Robust Image Registration Methods for Varying Image Quality and Imaging Conditions in Augmented Reality Systems

THESIS

Submitted in partial fulfilment
of the requirements for the degree of
DOCTOR OF PHILOSOPHY

by

Neetika Gupta

Under the Supervision of
Prof. Mukesh Kumar Rohil

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI

MARCH 2019

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI

CERTIFICATE

This is to certify that the thesis entitled **Robust Image Registration Methods for Varying Image Quality and Imaging Conditions in Augmented Reality Systems** and submitted by **Neetika Gupta** ID No. **2013PHXF0010P** for award of Ph.D. of the Institute embodies original work done by her under my supervision.

Signature of the Supervisor

Prof. Mukesh Kumar Rohil

Associate Professor, Department of Computer Science & Information Systems

Birla Institute of Technology and Science, Pilani

Pilani Campus

Vidya Vihar, Pilani – 333031 (Rajasthan) INDIA

Date

Acknowledgements

Foremost, I wish to express my sincere gratitude to my supervisor Prof. Mukesh Kumar Rohil, Associate Professor, Department of Computer Science & Information Systems, Birla Institute of Technology and Science, Pilani, Rajasthan, India, for his guidance, motivation, support and immense knowledge.

My sincere thanks also goes to Prof. Souvik Bhattacharyya, Vice Chancellor; Prof. S.C. Sivasubramanian, former Acting Registrar; Mr. Ernest Samuel Ratnakumar J, Registrar; Prof. Sanjay Kumar Verma, former Dean, Academic Research Division; Prof. Srinivas Krishnaswamy, Dean, Academic–Graduate Studies & Research Division, Birla Institute of Technology and Science, Pilani; Prof. Ashoke Kumar Sarkar, Director; Prof. Hemant R. Jadhav, former Associate Dean, Academic Research Division; Prof. Jitendra Panwar, Associate Dean, Academic–Graduate Studies & Research Division, Birla Institute of Technology and Science, Pilani, Pilani Campus (Raj.), for providing me an opportunity to pursue and complete my PhD work from this esteemed institution.

I would also like to thank the members of the Doctoral Research Committee, Prof. Poonam Goyal, Head of the Department; Prof. J.P. Mishra; Prof. Rahul Banerjee; Prof. Navneet Goyal; former Head of the Department; Prof. Shan Balasubramanian, Professor; Prof. Sudeept Mohan, Professor; Prof. Mukesh Kumar Rohil, Associate Professor and Convener, Doctoral Research Committee, and Dr. Amit Dua, Assistant Professor, Department of Computer Science & Information Systems, Birla Institute of Technology and Science, Pilani, Pilani Campus (Raj.), for their guidance and support in fulfilling the various requirements of the Ph.D. programme. I also thank the members of the Doctoral Advisory Committee, Dr. Sundaresan Raman, Assistant Professor; Department of Computer Science & Information Systems, Birla Institute of Technology and Science, Pilani, Pilani Campus (Raj.) and Dr. Tathagata Ray, Head of The Department, Department Computer Science and Information Systems, Birla Institute of Technology and Science, Pilani, Hyderabad Campus (A.P.), for their valuable remarks.

I owe a lot to my parents, Mr. O.P. Gupta and Mrs. Usha Gupta, who encouraged and helped me at every stage of my personal and academic life. I thank them for their love and moral support during the entire period of this research work.

