

**PERCEPTUAL VISUAL INFORMATION
PROCESSING USING ROI BASED ENCODING
SCHEMES**

MEERA THAPAR KHANNA



DEPARTMENT OF ELECTRICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY DELHI

JUNE 2020

©Indian Institute of Technology Delhi (IITD), New Delhi, 2020

**PERCEPTUAL VISUAL INFORMATION
PROCESSING USING ROI BASED ENCODING
SCHEMES**

by

MEERA THAPAR KHANNA

DEPARTMENT OF ELECTRICAL ENGINEERING

Submitted

in fulfillment of the requirements of the degree of Doctor of Philosophy

to the



INDIAN INSTITUTE OF TECHNOLOGY DELHI

JUNE 2020

To my family

Certificate

This is to certify that the thesis titled **Perceptual Visual Information Processing using ROI based Encoding Schemes** being submitted by **Mrs. Meera Thapar Khanna** to the Department of Electrical Engineering, Indian Institute of Technology Delhi, for the award of **Doctor of Philosophy** is a record of bona-fide research work carried out by her under my guidance and supervision. In my opinion, the thesis has reached the standards fulfilling the requirements of the regulations relating to the degree. The work presented in this thesis has not been submitted elsewhere, either in part or full, for the award of any other degree or diploma.

Professor Santanu Chaudhury
Department of Electrical Engineering
Indian Institute of Technology Delhi
New Delhi - 110016, India.

Professor Brejesh Lall
Department of Electrical Engineering
Indian Institute of Technology Delhi
New Delhi - 110016, India.

Acknowledgements

First and foremost, I give thanks to Almighty God for providing me the strength, health and ability to undertake and complete this work.

I would like to express my sincere gratitude and heartily thanks to my supervisors Prof. Santanu Chaudhury and Dr. Brejesh Lall. This thesis would not have been possible without their guidance, critical reviews and encouragement throughout the work. Their intelligent ideas, motivation and in-depth/immense understanding helped me during all the times of research. I would like to thank Prof. Santanu Chaudhury for introducing me to this exciting field of research. His enthusiasm, scholarly inputs and integral view on research, has made a deep impression on me. It is a great opportunity for me to do my doctoral research under his guidance and to learn a lot from his research expertise. I owe lots of gratitude to him for having me shown this way of research. I also thank Dr. Brejesh Lall for his continuous support, cooperation, and for facilitating all the requirements, going out of his way. I express my heart-felt gratitude to him for always being there for me. I feel privileged to be associated with persons like them during my life.

I thank the members of my SRC committee: Prof. Arun Kumar, Dr. Sumantra Dutta Roy and Dr. Mukul Sarkar for their insightful suggestions and advice which encouraged me to widen my research from various perspectives. Nevertheless, I am also grateful to Dr Amrita Basu, Director, School of Cognitive Science, Jadavpur University, Kolkata for helping me to carry out a part of my research work at her institute. Her invaluable directions, technical support and discussions helped me a lot to complete part of my work. I sincerely admire the contribution of Protik Nath and Moumita Bhowmick students in her lab who were always ready to give their

help when required.

I am indebted to my fellow doctoral students for their valuable feedback, discussions, cooperation and of course a fun filled environment with their friendship. I sincerely appreciate the contribution of Chetan Ralekar, who helped me to conduct eye tracking experiments in the Neuro Computing lab. A special thanks to all the participants some of them my friends, fellow doctoral or postgraduate students who performed survey experiments for me and helped me to complete survey part of my work. I would like to thank the administrative staff of the Multimedia lab Usha Mam, Manrika, Deepti, Amit ji and Sushila ji for the timely logistic support.

I owe my deepest gratitude to my parent in laws and parents family for their love, care, moral support and sacrifices many a time during my work. A special thanks to my mother in law, Meena Khanna, who supported me in every possible way, even sometimes by giving up my family responsibilities to complete my work. I would like to express the deepest appreciation to my mother and father who raised me up with good values and supported me in all my pursuits.

