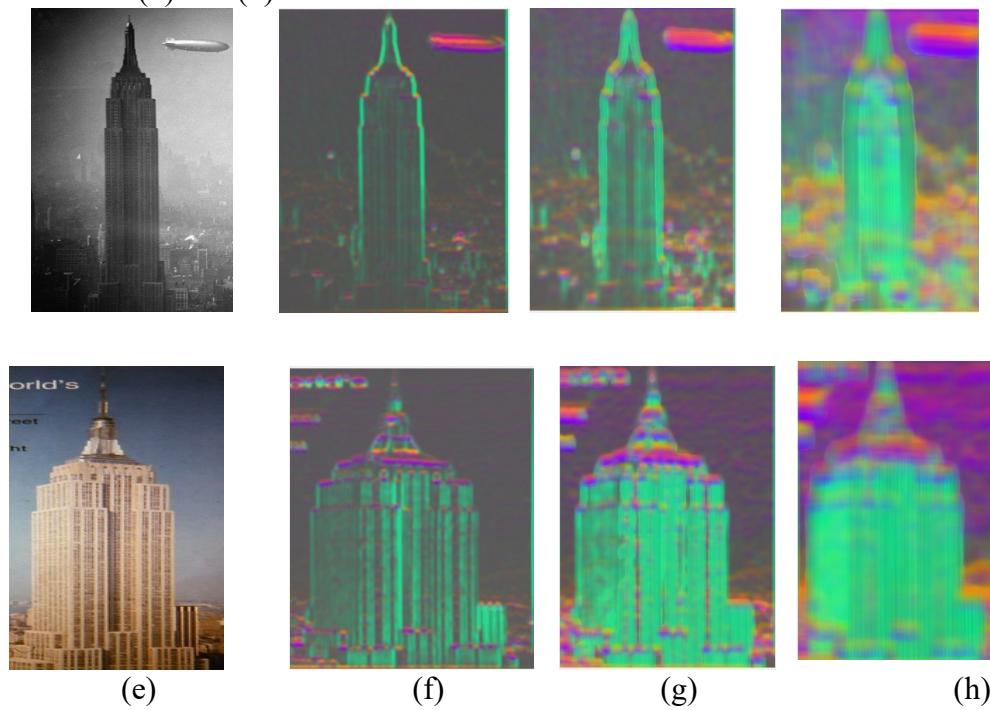


Figure A.2: Dense SIFT descriptor for the Empire Building landmark images of (a) 1936 and (e) 1984. The dense SIFT descriptors for window size 8 surrounding the pixel shown in (b) and (f), window size 16 shown in (c) and (g), window size 32 shown in (d) and (h).

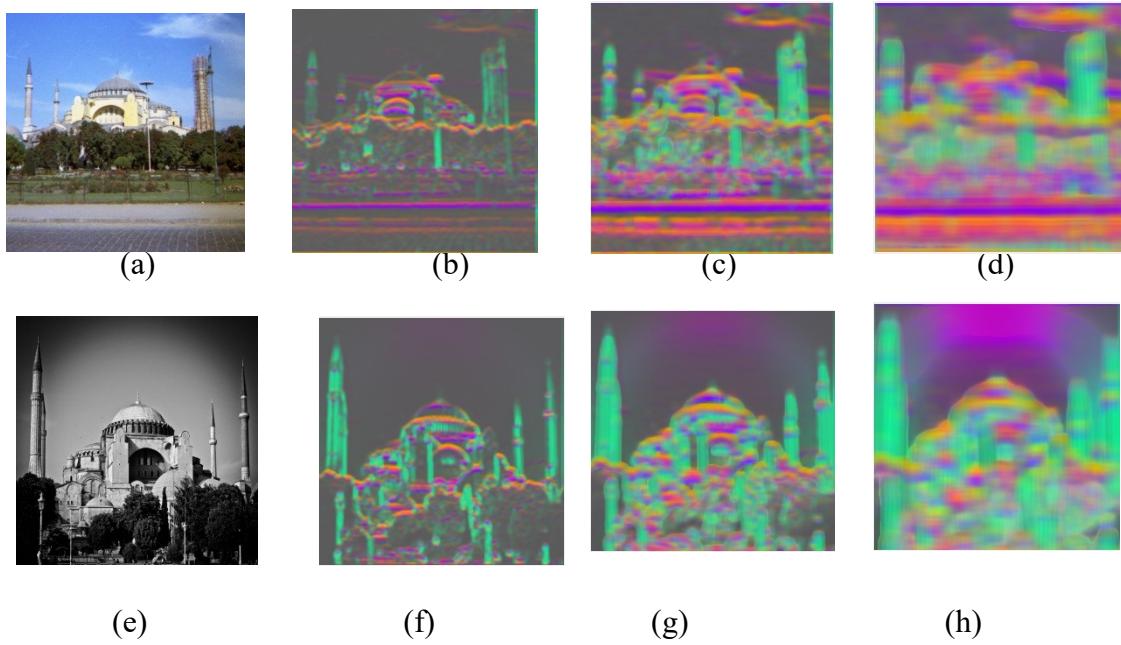


Figure A.3: Dense SIFT descriptor for the Hagia-Sophia landmark images of (a) 1963 and (e) 2011. The dense SIFT descriptors for window size 8 surrounding the pixel shown in (b) and (f), window size 16 shown in (c) and (g), window size 32 shown in (d) and (h).

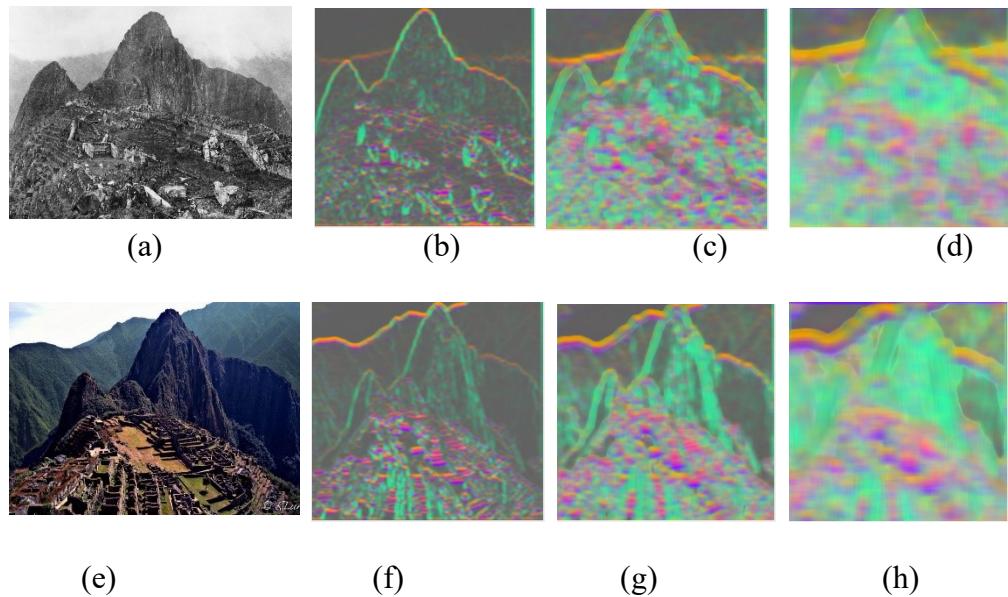
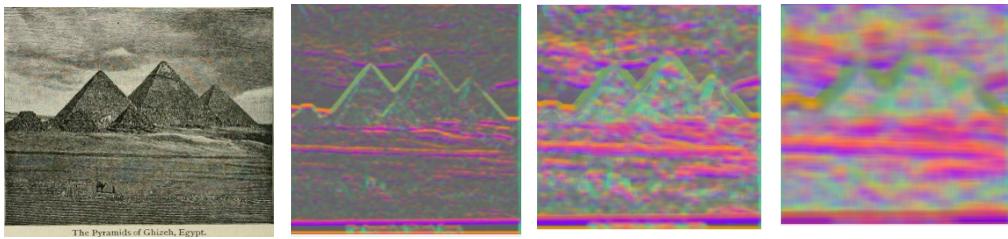


Figure A.4: Dense SIFT descriptor for the Mach Picchu landmark images of (a) 1912 and (e) 1985. The dense SIFT descriptors for window size 8 surrounding the pixel shown in (b) and (f), window size 16 shown in (c) and (g), window size 32 shown in (d) and (h).

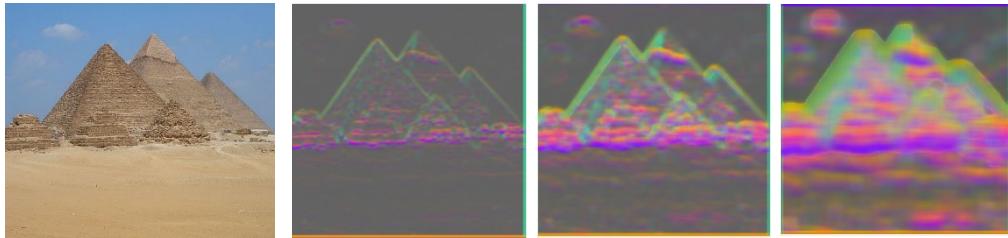


(a)

(b)

(c)

(d)



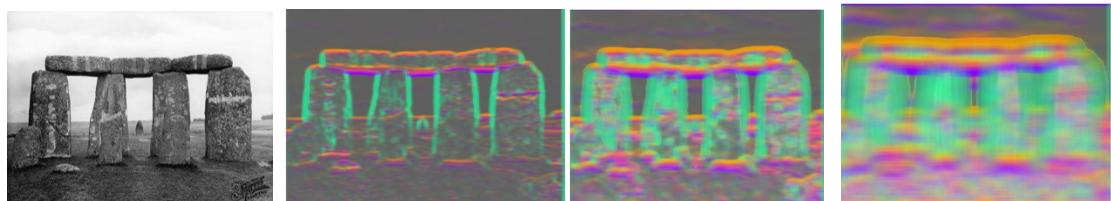
(e)

(f)

(g)

(h)

Figure A.5: Dense SIFT descriptor for the Pyramids of Giza landmark images of (a) 1897 and (e) 2006. The dense SIFT descriptors for window size 8 surrounding the pixel shown in (b) and (f), window size 16 shown in (c) and (g), window size 32 shown in (d) and (h).