(Neetika Gupta)

Abstract

Augmented Reality (AR) is a widely accepted technology that aims in building a seamless bridge between the real and virtual world. Therefore, harmonious coexistence of virtual and real world makes AR experience to serve as an improved user's environment where virtual information is used as a tool for providing assistance in a particular activity. This technology exists in various forms, differentiating the building mechanism in terms of objectives and application use cases, the type of computations involved in making a successful AR experience are generally categorized as 1) Marker based AR: Marker based AR applications use user's device camera to distinguish a previously defined and structured marker in the real scene. Markers used in such a scenario are simple and distinctive patterns, such as a QR code, barcode etc., as they can be easily recognized and requires less computational power at the end-user device for tracking and recognition. The position and orientation of these markers defines the actual placement of virtual objects (graphic content and/or information in form of text, videos etc.) in the real environment. 2) Markerless approach: This approach make use of natural features from the image scene to identify the location of virtual objects in the real environment, i.e., without making any use of fiducial markers, markerless approach uses stable extracted features from the real scene, making the approach applicable to a wide variety of scenarios where placing a marker every time we wish to enhance the real surrounding with additional virtual information becomes a cumbersome task.

To achieve accurate and desired outcome from a markerless AR system, image registration plays a vital role in defining distinctive features from an image and tracking those features in subsequent image frames to evaluate the right position and orientation of virtual objects that are to be rendered in the real environment. However, image registration methods proposed till date still requires improvement when dealing with affine transformations, varying image quality and other changing imaging conditions.

This research work improves the image registration procedure for attaining accurate markerless AR by firstly performing a comparative analysis of six widely used feature detectors namely, Harris-Affine, Hessian-Affine, Maximally Stable Extremal Regions (MSER), Scale Invariant Feature Transform (SIFT), Affine-SIFT (ASIFT) and Speeded Up Robust Features (SURF). Due to lack of any prior research that explains the performance of these methods for varying image quality and imaging conditions, present research analyzed the performance of the six feature detectors based upon the quality of images used for experimentation and varying imaging conditions. The Image Quality Assessment (IQA) metrics used for quality evaluation of images are classified as No-Reference Image Quality Assessment (NR-IQA) metrics and Full-Reference Image Quality Assessment (FR-IQA) metrics. Changing imaging conditions are taken into consideration by selecting the image dataset containing 48 images, grouped into eight image-sets with six images in each set.

Each set defines alterations in an image scene with respect to viewpoint change, scale change, image blur, illumination change, and JPEG compression.

Present research also explores various features that deteriorates image/video quality and thus aimed in developing a fundamental analysis of spatial and temporal artifacts. Keeping all artifacts in mind, an NR-IQA model is designed for better estimation of distortions in an image. The proposed NR-IQA model achieves as efficiency boost of 37.6%, 0.6% and 5% as compared to NIQE, BRISQUE and BLIINDS-II NR-IQA metrics respectively. In addition, a No-Reference Video Quality Assessment model (NR-VQA) is designed by defining and estimating different feature distortion quantifications.

Based on the comparative study of six feature detectors, MSER detector is chosen for further research. The MSER detector retains low computational complexity, making it an appropriate selection for performing image registration in a markerless AR system. As AR, for all practical purposes, requires extensive computation for accurate view alignment and also demands real time performance, therefore, an improved method of feature detection using MSER is designed. The approach, Linear-MSER, used for feature detection uses the process of extracting the regions of interest using a true flood fill approach for building and maintaining the component tree. The present improved work, MSLinear-MSER, implements Linear-MSER at multiple scales of an image using octave formation in order to increase the affine invariance properties of the detector while achieving linear time complexity. The two detectors, Linear-MSER and MSLinear-MSER, are then combined one by one with SIFT and SURF descriptors for performance comparison. Performance of the four methods, namely Linear-MSER+SIFT, Linear-MSER+SURF, MSLinear-MSER+SIFT, and MSLinear-MSER+SURF, is then evaluated for varying imaging conditions. The results are compared along three parameters namely: time complexity, number of correct correspondences between image pairs and affine invariance property. MSLinear-MSER+SIFT performs best among the four methods in terms of time complexity and number of correct matches between an image pair when executed at 6 octaves and 5 levels. This observation is true for all the image-sets and many of the images in these images-sets have been affine transformed in one way or other. Using this method a prototype of an AR system is also developed to demonstrate the efficiency of MSLinear-MSER+SIFT detector and its efficiency in terms of correct augmentation using precision metric yields an accuracy of 0.9729.

