

SALIENCY DETECTION IN IMAGES AND VIDEOS

ADITI KAPOOR



**AMAR NATH AND SHASHI KHOSLA SCHOOL OF INFORMATION
TECHNOLOGY**

INDIAN INSTITUTE OF TECHNOLOGY DELHI

January 2017

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

Abstract	i
List of Figures	v
List of Tables	xi
1 Introduction	1
1.1 Salient Object Detection in images	1
1.1.1 Center Surround Based Methods	2
1.1.2 CRF based method	3
1.1.3 Bayesian methods	4
1.1.4 Super-pixel based methods	4
1.1.5 Frequency based Methods	5
1.1.6 Saliency as a fuzzy event	8
1.1.7 Context-aware saliency	10
1.2 Salient Object Detection in videos	11
1.2.1 Shot based Method	11
1.2.2 MSER based semi-supervised method	12
1.2.3 Frequency based video summarization	13
1.2.4 Saliency for Surveillance	14
1.2.5 Spatiotemporal method for video saliency	15
1.2.6 Salient object detection for a video database system	15
1.3 Some applications of saliency	16
1.4 Issues and Directions of the Thesis	18
1.5 Contributions of the Thesis	19

2	Fuzzy color rule based approach for saliency detection	21
2.1	Overview	21
2.2	CIELAB quantization	23
2.2.1	Fuzzy color quantization	24
2.3	Background detection	27
2.4	Color based Features	27
2.4.1	Color Spread	29
2.4.2	Color Proximity	30
2.4.3	Features for black and white backgrounds	32
2.5	Fuzzy rules for saliency detection	33
2.5.1	Selection of salient color	33
2.5.2	Selection of salient components	35
2.5.3	Human face as an additional feature	35
2.6	Learning the Fuzzy Rules	37
2.6.1	Case Based Reasoning approach	38
2.6.1.1	Case Based learning	39
2.6.2	Genetic algorithm based learning	40
2.7	Case Study	43
2.7.1	Evaluation on MSRA dataset	44
2.7.2	Evaluation on Berkley-300 database	45
2.7.3	Evaluation on ECSSD database	47
2.7.4	Qualitative analysis	48
2.8	Summary of the chapter	54
3	Information Set based approach for saliency detection	55
3.1	Overview of Information Sets	55
3.1.1	Hanman-Anirban entropy to Shannon entropy	57
3.2	Information Sets for saliency	58
3.3	Information set features in CIELab Colorspace	60
3.4	Information set based features	60
3.4.1	Effective Information Features	61
3.4.2	Sigmoid Features	62

3.4.3	Energy Features	63
3.4.4	Multiquadratic Features	64
3.4.5	Analysis of features	65
3.5	Contrast map computation	65
3.5.1	Saliency cut computation	67
3.6	Color based clustering	68
3.6.1	Saliency maps	73
3.7	Combining color clustered and information based images	73
3.7.1	An alternate approach: Integrating information set based features with color clustering	74
3.8	Case Study	74
3.8.1	Evaluation on MSRA dataset	75
3.8.2	Evaluation on Berkley-300 database	77
3.8.3	Evaluation on ECSSD database	79
3.9	Summary of the chapter	82
4	Salient frame detection in videos	83
4.1	Overview	83
4.2	Fuzzy Rule based Method	85
4.2.1	Fuzzy segment distance	86
4.2.2	Histogram distance	89
4.2.3	Merging Fuzzy segment distance and histogram distance	90
4.3	Strip based salient frame detection	91
4.4	Case Study	94
4.5	Summary of the chapter	97
5	Salient activity detection in videos	101
5.1	Overview	101
5.2	Markov Logic Network Based Action Recognition	102
5.2.1	Markov Logic Networks	102
5.2.2	Markov Logic Networks in longer activity and unusual detection	107
5.2.2.1	Preprocessing	109