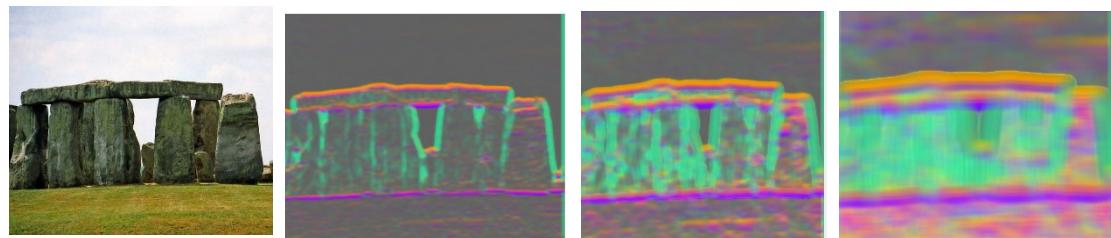


(a)

(b)

(c)

(d)



(e)

(f)

(g)

(h)

Figure A.6: Dense SIFT descriptor for the Stonehenge landmark images of (a) 1928 and (e) 1994. The dense SIFT descriptors for window size 8 surrounding the pixel shown in (b) and (f), window size 16 shown in (c) and (g), window size 32 shown in (d) and (h).

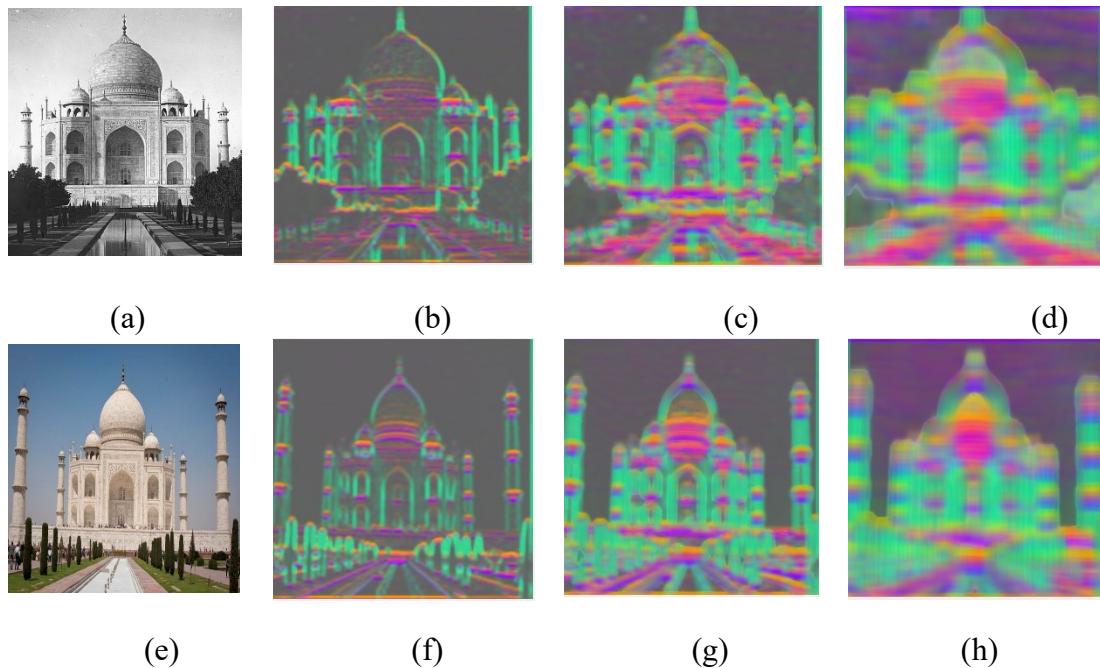


Figure A.7: Dense SIFT descriptor for the Taj Mahal landmark images of (a) 1910 and (e) 2011. The dense SIFT descriptors for window size 8 surrounding the pixel shown in (b) and (f), window size 16 shown in (c) and (g), window size 32 shown in (d) and (h).

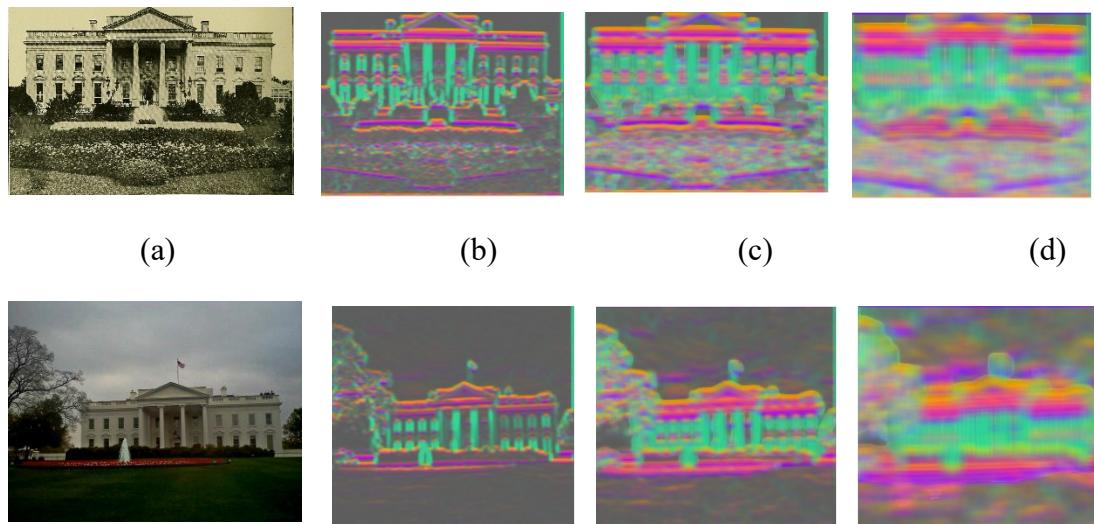


Figure A.8: Dense SIFT descriptor for the Whitehouse landmark images of (a) 1905 and (e) 2016. The dense SIFT descriptors for window size 8 surrounding the pixel shown in (b) and (f), window size 16 shown in (c) and (g), window size 32 shown in (d) and (h).

Appendix B: Landmark matching tables for HARRIS/HARRIS feature matching technique

Table B.1: Pyramids of Giza landmark matching using HARRIS/HARRIS matcher

HARRIS/HARRIS	185 8	188 0	194 0	196 9	197 8	199 2	199 6	200 6	201 1	unknow n
1858	1	0	0	0	0	0	0	0	0	0
1880	0	1	0	0	0	0	0	0	0	0
1940	0	0	1	0	0	0	0	0	0	0
1969	0	0	0	1	0	0	1	0	0	0
1978	0	0	0	0	1	0	0	0	0	0
1992	0	0	0	0	0	1	0	0	0	0
1996	0	0	1	0	0	0	1	0	0	0
2006	0	0	0	0	0	0	0	1	0	0
2011	0	0	0	0	0	0	0	0	1	0
unknown	0	0	0	0	0	0	0	0	0	1

Table B.2: Coliseum of Rome landmark matching using HARRIS/HARRIS matcher

HARRIS/HARRIS	185 5	189 0	189 6	197 6	200 5	200 7	2007 b	201 0	201 1	unknow n
1855	1	0	1	0	0	0	0	0	0	0
1890	0	1	0	0	0	0	0	0	0	0
1896	0	0	1	0	0	0	0	0	0	0
1976	0	0	0	1	0	0	0	0	0	0
2005	0	0	0	0	1	0	0	0	0	0
2007	1	0	0	0	0	1	0	1	0	1
2007b	0	0	0	0	0	0	1	0	0	0
2010	1	0	0	0	0	0	0	1	0	0
2011	0	0	0	0	0	0	0	1	1	0
unknown	1	0	0	0	1	0	0	0	0	1

Table B.3: Dome Rock landmark matching using HARRIS/HARRIS matcher

HARRIS/HARRIS	1900	1910	1915	1959	1966	1989	2009	2010	2012	2012b
1900	1	0	0	0	0	0	0	0	0	0
1910	0	1	0	0	0	0	0	0	0	0
1915	0	0	1	0	0	0	0	0	0	0
1959	0	0	0	1	0	0	0	0	0	0
1966	0	0	0	0	1	0	0	0	0	0
1989	0	0	0	0	0	1	0	0	0	0
2009	0	0	0	0	0	0	1	0	0	0
2010	0	0	0	0	0	0	0	1	0	0
2012	0	0	0	0	0	0	0	0	1	0
2012b	0	0	0	0	0	0	0	0	0	1

Table B.4: Empire State Building landmark matching using HARRIS/HARRIS matcher

HARRIS/HARRIS	1936	1940	1942	1955	1957	1959	1996	2007	2007b	2011
1936	1	0	0	0	0	0	0	1	0	0
1940	0	1	0	0	0	0	0	0	1	0
1942	0	0	1	0	0	0	0	0	0	0
1955	0	0	0	1	0	1	0	0	0	0
1957	0	0	0	0	1	0	0	0	0	0
1959	1	0	0	1	0	1	0	1	0	0
1996	0	0	0	0	0	0	1	0	0	0
2007	1	0	0	1	0	1	1	1	0	0
2007b	0	1	0	0	0	0	0	0	1	0
2011	0	0	0	0	0	1	0	0	0	1