In this research work, a novel feature descriptor based on circular and elliptical local sampling of pixels is also proposed, determining the neighborhood of the extracted features using circular and elliptical sampling. The main advantage of the approach is fast and robust matching results under varied imaging conditions of viewpoint change, scale change, illumination change etc. The proposed descriptor is tested on standard benchmark for evaluation and is proven to be 1.6 times faster as compared to the conventional SIFT descriptor while maintaining sufficient number of correspondences between an image pair.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Abbreviation	**Expanded Form**
2D	Two-Dimensional
3D	Three-Dimensional
6DOF	Six-Degree-of-Freedom
AR	Augmented Reality
ASIFT	Affine-SIFT
BIQI	Blind Image Quality Index
BLIINDS-II	BLind Image Integrity Notator using DCT Statistics-II
BRIEF	Binary Robust Independent Elementary Features
BRISK	Binary Robust Invariant Scalable Keypoints
BRISQUE	Blind/Reference-less Image Spatial Quality Evaluator
CCD	Charge-Coupled Device
CMoS	Complimentary Metal-Oxide Semiconductor
CS-LBP	Center-Symmetric Local Binary Patterns
CS-LTP	Center-Symmetric Local Ternary Pattern
DCT	Discrete Cosine Transform
DIIVINE	Distortion Identification-based Image Verity and Integrity Evaluation
DMOS	Differential Mean Opinion Score
DoG	Difference-of-Gaussian
EBR	Edge Based Region detector
FAST	Features from Accelerated Segment Test
FF	Fast Fading distortion
FR-IQA	Full-reference IQA
GB	Gaussian Blur distortion
GBIM	Generalized Block-edge Impairment Metric
GIS	Geographic/Geographical Information Systems
GLOH	Gradient location and orientation histogram
GPS	Global Positioning System
HMD	Head Mounted Display
HVS	Human Visual System
IBR	Intensity Based Region detector
IQA	Image Quality Assessment
J	JPEG compressed images
J2	JPEG2000 compressed images
KLT	Kanade-Lucas-Tomasi
LBP	Local Binary Pattern

Linear-MSER	MSER [Nister and Stewenius 2008]
LIVE	Laboratory for Image & Video Engineering
LoG	Laplacian-of-Gaussian
LTP	Local Ternary Pattern
MGV	Multivariate Gaussian
MLR	Multi-Linear Regression
MOS	Mean-Opinion Score
MSCN	Mean Subtracted Contrast Normalized
MSCR	Maximally Stable Color Regions
MSE	Mean Square Error
MSER	Maximally Stable Extremal Regions
MSLinear-MSER	Multi-scale Linear-MSER
MS-SSIM	Multi-Scale Structure Similarity Index
MSV	Maximally Stable Volumes
NIQE	Naturalness Image Quality Evaluator
NK	Normalized Cross-Correlation
NMR	Nuclear Magnetic Resonance
NN	Neural Network
NR-IQA	No-Reference Image Quality Assessment
NR-VQA	No-Reference Video Quality Assessment
NSS	Natural Scene Statistics
OIQA	Objective Image Quality Assessment
ORB	Oriented FAST and rotated BRIEF
OS-LTP	Orthogonal-Symmetric Local Ternary Pattern
PCA	Principal Component Analysis
PnP	Perspective-n-Point
QoS	Quality of Service
RANSAC	Random Sample Consensus
RR-IQA	Reduced-Reference Image Quality Assessment
RR-IQA	Reduced-Reference
SIFT	Scale Invariant Feature Transform
SLAM	Simultaneous Localization and Mapping
SSEQ	Spatial and Spectral entropies based Image Quality Assessment
SSIM	Structure Similarity Index
SURF	Speeded Up Robust Features
SVM	Support Vector Machine
SVR	SVM Regressor
VR	Virtual Reality
WN	White Noise distortion