At the end, I acknowledge the most important contribution of two persons during this course of work. A special thanks to my sweet little angel, my daughter Alena, for cheering me up always. She did not like to see me working at home, especially when I hold the laptop, but still allow me to work with the hope I will complete it soon and play a lot with her. I owe my sincere gratitude towards my better half Vikram for his unconditional love, care, eternal support and understanding of my goals and aspirations. His friendship, patience and sacrifice will remain my inspiration throughout my life. From the bottom of my heart, I say thanks to both of you for being with me always.

Meera Thapar Khanna

Abstract

Visual information processing is the ability to understand and interpret what we see. Humans generally pay attention to selected areas in an image as compared to the entire image. These areas are known as salient regions that catch human attention. Perceptual visual information processing can be defined as dealing with this perceptually important information in different applications such as ROI based compression, segmentation, enhancement etc. Modeling visual attention, specially saliency based attention has been a very active area of research over the past few years. A variety of different attention models are available now that have been used successfully in various applications in computer vision, mobile robotics, and cognitive systems.

In this dissertation, we explore the concept of visual saliency and develop a computationally efficient approach to estimate saliency in digital images; and presents different techniques for utilization of visual saliency in ROI based image and video coding applications. We also explore the concept of memorability in images and try to preserve memorable regions in ROI based coding. Following are the key contributions of the thesis:

- Saliency based video coding architecture.
- Defocus cue preserving image encoding.
- Behavioral study to validate the proposed encoding scheme.
- Saliency and defocus cue preserving encoding.
- Memorability preserving encoding.

We introduce a novel, macro-block level visual saliency guided video compression scheme. This model is divided into 2 steps viz. salient region detection and frame foveation. Visual saliency is modeled as a combination of low level and high-level features which become important at the higher-level visual cortex. A relevance vector machine is used to train over 3-dimensional feature maps representing global, local, and rarity measures of conspicuity, and provides probabilistic values to form the final saliency map. These saliency values are employed for non-uniform bit-allocation over frames. To achieve these goals, we propose a novel video compression architecture to incorporate saliency and save a huge amount of computation. This framework uses mutual information between consecutive frames for indicating frames requiring re-computation of saliency, and use motion vectors for propagating saliency values.

We propose a novel approach to preserve depth information in an image while ROI coding. We use defocus depth cue as a saliency measure to detect salient regions on depth information basis. We use frequency information to detect salient regions on the basis of blurring present in the images. Then, make changes in the existing JPEG compression architecture to incorporate these saliency values. There is a fixed quality value and quantization table for each block of an image in JPEG standard. We make the modification to alter the quality value for each block on the basis of defocus cue in that block.

We combine the saliency feature with the defocus information to use as a saliency measure for ROI based video compression. We make use of deep learning techniques CNN and RBM for predicting memorability in images and use generated memorability maps for ROI based image coding.

We do an intensive literature review and identify the current challenges at the onset of our research. We present optimistic solutions to some problems and also highlight future directions. We test the proposed ROI based encoding schemes for different databases including images and videos. During the course of this research, we conduct an eye-tracking experiment to validate the results obtained using proposed scheme. For subjective evaluation, a behavioral study is also conducted to prove the efficacy of our results. This study is conducted by a group of people. Experimental results are presented in a comprehensive and clear way within each chapter

providing the significance of each computational attention model. Finally, the discussion is opened on the basis of the theoretical as well as practical achievements, and possible future extensions are proposed.

सार

दृश्य सूचना प्रसंस्करण समझने और व्याख्या करने की वह क्षमता है जो हम देखते हैं। मनुष्य आम तौर पर पूरी छवि की तुलना में छवि में कुछ चयनित क्षेत्रों पर ध्यान देते हैं। इन क्षेत्रों को उन प्रमुख क्षेत्रों के रूप में जाना जाता है जो मानव ध्यान आकर्षित करते हैं। अवधारणात्मक दृश्य सूचना प्रसंस्करण को इस अवधारणात्मक महत्वपूर्ण जानकारी से निपटने वाली विभिन्न अनुप्रयोगों जैसे आर ओ आई आधारित कंप्रेशन, विभाजन, एनहॅन्समेंट आदि के रूप में परिभाषित किया जा सकता है। पिछले कुछ वर्षों से, दृश्य ध्यान मोडलिंग, विशेष रूप से सेलियेन्सी आधारित ध्यान, अनुसंधान का एक बहुत सक्रिय क्षेत्र रहा है। आजकल विभिन्न प्रकार के ध्यान मॉडल उपलब्ध हैं, जिनका उपयोग विभिन्न अनुप्रयोगों जैसे कंप्यूटर दृष्टि, मोबाइल रोबोटिक्स और कॉग्निटिव प्रणालियों में सफलतापूर्वक किया गया है।