Table B.5: Hagia-Sophia landmark matching using HARRIS/HARRIS matcher

HARRIS/HARRIS	1890	1900	1910	1977	1994	1996	2004	2008	2010	2011
1890	1	0	0	0	0	0	0	0	0	0
1900	0	1	0	0	0	0	0	0	0	0
1910	0	0	1	0	0	0	0	0	0	0
1977	0	0	0	1	0	0	0	0	0	0
1994	0	0	0	0	1	0	0	0	0	0
1996	0	0	0	0	0	1	0	0	0	0
2004	0	0	0	0	0	0	1	0	0	0
2008	0	0	0	0	0	0	0	1	0	0
2010	0	0	0	0	0	0	0	0	1	0
2011	0	0	0	0	0	1	0	0	0	1

Table B.6: Machu-Picchu landmark matching using HARRIS/HARRIS matcher

HARRIS/HARRI S	190	191	198	198	199	199	200	200	200	201
	0	1	0	5	6	9	5	7	9	1
1900	1	0	0	0	0	0	0	0	0	0
1911	0	1	0	0	0	0	0	0	0	0
1980	0	0	1	1	0	0	0	0	0	0
1985	0	0	0	1	0	0	0	0	1	0
1996	0	0	0	0	1	0	0	0	0	0
1999	0	0	0	0	0	1	0	0	0	0
2005	0	0	0	1	0	0	1	0	1	0
2007	0	0	0	0	0	0	0	1	0	1
2009	0	0	0	0	0	0	0	0	1	0
2011	0	0	0	0	0	0	0	1	0	1

Table B.7: Stonehenge landmark matching using HARRIS/HARRIS matcher

HARRIS/HARRIS	188	190	192	195	1955	200	200	200	201	unknow
	7	8	8	5	b	7	8	9	1	n
1887	1	0	0	0	0	0	0	0	0	0
1908	0	1	0	0	0	0	0	0	0	0
1928	0	0	1	1	0	0	0	0	0	0
1955	0	0	1	1	0	0	0	0	0	0
1955b	0	0	0	0	1	0	0	0	0	1
2007	0	0	0	0	0	1	0	0	0	0
2008	0	0	0	0	0	0	1	0	0	0
2009	0	0	0	0	0	0	0	1	0	0
2011	0	0	0	0	0	0	0	0	1	0
unknown	0	0	0	0	1	0	0	0	0	1

Table B.8: Taj-Mahal landmark matching using HARRIS/HARRIS matcher

HARRIS/HARRIS	1874	1890	1910	1910b	1943	1966	1986	1993	2006	2011
1874	1	0	0	0	0	0	0	0	0	0
1890	0	1	0	0	0	0	0	0	0	0
1910	0	0	1	0	0	0	0	0	0	0
1910b	0	0	0	1	0	0	0	0	0	0
1943	1	0	0	0	1	0	0	0	0	0
1966	0	0	0	0	0	1	0	0	0	0
1986	0	0	0	0	0	0	1	1	0	0
1993	0	0	1	0	1	0	1	1	0	0
2006	0	0	0	0	0	0	0	0	1	0
2011	0	0	0	0	0	0	0	0	0	1

Table B.9: White House of Washington landmark matching using HARRIS/HARRIS matcher

HARRIS/HARRIS	1863	1945	1948	1965	1966	1993	2009	2009b	2012	2012b
1863	1	0	0	0	0	0	0	0	0	0
1945	0	1	0	0	1	0	0	0	0	0
1948	0	0	1	0	0	0	0	0	0	0
1965	0	0	0	1	0	0	0	0	0	0
1966	0	0	0	0	1	1	0	0	0	0
1993	0	0	0	0	0	1	0	0	0	0
2009	0	0	0	0	0	0	1	0	0	0
2009b	0	0	0	0	0	1	0	1	0	0
2012	0	0	0	0	0	0	0	0	1	0
2012b	0	0	0	0	0	0	0	0	0	1

Appendix C: Landmark matching tables for SIFT/SIFT feature matching technique

Table C.1: Pyramids of Giza landmark matching using SIFT/SIFT matcher

SIFT/SIFT	1858	1880	1940	1969	1978	1992	1996	2006	2011	unknown
1858	1	0	0	0	0	0	0	0	0	0
1880	0	1	0	0	0	0	1	0	0	0
1940	0	1	1	0	0	0	1	0	0	0
1969	0	0	1	1	1	0	0	0	0	0
1978	0	0	0	0	1	0	0	0	0	0
1992	0	0	0	0	0	1	0	0	0	0
1996	0	0	1	0	0	0	1	0	0	0
2006	0	0	0	0	0	0	0	1	0	0
2011	0	0	0	0	0	0	0	0	1	0
unknown	0	0	0	0	0	0	0	0	0	1

Table C.2: Coliseum of Rome landmark matching using SIFT/SIFT matcher

SIFT/SIFT	T	1855	1890	1896	1976	2005	2007	2007b	2010	2011	unknown
1855	1	0	1	0	1	1	0	1	0	1	1
1890	1	1	1	0	1	0	0	0	0	0	1
1896	1	0	1	0	0	1	0	0	0	0	0
1976	1	1	0	1	1	1	0	1	0	0	0
2005	1	1	0	1	1	1	0	1	1	1	1
2007	0	1	0	0	0	1	0	0	0	1	1
2007b	0	0	0	0	1	1	1	1	0	0	1
2010	0	0	0	0	1	1	0	1	1	1	1
2011	0	0	0	1	1	1	0	1	1	1	1
unknown	0	0	1	0	1	1	0	1	1	1	1

Table C.3: Dome Rock landmark matching using SIFT/SIFT matcher

SIFT/SIFT	1900	1910	1915	1959	1966	1989	2009	2010	2012	2012b
1900	1	0	1	1	0	0	0	0	0	0
1910	0	1	0	0	0	0	0	0	0	0
1915	1	0	1	1	1	0	0	0	0	0
1959	0	0	1	1	0	0	0	0	0	0
1966	0	0	1	0	1	1	0	0	0	0
1989	0	0	0	0	0	1	1	1	0	1
2009	0	0	0	0	0	0	1	0	0	1
2010	0	0	0	0	0	1	0	1	0	1
2012	0	0	0	0	0	0	0	0	1	0
2012b	0	0	0	0	0	1	1	1	0	1

Table C.4: Empire State Building landmark matching using SIFT/SIFT matcher

SIFT/SIFT	1936	1940	1942	1955	1957	1959	1996	2007	2007b	2011
1936	1	0	0	0	0	0	0	0	0	0
1940	0	1	1	0	0	1	0	0	0	0
1942	0	1	1	0	0	0	0	0	0	1
1955	1	0	0	1	0	0	0	0	0	0
1957	0	0	0	1	1	1	0	0	0	0
1959	0	1	0	0	0	1	1	0	0	1
1996	0	0	1	1	0	0	1	1	0	1
2007	0	0	0	0	0	0	1	1	0	1
2007b	0	0	0	0	0	1	0	0	1	1
2011	0	0	0	0	0	1	0	1	0	1

Table C.5: Hagia-Sophia landmark matching using SIFT/SIFT matcher

SIFT/SIFT	1890	1900	1910	1977	1994	1996	2004	2008	2010	2011
1890	1	0	1	0	0	0	0	0	0	0
1900	1	1	0	0	0	0	0	0	0	0
1910	1	1	1	0	0	0	0	0	0	0
1977	0	0	0	1	0	0	0	1	0	0
1994	0	0	0	0	1	1	1	1	0	0
1996	0	0	0	0	1	1	0	0	0	1
2004	0	0	0	0	1	0	1	0	0	0
2008	0	0	0	0	0	0	0	1	0	0
2010	0	0	0	0	0	0	0	0	1	0
2011	0	0	0	0	0	1	0	0	1	1

Table C.6: Machu-Picchu landmark matching using SIFT/SIFT matcher

SIFT/SIFT	1900	1911	1980	1985	1996	1999	2005	2007	2009	2011
1900	1	0	0	0	0	0	0	0	0	0
1911	0	1	1	0	0	0	1	0	1	0
1980	0	0	1	0	0	0	0	0	0	0
1985	0	0	0	1	1	0	0	0	0	0
1996	0	0	0	1	1	0	0	0	1	0
1999	0	0	0	0	0	1	0	0	0	0
2005	0	0	0	0	0	0	1	0	1	0
2007	0	1	0	0	0	0	0	1	0	1
2009	0	1	1	0	1	1	1	1	1	1
2011	0	0	0	0	0	0	0	1	0	1

Table C.7: Stonehenge landmark matching using SIFT/SIFT matcher

SIFT/SIF T	1887	1908	1928	1955	1955b	2007	2008	2009	2011	unknown
1887	1	1	0	0	0	0	0	0	0	0
1908	1	1	1	0	1	0	0	0	0	0
1928	0	1	1	1	0	1	0	1	0	0
1955	0	0	1	1	0	0	0	1	0	0
1955b	0	0	0	0	1	1	0	0	0	1
2007	0	0	0	0	1	1	0	0	0	0
2008	0	0	0	0	0	0	1	0	0	0
2009	0	0	1	1	0	1	0	1	0	1
2011	0	0	0	0	0	0	0	0	1	0
unknown	0	0	0	0	1	0	0	0	0	1

