PERCEPTUAL VISUAL INFORMATION PROCESSING USING ROI BASED ENCODING SCHEMES

MEERA THAPAR KHANNA

DEPARTMENT OF ELECTRICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY DELHI

JUNE 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, co-operation 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, एनहॅन्स्मेंट 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 है। 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 है। अंत में, सैद्धांतिक और साथ ही साथ व्यावहारिक उपलब्धियों के आधार पर चर्चा खुली है, और संभावित भविष्य प्रसार प्रस्तावित हैं।

Contents

List of Figures

List of Tables

Acronyms

JPEG	**J**oint **P**hotographic **E**xperts **G**roup
HD	**H**igh **D**efinition
HVS	**H**uman **V**isual **S**ystem
PSNR	**P**eak **S**ignal **T**o **N**oise **R**atio
SSIM	**S**tructural **S**imilarity **I**ndex
RMSE	**R**oot **M**ean **S**quared **E**rror
ROI	**R**egion **O**f **I**nterest
MI	**M**utual **I**nformation
MB	**M**acroblock
QP	**Q**uantization **P**arameter
ANOVA	**An**alysis **O**f **V**ariance
RBM	**R**estricted **B**oltzmann **M**achine
CNN	**C**onvolutional **N**eural **N**etwork
SVR	**S**upport **V**ector **R**egression
RAISE	**Ra**w **I**mage **D**ataset
LCD	**L**iquid **C**rystal **D**isplay
SR	**S**pectral **R**esidual
RGB	**R**ed **G**reen **B**lue
ICA	**I**ndependent **C**omponent **A**nalysis
HMM	**H**idden **M**arkov **M**odels
CRF	**C**onditional **R**andom **F**ields
RC	**R**ate **C**ontrol
FMO	**F**lexible **M**acroblock **O**rdering
SNR	**S**ignal **T**o **N**oise **R**atio
GMM	**G**aussian **M**ixture **M**odel
SIFT	**S**cale **I**nvariant **F**eature **T**ransform
DCT	**D**iscrete **C**osine **T**ransform
PCT	**P**ulse **DCT**
SVM	**S**upport **V**ector **M**achine
ROC	**R**eceiver **O**perating **C**haracteristics
AUC	**A**rea **U**nder **T**he **C**urve
NROI	**N**on **R**egion **O**f **I**nterest
RDO	**R**ate **D**istortion **O**ptimization
MOS	**M**ean **O**pinion **S**core
SD	**S**tandard **D**eviation
CAVLC	**C**ontext **A**daptive **V**ariable **L**ength **C**oding
QCIF	**Q**uarter **C**ommon **I**ntermediate **F**ormat
XGA	**E**xtended **G**raphics **A**rray
DoG	**D**ifference **O**f **G**aussian
FA	**F**alse **A**larm