इस शोध प्रबंध में, हम दृश्य सेलियेन्सी की अवधारणा का पता लगाते हैं और डिजिटल छवियों में सेलियेन्सी का अनुमान लगाने के लिए एक कुशल गणना का दृष्टिकोण विकसित करते हैं; और आर ओ आई आधारित छवि और वीडियो कोडिंग अनुप्रयोगों में दृश्य सेलियेन्सी का उपयोग करने के लिए विभिन्न तकनीकों को प्रस्तुत करते हैं। हम छवियों में अविस्मरणीय की अवधारणा का पता लगाने और आर ओ आई आधारित कोडिंग में अविस्मरणीय क्षेत्रों को संरक्षित करने का प्रयास भी करते हैं। शोध पत्र के प्रमुख योगदान निम्नलिखित हैं:

- सेलियेन्सी आधारित वीडियो कोडिंग आर्किटेक्चर।
- डिफोकस क्यू संरक्षित छवि एन्कोडिंग।
- प्रस्तावित एन्कोडिंग योजना को मान्य करने के लिए व्यवहारिक अध्ययन।
- सेलियेन्सी और डिफोकस क्यू संरक्षित एन्कोडिंग।
- यादगार को संरक्षित करने वाली एन्कोडिंग।

हम एक नई, मैक्रो-ब्लॉक स्तर दृश्य सेलियेन्सी निर्देशित वीडियो कंप्रेशन योजना पेश करते हैं। यह नमूना दो चरणों में विभाजित है: मुख्य क्षेत्र का पता लगाना और फ्रेम फोविएशन। दृश्य सेलियेन्सी को निम्न स्तर और उच्च-स्तरीय विशेषताओं के संयोजन के रूप में तैयार किया जाता है, जो उच्च-स्तरीय दृश्य प्रांतस्था में महत्वपूर्ण हो जाते हैं। एक प्रासंगिक वेक्टर मशीन का उपयोग 3-आयामी विशेषता नक्शों, जो वैश्विक, स्थानीय और दुर्लभता के उपायों का प्रतिनिधित्व करते हैं, को प्रशिक्षित करने के लिए और अंतिम सेलियेन्सी मानचित्र बनाने के लिए संभाव्य मान प्रदान करने के लिए किया जाता है। ये सेलियेन्सी मूल्य फ्रेम पर गैर-समान बिट-आवंटन के लिए कार्यरत हैं। इन लक्ष्यों को प्राप्त करने के लिए, हम एक नई वीडियो कंप्रेशन आर्किटेक्चर का प्रस्ताव करते हैं, जो सेलियेन्सी को शामिल करता है और संगणना की एक बड़ी प्रक्रिया से बचाता है। यह बनावट उन फ्रेम्स को इंगित करने के लिए लगातार फ्रेम्स के बीच पारस्परिक जानकारी का उपयोग करता है जिन को सेलियेन्सी की पुनः गणना आवश्यक है, और सेलियेन्सी के मूल्यों के प्रसार के लिए गति वेक्टर का उपयोग करता है।