Table C.8: Whitehouse of Washington landmark matching using SIFT/SIFT matcher

SIFT/SIF T	186	194	194	196	196	199	200	2009	201	2012
	3	5	8	5	6	3	9	b	2	b
1863	1	0	1	0	0	0	0	0	0	0
1945	0	1	1	0	0	0	1	0	0	0
1948	1	0	1	0	0	1	1	1	1	0
1965	0	0	0	1	1	0	0	0	0	0
1966	0	0	0	1	1	0	0	0	0	0
1993	0	0	0	0	0	1	0	1	1	0
2009	0	0	0	0	0	0	1	1	0	1
2009b	0	0	0	0	0	1	1	1	1	1
2012	0	0	1	0	0	1	0	0	1	0
2012b	0	0	1	0	0	0	1	1	1	1

Appendix D: Landmark matching tables for SURF/SURF feature matching technique

Table D.1: Pyramids of Giza landmark matching using SURF/SURF matcher

SURF/SURF F	185	188	194	196	197	199	199	200	201	unknow
	8	0	0	9	8	2	6	6	1	n
1858	1	0	0	0	0	0	0	0	0	0
1880	0	1	0	0	0	0	0	0	0	0
1940	0	0	1	0	0	0	1	0	0	0
1969	0	0	0	1	0	0	0	0	0	0
1978	0	0	0	0	1	0	0	0	0	0
1992	0	0	0	0	0	1	0	0	0	0
1996	0	0	1	0	1	0	1	0	0	0
2006	0	0	0	0	0	0	0	1	0	1
2011	0	0	0	0	0	0	0	0	1	0
unknown	0	0	0	0	0	0	0	0	0	1

Table D.2: Coliseum of Rome landmark matching using SURF/SURF matcher

SURF/SURF F	185 5	189 0	189 6	197 6	200 5	200 7	2007 b	201 0	201 1	unknow n
1855	1	1	1	0	0	0	0	0	0	0
1890	0	1	1	0	0	0	0	0	0	0
1896	1	0	1	0	0	0	0	0	0	0
1976	0	0	0	1	1	0	0	0	0	0
2005	0	1	0	1	1	1	0	1	1	1
2007	0	0	0	0	1	1	0	1	1	1
2007b	0	0	0	0	0	0	1	0	0	1
2010	1	0	0	0	1	1	0	1	1	1
2011	0	0	0	0	1	1	0	1	1	1
unknown	0	0	0	0	1	1	1	1	1	1

Table D.3: Dome Rock landmark matching using SURF/SURF matcher

SURF/SURF	1900	1910	1915	1959	1966	1989	2009	2010	2012	2012b
1900	1	0	1	1	1	0	0	0	1	0
1910	0	1	0	0	0	0	0	0	0	0
1915	1	0	1	1	1	0	0	0	0	0
1959	1	0	1	1	1	0	0	0	0	0
1966	1	0	0	0	1	0	0	1	1	0
1989	0	0	0	0	0	1	1	1	0	1
2009	0	0	0	0	0	1	1	0	0	1
2010	1	0	1	0	1	1	0	1	1	0
2012	0	0	0	0	0	0	0	1	1	0
2012b	0	0	0	0	0	1	1	1	0	1

Table D.4: Empire State Building landmark matching using SURF/SURF matcher

SURF/SURF	1936	1940	1942	1955	1957	1959	1996	2007	2007b	2011
1936	1	1	0	0	0	0	0	1	0	1
1940	1	1	1	1	1	0	0	0	0	0
1942	0	0	1	1	0	0	1	0	0	0
1955	0	0	0	1	0	0	1	1	1	1
1957	1	0	1	1	1	1	0	1	0	0
1959	0	1	0	1	1	1	0	1	1	1
1996	0	0	0	0	0	1	1	1	1	1
2007	0	0	0	0	0	1	1	1	0	1
2007b	0	1	1	0	0	1	1	0	1	0
2011	0	0	0	0	0	0	0	1	0	1

Table D.5: Hagia-Sophia landmark matching using SURF/SURF matcher

SURF/SURF	1890	1900	1910	1977	1994	1996	2004	2008	2010	2011
1890	1	0	0	0	0	0	0	0	0	0
1900	0	1	0	0	0	0	0	0	0	0
1910	1	1	1	0	0	0	0	0	0	0
1977	0	0	0	1	0	0	0	1	0	0
1994	0	0	0	0	1	1	1	0	0	0
1996	0	0	0	0	1	1	0	0	1	1
2004	0	0	0	0	1	0	1	0	0	0
2008	0	0	0	1	0	0	1	1	0	1
2010	0	0	0	0	0	0	0	0	1	1
2011	0	0	0	0	0	1	1	0	0	1

Table D.6: Machu-Picchu landmark matching using SURF/SURF matcher

SURF/SURF	1900	1911	1980	1985	1996	1999	2005	2007	2009	2011
1900	1	0	0	1	0	0	0	0	0	0
1911	0	1	1	0	1	0	0	0	0	0
1980	0	0	1	0	1	0	0	0	1	0
1985	1	0	0	1	1	0	0	0	0	0
1996	0	1	0	1	1	0	0	0	1	1
1999	0	0	0	0	0	1	0	0	0	0
2005	0	1	1	0	0	0	1	0	1	0
2007	0	0	0	0	0	0	0	1	1	1
2009	0	0	0	0	1	0	1	1	1	1
2011	0	0	0	0	1	0	0	0	1	1

Table D.7: Stonehenge landmark matching using SURF/SURF matcher

SURF/SURF F	188	190	192	195	1955	200	200	200	201	unknow n
	7	8	8	5	b	7	8	9	1	
1887	1	0	0	0	0	0	0	0	0	0
1908	0	1	0	0	0	0	0	0	0	0
1928	0	1	1	1	1	1	0	1	0	0
1955	1	0	1	1	0	0	0	1	0	0
1955b	0	0	1	0	1	0	0	0	0	1
2007	0	0	0	0	0	1	0	0	0	0
2008	0	0	0	0	0	0	1	0	0	0
2009	0	0	0	0	0	0	0	1	0	0
2011	0	0	0	0	0	0	0	0	1	0
unknown	0	0	1	0	1	0	0	0	0	1

Table D.8: Taj-Mahal landmark matching using SURF/SURF matcher

SURF/SURF	1874	1890	1910	1910b	1943	1966	1986	1993	2006	2011
1874	1	0	0	0	1	0	0	0	0	0
1890	0	1	0	0	1	0	0	1	0	0
1910	0	0	1	1	0	0	0	0	0	0
1910b	0	0	0	1	0	0	0	0	0	0
1943	1	1	1	1	1	0	1	0	1	0
1966	0	0	0	0	0	1	0	0	0	0
1986	0	0	0	0	0	0	1	1	1	0
1993	0	0	0	0	0	0	1	1	1	0
2006	0	0	0	0	0	0	1	1	1	1
2011	0	0	0	0	0	0	0	1	1	1

Table D.9: Whitehouse of Washington landmark matching using SURF/SURF matcher

SURF/SURF	1863	1945	1948	1965	1966	1993	2009	2009b	2012	2012b
1863	1	0	0	0	0	0	0	0	0	0
1945	0	1	0	0	0	0	0	0	0	0
1948	0	0	1	0	0	1	0	1	1	0
1965	0	0	0	1	0	0	0	0	0	0
1966	0	0	0	0	1	0	0	0	0	0
1993	0	0	1	1	0	1	1	1	1	0
2009	0	1	0	0	0	0	1	0	1	1
2009b	0	0	0	0	0	1	1	1	1	1
2012	0	0	0	0	0	1	1	1	1	1
2012b	0	0	0	0	0	1	1	1	1	1