List of Symbols/Notations

Notation	**Detail**
(x_n^{kp}, y_n^{kp})	n^{th} pixel coordinate obtained at k^{th} concentric circle or concentric ellipse at delta p
(x', y')	Transformed coordinates of a point in 2D space with coordinate position (x, y)
(x_n, y_n)	x and y coordinate of extracted keypoint k_i at n^{th} index
C_N	Coefficient Normalization factor
$D_j(x_n^{kp}, y_n^{kp})$	Gradient of image pixels at $Ɨ(x_n^{(k+1)p}, y_n^{(k+1)p}) - Ɨ\left(x_n^{(k-1)p}, y_n^{(k-1)p}\right)$ position
D_v	Decision variable used in curve tracking
$D_\theta(x_n^{kp}, y_n^{kp})$	Gradient of image pixels at $Ɨ\left(x_n^{k(p+1)}, y_n^{k(p+1)}\right) - Ɨ\left(x_n^{k(p-1)}, y_n^{k(p-1)}\right)$ position
E_D	Local Spectral Entropy
E_L	Spatial Entropy
E_c	Ellipse Eccentricity
I_d	Diagonal octant move
I_{it}	Image Sequence with frame i corresponding to threshold t
I_s	Square octant move
$L_k(a, b)$	Luminance comparison between image a and image b done at scale k
$M_{x,y}$	Spectral Probability Map
$\hat{P}(x, y)$	Processed Image for local mean removal and divisive normalization at x and y coordinate
j_k and n_k	Major and minor axis of the k^{th} concentric ellipse
k_i	Extracted Keypoint in an image at i^{th} index
r_k	Radius of the k^{th} concentric circle
ε_i	Residual Error
θ_{xmj}	Angle between x axis and ellipse major axis
μ_a and μ_b	Mean of image a and image b respectively
σ_a and σ_b	Variance of image a and image b respectively
σ_n	Scale of the extracted keypoint
Δf	Frame Difference
$Ɨ(x_n^{kp}, y_n^{kp})$	n^{th} interpolated pixel coordinate obtained at k^{th} concentric circle or concentric ellipse at delta p
R	Neural Network goodness of fit
$C(a, b)$	Contrast comparison between image a and image b
$C(x, y) = 0$	Equation of a general conic ($C(x, y) = Gx^2 + Hxy + Iy^2 + Jx + Ky + L = 0$)
DD	Distance between ellipse center and closest Directix point
DS	Distance between center of ellipse and either of the focus point
FD_i	Feature descriptor for i^{th} extracted keypoint

G, H, I, J, K and L	Six coefficients of general conic equation
HM	Second Moment Matrix used by Hessian-Affine detector
HaM	Second Moment Matrix used by Harris-Affine detector
$L(a, b)$	Luminance comparison between image a and image b
MJ	Semi-Major Axis Length
MN	Semi-Minor Axis Length
$S(a, b)$	Structure comparison between image a and image b
$T(n)$	Worst-case time complexity of an algorithm
$T(x, y)$	$x \times y$ DCT coefficient matrix
$Tilt(f_2, f_3)$	Transition tilt between two slanted views of an image scene (f_2) and (f_3)
$angle\left(x_n^{kp}, y_n^{kp}\right)$	Local gradient orientation for n^{th} interpolated pixel coordinate obtained at k^{th} concentric circle or concentric ellipse at delta p
dx	Horizontal Haar wavelet response
dy	Vertical Haar wavelet response
$m\left(x_n^{kp}, y_n^{kp}\right)$	Gradients magnitude for n^{th} interpolated pixel coordinate obtained at k^{th} concentric circle or concentric ellipse at delta p
r	Pearson Coefficient
$\mu(x, y)$	Mean value of an image
$\sigma(x, y)$	Contrast value of an image