हम आर ओ आई कोडिंग करते समय एक छवि में गहराई से जानकारी को संरक्षित करने के लिए एक नया दृष्टिकोण प्रस्तावित करते हैं। हम गहराई से जानकारी के आधार पर मुख्य क्षेत्रों का पता लगाने के लिए डीफोकस डेप्थ क्यू का उपयोग सेलियेन्सी के रूप में करते हैं। हम छवियों में धुंधलापन मौजूद होने के आधार पर आवृत्ति जानकारी का उपयोग मुख्य क्षेत्रों का पता लगाने के लिए करते हैं। फिर, सेलियेन्सी मूल्यों को शामिल करने के लिए मौजूदा जे पी ई जी कंप्रेशन आर्किटेक्चर में परिवर्तन करते हैं। जे पी ई जी मानक में छवि के प्रत्येक ब्लॉक के लिए एक निश्चित गुणवत्ता मूल्य और परिमाणीकरण तालिका है। हम प्रत्येक ब्लॉक के लिए गुणवत्ता मूल्य को बदलने के लिए उस ब्लॉक में डिफोकस क्यू के आधार पर संशोधन करते हैं।

हम एक सेलियेन्सी माप को आर ओ आई आधारित वीडियो कंप्रेशन में उपयोग करने के लिए डिफोकस जानकारी के साथ सेलियेन्सी माप को जोड़ते हैं। हम गहन शिक्षण तकनीकों सी एन एन और आर बी एम का उपयोग छवियों में अविस्मरणीय की भविष्यवाणी के लिए करते हैं और आर ओ आई आधारित छवि कोडिंग के लिए उत्पन्न यादगार नक्शों का उपयोग करते हैं।

हम एक गहन साहित्य समीक्षा करते हैं और अपने शोध में वर्तमान चुनौतियों की पहचान करते हैं। हम कुछ समस्याओं के लिए आशावादी समाधान प्रस्तुत करते हैं और भविष्य की दिशाओं पर भी प्रकाश डालते हैं। हम प्रस्तावित आर ओ आई आधारित एन्कोडिंग योजनाओं का परीक्षण छवियों और वीडियो सहित विभिन्न डेटाबेस के लिए

करते हैं। इस शोध के दौरान, हम प्रस्तावित योजना का उपयोग करके प्राप्त परिणाम को सत्यापित करने के लिए एक लक्ष्यानुसरण प्रयोग करते हैं। आत्मगत मूल्यांकन के लिए, एक व्यवहारिक अध्ययन भी हमारे परिणामों की प्रभावकारिता साबित करने के लिए आयोजित किया गया है। यह अध्ययन लोगों के एक समूह द्वारा किया जाता है। प्रत्येक कम्प्यूटेशनल अवधान मॉडल का महत्व प्रदान करने के लिए प्रायोगिक परिणाम प्रत्येक अध्याय के भीतर एक व्यापक और स्पष्ट तरीके से प्रस्तुत किये गए हैं। अंत में, सैद्धांतिक और साथ ही साथ व्यावहारिक उपलब्धियों के आधार पर चर्चा खुली है, और संभावित भविष्य प्रसार प्रस्तावित हैं।

Contents

Certificate	iii
Acknowledgements	v
Abstract	vii
List of Figures	vii
List of Tables	xiii
<i>Acronyms</i>	xv
1 Introduction	1
1.1 Scope and Objectives	2
1.2 Overview of the Thesis	3
1.3 Database Used	6
1.3.1 MSRA Database	6
1.3.2 Video Sequences	7
1.3.3 Blur Detection Database	8
1.3.4 Raise Database	9
1.3.5 10K US Face Database	10
1.3.6 Image Memorability Database	11
1.3.7 Object Memorability Database	12
1.3.8 Caltech Faces Database	12

1.3.9	Stanford Dogs Database	13
1.4	Layout of Report	14
2	Perceptual Video Coding: A Review	15
2.1	Visual Saliency Models	16
2.1.1	Transform Domain Approaches	16
2.1.2	Pixel Domain Approaches	19
2.2	ROI based Encoding	22
2.2.1	Pre-Processing	22
2.2.2	Embedded Encoding Schemes	23
2.3	Depth Perception	26
2.4	Memorability of Images	29
2.4.1	Image Memorability	29
2.4.2	Memorability and Visual Attention	30
2.5	Quality	31
2.5.1	Objective Assessment	31
2.5.2	Subjective Assessment	32
2.6	Motivation for the Current Work	33
3	Saliency Preserving Video Encoding Scheme	37
3.1	Human Visual System	37
3.1.1	Eye: Photoreceptor Cells	37
3.1.2	Visual Cortex	38
3.1.3	Attention Mechanisms	38
3.1.4	Saliency Map	39
3.2	Saliency Algorithm	40
3.2.1	Global Conspicuity	41
3.2.2	Local Conspicuity	43
3.2.3	Rarity Conspicuity	45