Appendix E: Automatic image registration applied on landmark images



Figure E.1: Registered images of the Dome Rock landmark image set

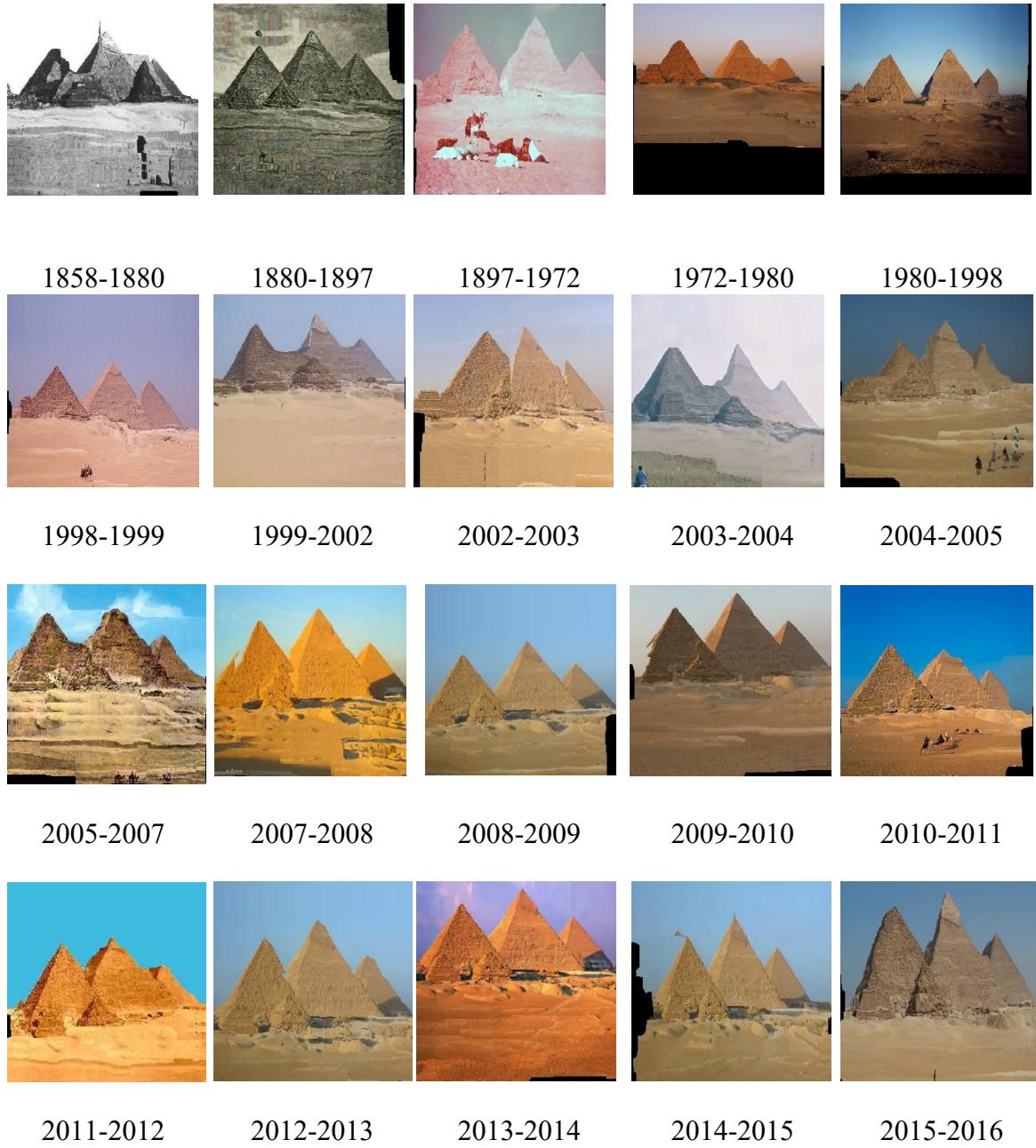
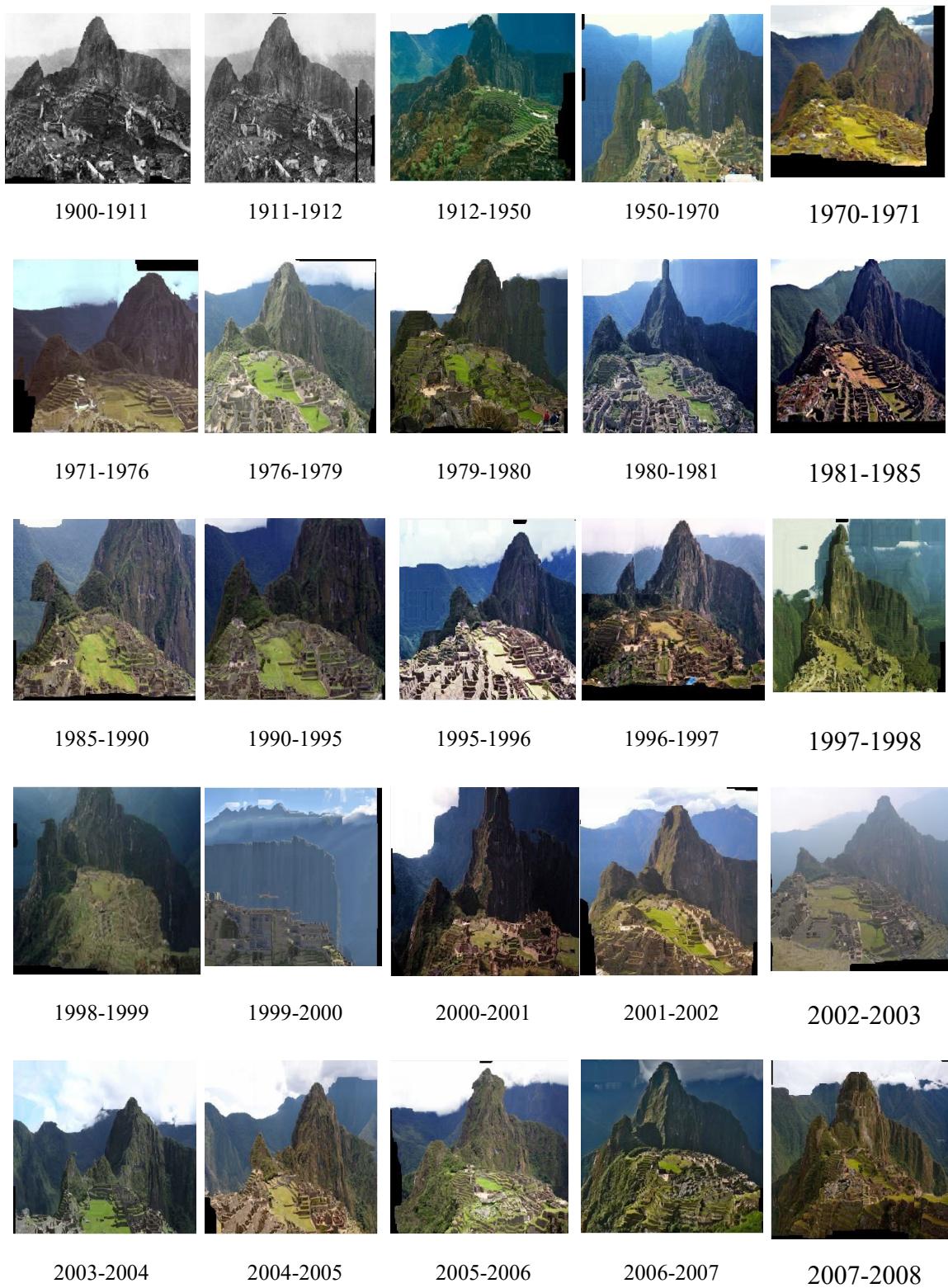


Figure E.2: Registered images of the Pyramids of Giza landmark image set





2008-2009

2009-2010

2010-2011

2011-2012

2012-2013



2013-2014

2014-2015

2015-2016

Figure E.3: Registered images of the Machu-Picchu landmark image set



Figure E.4: Registered images of the Empire State Building landmark image set



Figure E.5: Registered images of the Hagia-Sophia landmark image set

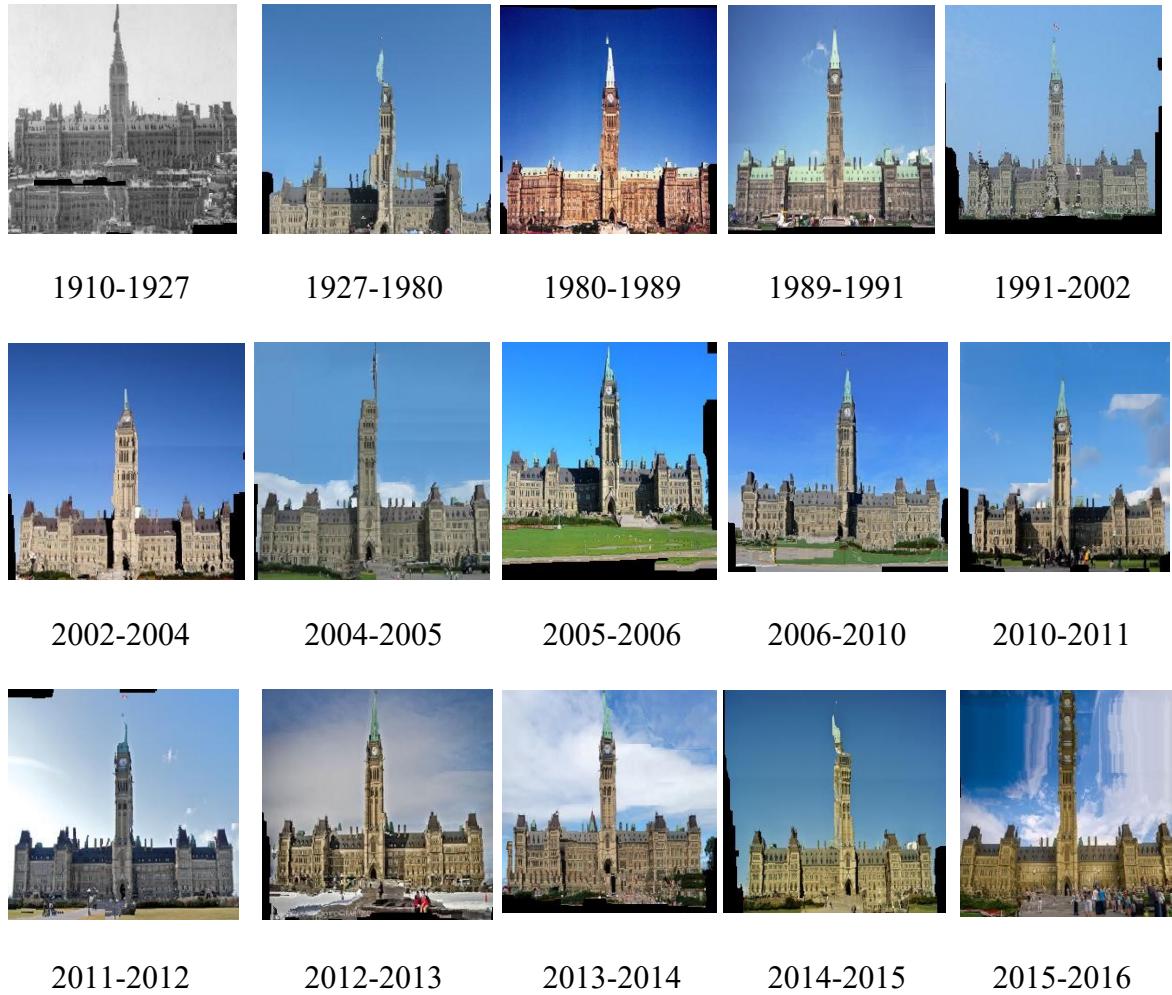


Figure E.6: Registered images of the Parliament Hill of Canada landmark image set



