SALIENCY DETECTION IN IMAGES AND VIDEOS

ADITI KAPOOR

AMAR NATH AND SHASHI KHOSLA SCHOOL OF INFORMATION TECHNOLOGY

INDIAN INSTITUTE OF TECHNOLOGY DELHI

January 2017

SALIENCY DETECTION IN IMAGES AND VIDEOS

by

ADITI KAPOOR

AMAR NATH AND SHASHI KHOSLA SCHOOL OF INFORMATION TECHNOLOGY

Submitted

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

to the

Indian Institute of Technology Delhi

January 2017

Certificate

This is to certify that the thesis titled **Saliency Detection in Images and Videos** being submitted by **Ms. Aditi Kapoor** for the award of **Doctor of Philosophy** in **Amar Nath and Shashi Khosla School of Information Technology** is a record of bona-fide work carried out by her under our guidance and supervision at the **Amar Nath and Shashi Khosla School of Information Technology, Indian Institute of Technology Delhi**. To the best of our knowledge, 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

K. K. Biswas
Professor
Deptt. of Computer Science & Engineering
Indian Institute of Technology Delhi.

M. Hanmandlu
Professor
Deptt. of Electrical Engineering
Indian Institute of Technology Delhi.

Acknowledgments

The experience of research at IIT Delhi for this degree has been an enlightening experience. I am indebted to several people who contributed directly or indirectly during the last few years on my way to realizing this thesis. I take this opportunity to thank as many of them as I can. First of all, I wish to express my deepest gratitude to my supervisors Prof. K. K. Biswas and Prof. M. Hanmandlu for their enlightening guidance and constant encouragement during the development and completion of this dissertation. I will always be thankful for their continuous support and their encouragement for allowing me the freedom to explore my interests, with understanding and patience. I thank them for the numerous revisions and suggestions and for setting the highest of standards for any kind of submission. I am forever grateful to them for always taking the time to teach me the basics and encouraging me to attend meets/workshops that aided in the development of ideas and a better grip on the fundamentals. It has been a privilege to work under their guidance.

I would also like to thank my research committee members Prof. Subhashis Banerjee, Prof. Prem Kalra and Prof. Brejesh Lall for their valuable guidance and support. Their words of encouragement helped in maintaining the required levels of motivation. Further I would like to thank Prof. S.K. Gupta and Prof. Saroj Kaushik for their constant encouragement and help. I thank Prof. Parag Singla for his valuable help and guidance especially in Markov Logic Networks. I would also like to thank all faculty members of Computer Science and Engineering and School of Information Technology especially Prof. Sanjiva Prasad, Prof. S.N. Maheshwari, Prof. Huzur Saran, Prof. M. Balakrishnan, Prof. Amitabha Bagchi, Prof. Sorav Bansal and Prof. Sumantra Dutta Roy who have encouraged me with their interactions during these years.

I am grateful to my friends and colleagues Parul Shukla and Sonia Khetarpaul who enriched my life during the PhD program with their discussions, encouragement and support. Our discussions along with numerous cups of tea throughout these years provided great companionship and helped in maintaining the required energy and enthusiasm. I would further like to thank Nisha Jain, Swati Sharma, Sunita Tiwari, Priti Jagwani and Mona Jain for their companionship. I would also like to thank the different M.Tech and summer intern students with whom I worked during my PhD. I would like to thank Yamuna Shukla and Manish Agarwal for their valuable discussions. I would like to extend my thanks to all the staff members of both School of Information Technology and Computer Science and Engineering especially Mr. Rajesh Kumar

for his continuous help in tackling all administrative processing. I would also like to thank the staff of Database and AI Lab, Mr. M. Rathinam and Mr. S.S. Negi for taking care of all lab related technicalities. I would also like to thank the team of IIT Delhi involved in maintaining this institute to be such a peaceful, pleasing and research-oriented institute.

Finally I would like to thank my family. I thank my parents who started this dream by giving me the freedom and encouragement at every point of my life with their unconditional love and support. I would also like to thank my parents-in laws who sustained this dream with their continuous support and encouragement. I would like to thank my husband and best friend Ankur who has always encouraged me to keep dreaming and set high goals and has helped me in everyway possible throughout this time. From his constant discussions about my projects to his unwavering help, to his endeavor of encouraging me to keep trying whenever I got discouraged, his support has sustained me throughout this time. Finally I would like to dedicate this thesis to my son Adhyayan for his joyful presence and immense patience.

Aditi Kapoor

Abstract

How humans perceive a certain image can vary greatly and depends on specific image or video characteristics. Our brains do not register everything that is presented to our visual field. Although a number of objects could be visible to the human eye at any point in time, the attention gets focused on a particular object or a group of objects which are more conspicuous by virtue of their contrast with the surrounding. The human brain perceives a region to be salient based on a number of attributes like color contrast with the boundary colors, relative intensity, relative size with respect to other color patches, its location within an image. However since these features vary widely across range of images, no crisp thresholds can be specified for an automatic salient region detector. In this thesis we address this issue by using soft computing approaches. Such regions of the image or a video are referred to as salient regions.

We start with developing a fuzzy rule based approach which makes use of fuzzy attributes mentioned above for salient region detection in images. As a first step, the boundary colors are quantified into four main categories. The rest of the image is divided in rectangular patches, and each patch is marked to be salient or non-salient. Salient patches are clubbed together to form the salient region. The fuzzy rules themselves are learned from a large diverse collection of images with marked salient regions, using Genetic Algorithm based evolutionary model. A case based reasoning approach has also proposed on similar lines. The second approach for saliency detection in images is based on information sets which utilize the information content of the images in terms of texture and color. The entropy of image patches are used to create saliency maps. We then integrate this information with color contrast maps for the detection of salient regions. We establish the effectiveness of both of our approaches by extensive comparisons with the state-of-the-art methods in terms of precision, recall and F- Measure on three publicly available datasets.

Next we consider the problem of saliency detection in videos. Human attention is directed to

any significant change in content in specific frames (termed salient frames or key frames). For detection of salient frames we propose two models. The first model is a fuzzy rule based one, based on fuzzified histogram distances and pixel intensity distances between selected frames. For the second model we make use of the fact that instead of scanning the whole of the frame, content change may be noticed even in specified strips along various directions. We create artificial images by collating strips from successive frames, and look for sharp changes in color components. We establish the effectiveness of these models using fidelity and compression ratio measures and comparing with other reported algorithms.

On a similar vein, any human activity which is in variance from the usual set attracts human attention. We have explored this aspect in our thesis, by proposing a Markov Logic Network based framework. Given a large set of activities we train the system so that it is able to recognize normal pattern of activities. The system automatically assigns low weights for action patterns that are significantly in variance with the usual ones. We test these models on our own created database. We have additionally performed experiments on action recognition and unusual activity detection for the THUMOS 2014 challenge.

Finally we present some applications of our saliency methods in images and videos through object discovery in videos, image retargeting, non-photorealistic rendering and image compression.

Contents

List of Figures

List of Tables