3.2.4	Integration of the Feature Maps	47
3.2.5	Results	49
3.3	Video Compression Architecture	52
3.3.1	Re-calculation of Saliency Values	52
3.3.2	Propagation of Saliency Values	55
3.3.3	Selection of Quantization Values	56
3.4	Experimental Results	58
3.4.1	Gaussian Blurring Method	59
3.4.2	QP Tuning Method	61
3.4.3	Discussions	63
3.5	Conclusion	67
4	Depth Preserving Image Encoding	69
4.1	Depth Perception	69
4.1.1	Depth Cues	71
4.2	Saliency Algorithm	74
4.2.1	Selection of Cue	74
4.2.2	Formulation	75
4.3	Encoding Scheme	76
4.4	Results and Discussions	78
4.5	Conclusion	81
5	Behavioral Study of Defocus Cue Preserving Image Encoding	83
5.1	Introduction	83
5.2	Proposed Behavioral Study	84
5.2.1	Behavioral Study Algorithm	85
5.3	Behavioral Study Analysis	86
5.3.1	One Way ANOVA Analysis	86
5.3.2	Two Way ANOVA Analysis	91

5.4	Conclusion	97
6	Combining Saliency and Depth Cue for Video Encoding	99
6.1	Non-Linear fusion Scheme	100
6.2	Results	102
6.3	Conclusion	111
7	Memorability Preserving Image Encoding	113
7.1	Introduction	113
7.2	Image Memorability	117
7.2.1	Experiment Setup	117
7.2.2	Analysis	119
7.3	Memorability Prediction	122
7.3.1	Restricted Boltzmann Machine	123
7.3.2	CNN	126
7.3.3	Fusion of RBM and CNN	127
7.3.4	Memorability Maps	128
7.4	Memorability Preserving Compression	130
7.4.1	Compression Scheme	130
7.5	Experiments	131
7.5.1	Database	131
7.5.2	Prediction Results	132
7.5.3	Memorability based Compression Results	134
7.6	Conclusions	139
8	Conclusion and Future Directions	141
8.1	Summary of Thesis	141
8.2	Our Contributions	143
8.3	Future Directions	144

Bibliography**147****Publications****164****Biography****166**

List of Figures

1.1	Sample images from the MSRA Salient Object Detection Database	7
1.2	Sample frames from different Video Sequences	8
1.3	Sample images from the Blur Detection Database	9
1.4	Sample images from the RAISE Database	10
1.5	Sample images from the 10K US Face Database	11
1.6	Sample images from the Image Memorability Database	11
1.7	Sample images from the Object Database	12
1.8	Sample images from the Caltech Faces Database	13
1.9	Sample images from the Stanford Dogs Database	13
2.1	Computation of crisply perceived radial region 'r'	16
2.2	Failure to detect large salient objects	18
2.3	Illustration of Saliency Reversal	19
3.1	The human eye and density of photo-receptor cells on the retina	38
3.2	Examples of (a) bottom-up and (b) top-down attention mechanisms	39
3.3	Saliency not only depends on immediate context with respect to a feature but also on how unique of a response is elicited by that feature in the image	41
3.4	Feature map representing global conspicuity	42
3.5	Illustration of regions in concentric ring	43
3.6	(a) Image, (b) SIFT Key points, (c) Feature map by Local Conspicuity	45