Figure E.7: Registered images of the Stonehenge landmark image set



Figure E.8: Registered images of the Whitehouse landmark image set

Appendix F: User Experience

The implemented timescape image panorama offers an opportunity to the user to explore a number of famous landmarks around the world. The exploration could be used by the user through the possibility to select certain landmark from a given menu popped up by the module Timescapes Panorama Display (TPD) as shown in Fig F.1 bellow:

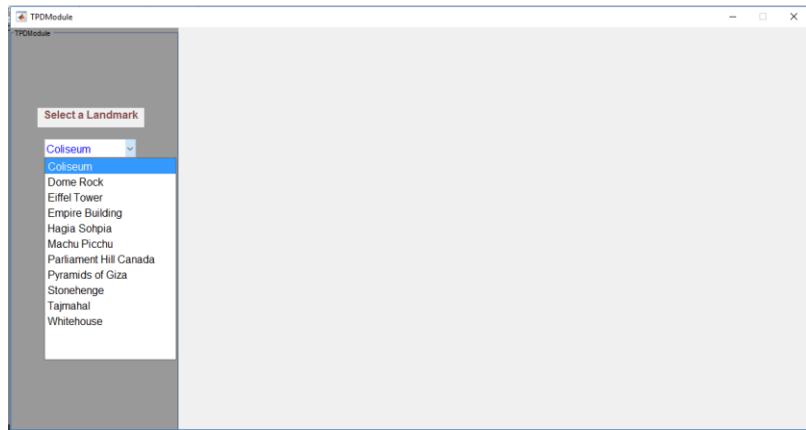


Figure F.1: Screenshot for a user request to select a landmark for timescape display

The system will respond by showing the reference image of the selected landmark which is the Coliseum of Rome, image of 2001 for Coliseum landmark as shown in Fig. F.2:

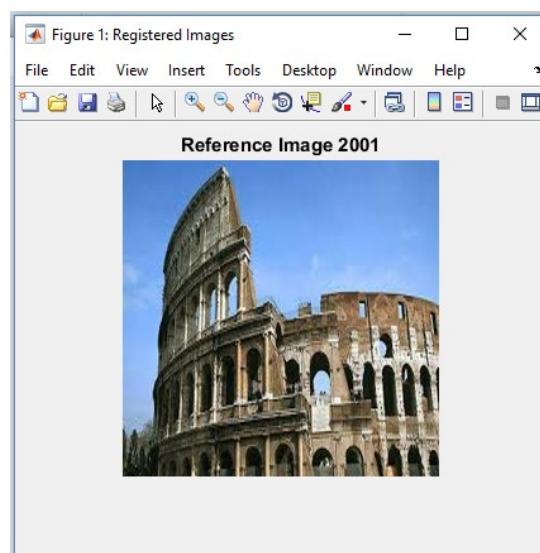


Figure F.2: Screenshot shows the reference image of the Coliseum of Rome landmark

Then, the system prompts the user to scroll the mouse wheel to navigate through the timeline panorama display of the selected landmark. This navigation may be to the oldest images by scrolling the mouse wheel upward as shown in Fig. F.3 bellow:

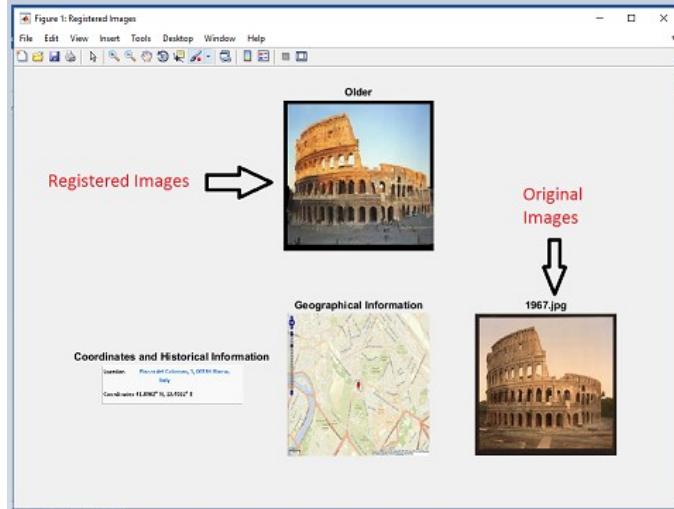


Figure F.3: User navigation to older images, system responds showing registered images with geographic information, historic information, and landmark coordinates

If the user desires to navigate to newest images of the landmark reaching to the present, the system responds as shown in Fig. F.4:

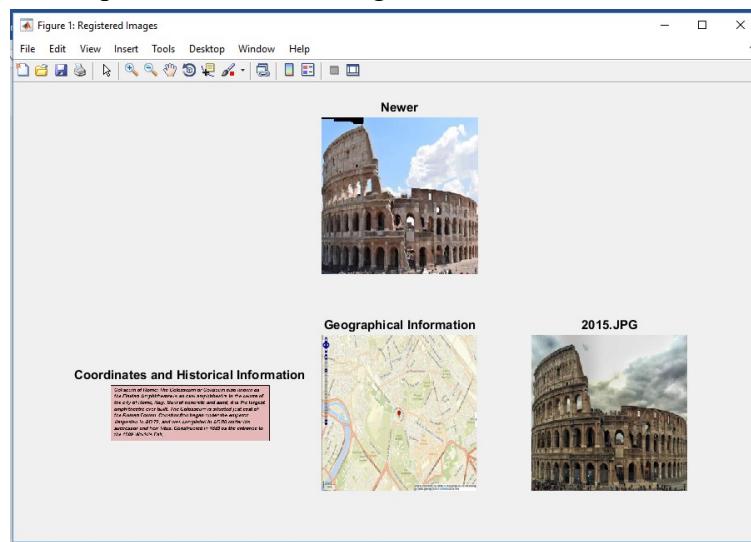


Figure F.4: User navigation to newer images, system responds showing registered images with geographic information, historic information, and landmark coordinates

We can notice that the system provides an opportunity to the user to navigate to the past or the present of many landmarks around the world, showing high quality registered images with supporting historical and geographical information; as well as, accompanied videos, documents, and maps depending on the availability of these media items on the Internet.

References

- [1] A. Stevenson, “Oxford Dictionary of English,” 3rd Ed. *Oxford University Press*, 2010.
- [2] M. Brown and D. Lowe. “Recognising panoramas”. In *Proceedings of the 9th International Conference on Computer Vision (ICCV03)*, Vol. 2, pp. 1218–1225, 2003.
- [3] P. Baudish, D. Tan, D. Steedly, E. Rudolph, M. Uyttendaele, C. Pal, and R. Szeliski. “Panoramic Viewfinder: Providing a Real-time Preview to Help Users Avoid Flaws in Panoramic Pictures” In *Proceedings of Conference for the Computer-Human Interaction Special Interest Group of the Human Factors Society of Australia*, 2005.
- [4] S. Agarwal, N. Snavely, Simon, I.S.M. Seitz, and R. Szeliski, ”Building Rome in a Day “, *Communications of the ACM* , Vol. 54 , No. 10 , 2011.
- [5] D. Capel, and A. Zisserman, “Automatic Mosaicing with Super-Resolution Zoom” , In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* , 1998.

- [6] M. Brown and D. G. Lowe, "Automatic panoramic image stitching using invariant features," *International Journal of Computer Vision*, Vol. 74, No. 1 , pp. 59-73 , 2007.
- [7] V. C.S. Chew and F.L. Lian. "Panorama Stitching Using Overlap Area Weighted Image Plane Projection and Dynamic Programming for Visual Localization" *In the 2012 IEEE/ASME International Conference on Advanced Intelligent Mechatronics*, Kaohsiung, Taiwan, 2012.
- [8] D. G. Lowe. "Distinctive image features from scale-invariant keypoints." *International Journal of Computer Vision*, Vol. 60, No. 2, pp. 91-110, 2004.
- [9] H. Bay, T. Tuytelaars, and L. Van Gool. "Surf: Speeded up robust features." In *Proceedings of the European Conference on Computer Vision*, 2006.
- [10] S. Leutenegger, M. Chli and R. Y. Siegwart. "BRISK: Binary Robust Invariant Scalable Keypoints" In *2011 IEEE International Conference on Computer Vision*, pp. 2548-2555, 2011.
- [11] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski. "ORB: an efficient alternative to SIFT or SURF" In *Proceedings of 2011 IEEE International Conference on Computer Vision*, pp. 2564-2571, 2011.
- [12] A. Whitehead, and J. Opp. "Timescapes: Putting History in Your Hip Pocket". In *Proceedings of Computers and Their Applications Conference (CATA, Hawaii, USA)*, 2013.
- [13] Online, https://en.wikipedia.org/wiki/Cultural_heritage , Accessed on March 11, 2016.