3.7	Illustration of original and blurred images after applying PCT algorithm. Represents that the blurred image corresponding feature map is sparser	47
3.8	Macroblock level ground truth preparation	48
3.9	ROC curves by proposed method and [61] by varying threshold on saliency map values	50
3.10	(a) Original image, Feature maps by (b) Global, (c) Local [67] and (d) Rarity [154] methods, (e) Proposed saliency map, (f) Saliency map generated by [61] and (g) [24]	50
3.11	More results. Original image, ground truth, proposed method saliency map in each set from left to right	51
3.12	Illustration of consistency between consecutive video frames: top row saliency maps and bottom row encoded results of ESPN Sequence	52
3.13	Video compression framework that incorporate saliency propagation	53
3.14	(a) Mutual Information plot for saliency maps, (b) Mutual Information plots for representing RGB and saliency overlap. Top: Airtel ad sequence having 9 scene cuts. Bottom: ESPN news sequence having 4 scene cuts	55
3.15	Image to represent weighted average of saliency values, different colors indicate the amount of overlap and thus weights	56
3.16	12.5% increase in Q_{step} per QP increment or doubling of Q_{step} for every 6 th increment of QP	57
3.17	Image to illustrate QPs for 16×16 MBs at QP = 24	62
3.18	Coded results for Dance Sequence	65
3.19	Coded results for ESPN Sequence	66
3.20	Coded results for Claire Sequence	66
4.1	The foreground frames the planes in the back. A popular composition device to add depth.	70

4.2	Giotto, Exorcism of the Demons at Arezzo (1299), fresco, 270 times 230 cm, San Francesco, Upper Church, Assisi, Italy. WikiArt.	71
4.3	Illustration of different depth cues present in an image.	74
4.4	Top row represents Original images, and bottom row corresponding Saliency Maps.	76
4.5	Compression results using standard JPEG and Proposed compression schemes.	81
5.1	Box plot for representing the response for the variable image type.	88
5.2	(a) Input Image, Compression results by (b) Standard JPEG scheme (c) Defocus cue based scheme at quality parameter 3, Compression results using (d) Standard JPEG scheme (e) Defocus cue based scheme at quality parameter 5.	89
5.3	Plot to show least square mean for three image types.	91
5.4	Interaction plot representing responses for two variables image quality and image type split by image quality.	94
5.5	Interaction plot representing responses for two variables image quality and image type split by image type.	95
6.1	ROC curves obtained by proposed method and by using [61] and [24] methods.	101
6.2	Image to show saliency map results. Original image, GBVS saliency map [61], AIM saliency map [24], Pure saliency map [59], Defocus saliency map [80], Fused Saliency map (top to bottom).	101
6.3	Rate and distortion comparison between different compression schemes for 2 videos.	105
6.4	(a) Original image, (b) Proposed method saliency map, Compression results using (c) H.264 method, (d) Pure saliency method [80], (e) Defocus saliency method [59], and (f) Proposed fusion saliency method.	107
6.5	(a) Original image, (b) Proposed method saliency map, Compression results using (c) H.264 method, (d) Pure saliency method [80], (e) Defocus saliency method [59], and (f) Proposed fusion saliency method.	108

6.6	(a) Original image, (b) Depth map of original image, Depth map obtained after (c) H.264 method, (d) Pure saliency method [59], (e) Defocus saliency method [80] and, (f) Fusion saliency method.	109
6.7	(a) Original image, (b) Depth map of original image, Depth map obtained after (c) H.264 method, (d) Pure saliency method [59], (e) Defocus saliency method [80] and, (f) Fusion saliency method.	110
7.1	Example memorability and saliency maps, (a) Original Image (b) Fixation Map (c) Memorability Map (d) Saliency Map.	115
7.2	Flow Chart of proposed memorability preserving encoding scheme.	116
7.3	An example eyetracking experimental sequence. It include the presentation of images, the collection of key press after image presentation, and the response of questions.	119
7.4	Fixations and saccades. Each circle denotes the fixation and number inside the circle indicates the order of fixation. The radius of the circle increases with the increase in fixation duration [120].	120
7.5	(a) Original Image, (b) Heat Map, (c) Fixation Points. Each circle denotes the fixation and the radius of the circle increases with the increase in fixation duration.	121
7.6	Illustration of regions in an image.	122
7.7	Flow Chart of proposed memorability prediction scheme.	123
7.8	RBM structure with m visible and n hidden units.	124
7.9	Illustration of the RBM features.	125
7.10	VGG16 Architecture.	127
7.11	Visualization of some of the sample memorability maps form the database. Top row represent Original images, second row represent the ground truth memorability maps, third row represent the fixation points, and last row represent memorability maps generated by the proposed scheme.	129

7.12	Visualization of some of the sample memorability maps of the compressed images. Red color shows the object or regions of high memorability and blue shows least memorable object or region in an image. Different color levels shows the magnitude of contribution.	135
7.13	Rate and distortion comparison between different compression schemes for 2 videos.	135
7.14	(a) Original Image, Reconstructed Images after (b)JPEG Compression, (c) Saliency based Compression, and (d) Memorability based Compression.	137
7.15	Illustration of compression results using different scheme in detail.	138

List of Tables

3.1	Video Sequences	60
3.2	Compression results using Gaussian blurring and QP tuning method for QP = 24	60
3.3	Compression results using Gaussian blurring and QP tuning method for QP = 28	60
3.4	Compression results using Gaussian blurring and QP tuning method for QP = 30	60
3.5	Compression results using Gaussian blurring and QP tuning method for QP = 32	61
3.6	MOS and Standard Deviation (SD) Table	65
4.1	Compression results using standard JPEG and Modified JPEG method	79
4.2	Compression results using standard JPEG and Modified JPEG method	79
4.3	Compression results using standard JPEG and Modified JPEG method	80
4.4	Compression results using standard JPEG and Modified JPEG method	80
5.1	Summary Table for one way ANOVA	87
5.2	Summary table for least square mean of different image types	89
5.3	Pairwise differences of least square mean of different image types	90
5.4	Summary Table for two way ANOVA	93
5.5	Least square mean of different quality parameters from H to A (high to low) . .	95
5.6	Pairwise differences of lsmean of different quality parameters from H to A (high to low)	96
5.7	Pairwise differences of lsmean of different Image Types	96
5.8	Pairwise differences of lsmean of different Image Types and Quality	97

6.1	Video Sequences	102
6.2	Comparison results of File Size using standard H.264, Pure saliency, Defocus saliency and Proposed method	103
6.3	Comparison results of PSNR using standard H.264, Pure saliency, Defocus saliency and Proposed method	103
6.4	Comparison results of RMSE (Depth Reconstruction) using H.264 method, Pure saliency method, Defocus saliency method and Proposed method	104
6.5	MOS and Standard Deviation (SD) values using different encoding methods	106
7.1	Eyetracking Features	120
7.2	Fixation duration and Memorability score of images	122
7.3	Results on 10K US Face Dataset	133
7.4	Results on Object Dataset	134
7.5	Results on Stanford Dogs and Caltech Face Dataset	134
7.6	PSNR and SSIM Results	136

Acronyms

JPEG	Joint Photographic Experts Group
HD	High Definition
HVS	Human Visual System
PSNR	Peak Signal To Noise Ratio
SSIM	Structural Similarity Index
RMSE	Root Mean Squared Error
ROI	Region Of Interest
MI	Mutual Information
MB	Macroblock
QP	Quantization Parameter
ANOVA	Analysis Of Variance
RBM	Restricted Boltzmann Machine
CNN	Convolutional Neural Network
SVR	Support Vector Regression
RAISE	Raw Image Dataset
LCD	Liquid Crystal Display
SR	Spectral Residual
RGB	Red Green Blue
ICA	Independent Component Analysis
HMM	Hidden Markov Models
CRF	Conditional Random Fields
RC	Rate Control
FMO	Flexible Macroblock Ordering
SNR	Signal To Noise Ratio
GMM	Gaussian Mixture Model
SIFT	Scale Invariant Feature Transform
DCT	Discrete Cosine Transform
PCT	Pulse DCT
SVM	Support Vector Machine
ROC	Receiver Operating Characteristics
AUC	Area Under The Curve
NROI	Non Region Of Interest
RDO	Rate Distortion Optimization
MOS	Mean Opinion Score
SD	Standard Deviation
CAVLC	Context Adaptive Variable Length Coding
QCIF	Quarter Common Intermediate Format
XGA	Extended Graphics Array
DoG	Difference Of Gaussian
FA	False Alarm