- [14] G. Thomas Tanselle, “Literature and Artifacts,” 1st Edition. , *Oak Knoll Press*, 1998.
- [15] UNESCO, “New World Heritage Sites in Danger,” In *34th Session of the UNESCO Committee*. 2010.
- [16] H.K. Ali, and A. Whitehead, “Feature Matching for Aligning Historical and Modern Images,” *International Journal of Computers and Their Applications*, Vol. 21, No. 3, pp. 188-201, 2014.
- [17] S., Marilyn; D. Cateforis, S. Addiss, "Art History," 2nd Ed, *Upper Saddle River, New Jersey: Pearson Education*. pp. 964–967, 2005.
- [18] B. Zitová, and J. Flusser, “Image registration methods: a survey,” In *Image and Vision Computing*, Vol. 21, pp. 977–1000, 2003.
- [19] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert, “High Accuracy Optical Flow Estimation Based on a Theory for Warping,” In *Proceedings of 8th European Conference on Computer Vision*, 2004.
- [20] J. Pearl, “Probabilistic reasoning in intelligent systems: networks of plausible inference,” *San Francisco: Morgan Kaufmann Publishers, Inc.*, 1988.
- [21] N. Komodakis and G. Tziritas, “Image Completion Using Efficient Belief Propagation Via Priority Scheduling and Dynamic Pruning,” *IEEE Transactions On Image Processing*, Vol. 16, No. 11, pp. 2649-2661, 2007.
- [22] C. Liu, J. Yuen, and A. Torralba, “SIFT Flow: Dense Correspondence across Scenes and Its Applications,” In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 33, No. 5, 2011.
- [23] R. Szeliski, “Video mosaics for virtual environments,” *IEEE Computer Graphics*

- and Applications*, Vol. 16, No. 2, pp. 22–30, 1996.
- [24] R. Szeliski, “Image Alignment and Stitching: A Tutorial,” *Foundations and Trends in Computer Graphics and Vision*, Vol. 2, No. 1, 2006.
- [25] Panorama Wikipedia website. <https://en.wikipedia.org/wiki/Panorama> . Accessed on April 11, 2016.
- [26] H. K. Ali and A. Whitehead, “Timescapes: Automatic Visual Temporal Panorama Generation of Historical Landmarks,” *Paper Submitted to the Journal of Computing and Cultural Heritage*, 2016.
- [27] Time-lapse Wikipedia. https://en.wikipedia.org/wiki/Time-lapse_photography . Accessed on April 13, 2016.
- [28] Dipity website. <http://www.dipity.com> . Accessed on April 15, 2016.
- [29] Rideau Timescapes Application Website. <http://timescapes.carleton.ca/index.php> . Accessed on April, 15, 2016.
- [30] D. M W Powers, “Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation,” *Journal of Machine Learning Technologies*, Vol. 2, No. 1, 2007.
- [31] H.K. Ali, and A. Whitehead, “Image Subset Selection Using Gabor Filters and Neural Networks,” *The International Journal of Multimedia & Its Applications*, Vol.7, No.2, pp.43-55, 2014.
- [32] J. Campbell , "Film and cinema spectatorship: melodrama and mimesis," *Polity*, 2005.
- [33] C. M. Falco, "Ibn al-Haytham and the Origins of Computerized Image Analysis," In *Proceedings of the International Conference on Computer Engineering &*

Systems, ICCES '07, xx - xxiii, 2007.

- [34] E. Rosten, R. Porter, and T. Drummond, "Faster and Better: A Machine Learning Approach to Corner Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 32, No. 1, pp. 105-119, 2010.
- [35] D. Puig and M. A. Garcia, "Automatic texture feature selection for image pixel classification," *Pattern Recognition*, Vol. 39, Issue 11, pp. 1996-2009, 2006.
- [36] C. Michael, "The Origins of Photojournalism in America," *Smithsonian Institution Press*. 1992.
- [37] Dry Plate Wikipedia. http://en.wikipedia.org/wiki/Dry_plate Accessed on April 9, 2014.
- [38] Photographic film Wikipedia. http://en.wikipedia.org/wiki/History_of_the_camera . Accessed on April 9, 2014.
- [39] M. R. Peres, "The Focal Encyclopedia of Photography," 4th Ed, *Focal Press*. 2007.
- [40] Camera Wikipedia. <http://en.wikipedia.org/wiki/Camera> . Accessed on April 9, 2014.
- [41] A. Alahi, R. Ortiz, and P. Vandergheynst, "FREAK: Fast Retina Keypoint," In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 510-517, 2012.
- [42] History of camera Wikipedia. http://en.wikipedia.org/wiki/History_of_the_camera . Accessed on April 9, 2014.
- [43] M. S. Nixon and A. S. Aguado, "Feature Extraction & Image Processing for Computer Vision," 3rd Ed. *Academic Press*. 2012.

- [44] C. Harris and M. Stephens, "A combined corner and edge detector," In *Proceedings of the Alvey Vision Conference*, pp. 147 – 151, 1988.
- [45] Y. Ke and R. Suthankar, "PCA-SIFT: A More Distinctive Representation for Local Image Descriptors," In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 2, pp. 506-513, 2004.
- [46] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "BRIEF: Binary Robust Independent Elementary Features," In *Proceedings of the 11th European conference on Computer vision: Part IV*, pp. 778-792, 2010.
- [47] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: an efficient alternative to SIFT or SURF," In *2011 IEEE International Conference on Computer Vision*, pp. 2564-2571, 2011.
- [48] B. D. Lucas and T. Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision," In *Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI)* Vancouver. British Columbia, pp.674-679. 1981.
- [49] S. F. Chang, W. Chen, and H. Sundaram, "Semantic Visual Templates: Linking Visual Features to Semantics," In *Proceedings of International Conference on Image Processing*.1998.
- [50] N. A. Mahmon and N. Ya'acob, "A Review on Classification of Satellite Image Using Artificial Neural Network (ANN)". In *IEEE 5th Control and System Graduate Research Colloquium*. 2014.

- [51] Z. Xiong, K. Chen, C. GU, Y. Liang, and F. Yu, “An Algorithm of Image Classification Based on BP Neural Network,” *IEEE International Conference of Computer Science and Automation Engineering (CSAE)*. 2012.
- [52] D. Xu and Z. Qu, “An Image Classification Method Based on Matching Similarity and TF-IDF Value of Region,” In *6th International Symposium on Computational Intelligence and Design*. 2013.
- [53] B. Li, W.i Li and D. Zhao, “Multi-Scale Feature Based Medical Image Classification,” In *3rd International Conference on Computer Science and Network Technology*. 2013.
- [54] H. Al-Sahaf, A. Song, K. Neshatian, and, M. Zhang, “Extracting Image Features for Classification By Two-Tier Genetic Programming,” In *IEEE World Congress on Computational Intelligence*.2012.
- [55] A. Kharrat, K. Gasmi, M. B. Messaoud, N. Benamrane and M. Abid, “Genetic Algorithm for Feature Selection of MR Brain Images Using Wavelet Co-occurrence,” In *Proceedings of International Conference on Graphic and Image Processing (ICGIP)*. 2011.
- [56] M. Coustaty, S. Uttama, and J.M. Ogier “Extraction of light and specific features for historical image indexing and matching,” In *Proceedings of 21st International Conference on Pattern Recognition (ICPR 2012)* , Tsukuba, Japan , 2012.
- [57] R. Paretí and N. Vincent, “Ancient initial letters indexing,” In *ICPR '06: Proceedings of the 18th International Conference on Pattern Recognition*, pp. 756–759, 2006.
- [58] H. Chouaib, F. Clopet, and N. Vincent, “Graphical drop caps indexing,” In *Eighth*

IAPR International Workshop on Graphics Recognition, La Rochelle, pp. 179–185, 2009.

- [59] J.-M. O. Thi Thuong Huyen Nguyen, M. Coustaty, “Bags of strokes based approach for classification and indexing of drop caps,” In *ICDAR 2011*, pp. 349–353, 2011.
- [60] S. Uttama, P. Loonis, M. Delalandre, and J.-M. Ogier, “Segmentation and retrieval of ancient graphic documents,” *LNCS*, pp. 88–98, 2006.
- [61] M.S. El-Hussainy, M.A. Baraka, and M. A. El-Hallaq, “A methodology for image matching of historical maps,” In *e-Perimetron*, Vol. 6, No. 2, pp.77-95, 2011.
- [62] E. Livieratos, “On the Study of the Geometric Properties of Historical Cartographic Representations,” In *Cartographica* , Vol. No. 41, Issue 2, 2006.
- [63] V. Mnih and G.Hinton, “Learning to label aerial images from noisy data,” In *Proceedings of International Conference on Machine Learning(ICML)*, 2012.
- [64] P. Chhabra, “An architecture for video change detection for unmanned aerial vehicle applications,” *MSc Thesis*, Heriot Watt University, Edinburgh, 2009.
- [65] C. Balletti and F. Guerra, “Image matching for historical maps comparison,” In *e-Perimetron*, Vol. 4, No. 3, pp. 180-186, 2009.
- [66] W. J. T. Mitchell, “What Is an Image?,” *New Literary History*, Vol. 15, No. 3, 1984.
- [67] J. Shi, “Statistical Dependence Measure Based Multi-Model Image Registration and Registration Assisted Non-Parametric Image Segmentation,” *PhD Thesis, University of Florida*. 2013
- [68] T.M. Buzug, J. Weese, C. Fassnacht, and C. Loren, “Image Registration: Convex

- Weighting Functions for Histogram-Bsaed Similarity Measures," In *CVRMed-MRCAS '97: First Joint Conference Computer Vision, Virtual Reality*, 1997.
- [69] M. Morrone, Concetta, and D.C. Burr, "Feature Detection in Human Vision: A Phase-Dependent Energy Model," In *Proceedings of the Royal Society of London. Series B, Biological Sciences*. pp.221-45. 1988.
- [70] H. Moravec, "Obstacle Avoidance and Navigation in the Real World by a Seeing Robot Rover," *Technical Report CMU-RI-TR-3, Carnegie-Mellon University, Robotics Institute*, 1980.
- [71] A. L. Yuille, "Deformable templates for face recognition," *Journal of Cognitive Neuroscience*, Vol. 3, No. 1, pp. 59-70. 1991.
- [72] D.E. Benn, M.S. Nixon, and J.N. Carter, "Extending concentricity analysis by deformable templates for improved eye extraction," In *Proceedings of the Second International Conference on Audio- and Video-Based Biometric Person Authentication*. pp. 1-6. 1999.
- [73] P. F. Felzenszwalb and D. P. Huttenlocher, "Pictorial Structures for Object Recognition," *International Journal of Computer Vision*, Vol. 61, Issue 1, pp 55-79. 2005.
- [74] J. S. Yedidia, W. T. Freeman, and Y. Weiss, "Understanding belief propagation and its generalizations," *Published in Exploring artificial intelligence in the new millennium book, Morgan Kaufmann Publishers Inc. San Francisco, CA, USA*, pp.239-269. 2003.
- [75] Structural Similarity Wikipedia, https://en.wikipedia.org/wiki/Structural_similarity
Accessed June 21, 2016.

- [76] H.K Ali and A. Whitehead, “Registration of Modern and Historic Imagery for Timescape Creation,” In *Proceedings of 13th Conference on Computer and Robot Vision*, Victoria, BC, 2016.
- [77] K. Mikolajczyk, , and C. Schmid, “A performance evaluation of local descriptors” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 27, No. 10, 2005.
- [78] W. Freeman and E. Adelson, “The Design and Use of Steerable Filters,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 9, pp. 891-906, 1991.
- [79] L. Florack, B. ter Haar Romeny, J. Koenderink, and M. Viergever, “General Intensity Transformations and Second Order Invariants,” In *Proceedings of 7th Scandinavian Conference Image Analysis*, pp. 338-345, 1991.
- [80] L. Van Gool, T. Moons, and D. Ungureanu. “Affine / photometric invariants for planar intensity patterns.” In *Proceedings of the 4th European Conference on Computer Vision, Cambridge, UK* , pp. 642-651, 1996.
- [81] P. L. Rosin. “Measuring corner properties,” *Computer Vision and Image Understanding*, Vol. 73, No. 2, pp.291 - 307, 1999.
- [82] S. Leutenegger, M. Chli and R. Y. Siegwart. “BRISK: Binary Robust Invariant Scalable Keypoints” In *2011 IEEE International Conference on Computer Vision*, pp. 2548-2555, 2011.

- [83] E. Tola, V. Lepetit, and P. Fua, “DAISY: An Efficient Dense Descriptor Applied to Wide-Baseline Stereo” *IEEE Transactions On Pattern Analysis and Machine Intelligence*, Vol. 32, No. 5, 2010.
- [84] B. Fan, F. Wu, and Z. Hu, “Rotationally Invariant Descriptors Using Intensity Order Pooling” *IEEE Transactions On Pattern Analysis and Machine Intelligence*, Vol. 34, No. 10, 2012.
- [85] L. V. N. Duy, “Review and Enhancement Optimization Methods in Image Registration,” *Master Thesis, Imperial College London*. 2009.
- [86] J. Nocedal and S. J. Wright. “Numerical optimization,” *Springer*, 2006.
- [87] K. Kutulakos, “Markov Random Fields,” *Notes for the Course Foundations of Computer Vision CSC 2503, Toronto University*, 2015.
- [88] Belief Propagation Wikipedia, https://en.wikipedia.org/wiki/Belief_propagation
Accessed June 15, 2016.
- [89] V. Venkateswar a,d R. Chellappa, “ Extraction of Straight Lines in Arial Images,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Vol. 14, No. 11, 1993.
- [90] P. Kamavisdar, S. Saluja, and S. Agrawal, “A Survey on Image Classification Approaches and Techniques,” *International Journal of Advanced Research in Computer and Communication Engineering*, Vol. 2, Issue 1, 2013.
- [91] D. A. Forsyth and J. Ponce, “Computer Vision: A Modern Approach,”, 2nd Ed, *Pearson*. 2011.

- [92] R. C. Gonzalez and R. E. Woods,” Digital Image Processing,” *Prentice-Hall*, 3rd Edition. 2008.
- [93] D. J. Ketcham, R.W. Lowe and J. W. Weber, “Image enhancement techniques for cockpit displays”, *Technical Report, Hughes Aircraft Company*. 1974.
- [94] D. Gabor, “Theory of communications”, *Journal of International Electrical Engineers*, Vo. 93, pp. 427-457. 1946.
- [95] L. Shen and L. Bai, “A review on Gabor wavelets for face recognition,” *Pattern Analysis and Applications*. Vol. 9, Issue 2, pp. 273–292. 2006.
- [96] J. Kamarainen, V. Kyrki and H. Kalviainen, “Fundamental frequency Gabor filters for object recognition”, In *Proceedings of the 16th International Conference on Pattern Recognition*, pp. 628-631. 2002.
- [97] S. Berisha, “Image classification using Gabor filters and machine learning”, *MSc Thesis, Wake Forest University*, Winston-Salem, North Carolina. 2009.
- [98] H. K. Ali and A. Whitehead, “Subset Selection for Landmark Modern and Historic Images,” In *Proceedings of 2nd International Conference on Signal and Image Processing (SIGL 2015)*, Geneva, Switzerland, 2015.
- [99] MATLAB 2016 Documentation in Choosing a Multilayer Network Training Function <http://www.mathworks.com/help/nnet/ug/choose-a-multilayer-neural-network-training-function.html> Accessed on July 21, 2016.
- [100] L. Jacobson, “Introduction to Artificial Neural Networks Part 2 – Learning,” Online Article, <http://www.theprojectspot.com/tutorial-post/introduction-to-artificial-neural-networks-part-2-learning/8> Accessed on July 21, 2016.
- [101] M. Jiang, B. Zhang, and Z. Luo, “Fast Learning Algorithms for Feedforward

- Neural Networks,” In *Applied Intelligence*, Vol. 18, pp. 37–54, 2003.
- [102] Opencv Documentation. Common Interface of Descriptor Matchers. http://opencv.willowgarage.com/documentation/cpp/features2d_common_interfaces_of_descriptor_matchers.html Accessed on September 19, 2013.
- [103] H. P. Trivedi and S. A. Lioyd, “The Role of Disparity Gradient in Stereo Vision,” *Perception* Vol. 14, pp. 685-690, 1985.
- [104] H. K. Ali and A. Whitehead, “Modern to Historic Image Matching: ORB/SURF an Effective Matching Technique,” In *Proceedings of International Conference on Computers and Their Application (CATA 2014)* , Las Vegas, Nevada, USA, 2014.
- [105] Z. Zexu, and C. Pinguan, “A Reliable Method of Image Registration Based on Optical Flow Field and Feature Extraction,” *Chinese Journal of Electronics*, Vol. 17, No. 1, pp.12-15, 2008.
- [106] J. Sánchez, E. Meinhardt-Llopis, and G. Facciolo, “TV-L1 Optical Flow Estimation,” *Published in Image Processing On Line*, Vol. 3, pp. 137-150, 2013.
- [107] J.L. Baron, D.J. Fleet, and S.S. Beauchemin, “Performance of Optical Flow Techniques,” *International Journal of Computer Vision*, Vol. 12, No. 1, pp. 43-77, 1994.
- [108] Optical Flow on Wikipedia, https://en.wikipedia.org/wiki/Optical_flow Accessed on February 12 2016.
- [109] B.D. Lucas, and T. Kanade, “An iterative image registration technique with an application to stereo vision,” In *Proceedings of Imaging Understanding Workshop*, pp. 121-130, 1981.

- [110] B.K.P. Horn, and B.G. Schunck, "Determining optical flow," *Artificial Intelligence*, Vol 17, pp 185–203, 1981.
- [111] B. Glocker , N. Komodakis, G. Tziritas, N. Navab , and N. Paragios, “Dense Image Registration through MRFs and Efficient Linear Programming,” *Medical Image Analysis*, Vol. 12, No. 6, pp. 731-741, 2008.
- [112] A. Materka1, M. Strzelecki, R Lerski, and L. Schad, “Toward Automatic Feature Selection of Texture Test Objects for Magnetic Resonance Imaging,” In *Proceedings of the 11th Portuguese Conference on Pattern Recognition, Porto*, pp. 11-16, 2000.
- [113] R. W. Marcato, “Optimizing an Inverse Warper,” *MSc Thesis*, Massachusetts Institute of Technology, 1998.
- [114] Image Texture in Wikipedia, https://en.wikipedia.org/wiki/Image_texture
Accessed on September 16, 2016.
- [115] M. Pharr (Editor), “GPU Gems 2: Programming Techniques for High-Performance Graphics and General-Purpose Computation,” *Addison-Wesley Professional*, 2005.
- [116] M. Cooper, J. Foote, A. Girgensohn, and L. Wilcox, “Temporal Event Clustering for Digital Photo Collections,” In *Proceedings of ACM International Conference on Multimedia*, pp. 364-373, 2003.
- [117] N. Snavely, S. M. Seitz, and R. Szeliski, “Photo Tourism: Exploring Photo Collections in 3D,” *Journal ACM Transactions on Graphics (TOG)*, Vol. 25, Issue 3, pp. 835-846, 2006.

- [118] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, “Image Quality Assessment: From Error Visibility to Structural Similarity,” *IEEE Transactions On Image Processing*, Vol. 13, No. 4, pp.600-612, 2004.
- [119] R.J. Hyndman, and A.B. Koehler, “Another look at measures of forecast accuracy,” *International Journal of Forecasting*, Vol. 22, Issue 4, pp.679-688, 2006.
- [120] S. Cheng , V. Stanković and L. Stanković, “Improved Sift-Based Image Registration Using Belief Propagation,” In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, 2009.
- [121] D. Sharon, “Loopy Belief Propagation in Image-Based Rendering,” *Technical Report, Department of Computer Science, University of British Columbia*, 2004.
- [122] R. Szeliski, “Image Alignment and Stitching: A Tutorial,”, *Technical Report*, 2006.