5.2.2.2	Fuzzy classification of subactions	109
5.2.2.3	Markov Logic Network based classification	112
5.2.2.4	Case Study	115
5.2.3	Markov Logic Networks for small activity detection	119
5.2.3.1	Dataset creation	119
5.2.3.2	Training based on MLN	120
5.3	Large scale action recognition and unusual action detection: THUMOS 2014 challenge	122
5.3.1	Dataset	123
5.3.2	Task 1: Action Recognition	123
5.3.3	Task 2: Temporal Action and Unusual Detection	125
5.3.4	Result Analysis	126
5.4	Summary of the chapter	127
6	Saliency Applications	129
6.1	Introduction	129
6.1.1	Salient object discovery in videos	130
6.1.2	Perception based image compression	133
6.1.3	Image retargeting	136
6.1.4	Saliency based non-photorealistic rendering	139
6.2	Summary of the chapter	146
7	Conclusions and Future Work	147
7.1	Research contributions	147
7.2	Future Research Directions	150
	Bibliography	153
	List of Publications	167

List of Figures

2.1	CIELab color quantization chart	22
2.2	Circular color quantization	24
2.3	Fuzzification of angular zones	26
2.4	Example of fuzzy colors membership functions	26
2.5	CIELab quantization for angular zones z6 and z7	26
2.6	Membership functions for the percentage area of the boundary color	28
2.7	Example of background color detection: (a),(c) Original image; (b),(d) Back-ground colors highlighted	28
2.8	Membership functions for color spread	30
2.9	Membership functions for color proximity	32
2.10	Membership functions for RCCS	36
2.11	Examples of the connected components (Yellow rectangles indicate our results for the salient object and the red rectangles indicate other objects present in the image)	36
2.12	Comparison with other approaches in terms of precision, recall and F-Measure for MSR dataset	46
2.13	Comparison with other approaches in terms of precision, recall and F-Measure for MSR dataset	47
2.14	Visual comparison for MSR dataset with other approaches (a)GC [10] (b)MC [88] (c)MNP [7] (d)PCA [6] (e)HS [69] (f)LMLC [79] (g)CA [31] (h)HC [74] (i)RC [4] (j)LC [71] (k) Our GA approach (l)Ground Truth	48
2.15	Comparison with other approaches in terms of precision, recall and F-Measure for Berkley dataset	49

2.16	Visual comparison with other approaches for Berkley dataset (a)CA (b)GS_GD (c)GS_SP (D)CRF (e)Our GA approach (f)Ground Truth	50
2.17	F-Measure based comparisons for ECSSD dataset	50
2.18	Some GA based results (a)Original image with rectangle around our object detection result (b)Ground truth in the form of binary cut with rectangle around salient object	51
2.19	Complex cases of saliency detection: Variation in object sizes	52
2.20	Complex cases of saliency detection: non-centered objects	52
2.21	Complex cases of saliency detection: Human Faces	52
2.22	Complex cases of saliency detection: Multiple objects and scattered objects	53
2.23	Complex cases of saliency detection: luminance based results and multi-colored objects	53
3.1	Average L, a, b values and entropy features of Tomato image	61
3.2	Average L, a, b values and entropy features of Building image	62
3.3	Average L, a, b values and entropy features of Shuttlecock image	63
3.4	Average L, a, b values and entropy features of House image	64
3.5	Variation in center-surround bandwidth depending on the distance of the pixel from the image borders [2]	65
3.6	Examples of contrast maps: (a)Original image (b)Contrast map	69
3.7	Examples of contrast maps: (a)Original image (b)Contrast map	70
3.8	Examples of saliency cuts: (a)Original image (b)Contrast map (c)Saliency Cut (d)Ground Truth	71
3.9	Examples of saliency cuts: (a)Original image (b)Contrast map (c)Saliency Cut (d)Ground Truth	72
3.10	Combining color clustering and information based features	75
3.11	Clustering of information based features	75
3.12	Comparison with other approaches in terms of Precision, Recall and F-measure for MSR dataset	76
3.13	Comparison with other approaches (a)Original (b)FT [21] (c)HC [74] (d)FES [81] (e)DRFI [84] (f)GR [83] (g) Our Result (h)Ground Truth for MSR dataset	77

3.14	Comparison with other approaches in terms of Precision, Recall and F-measure for Berkley dataset	78
3.15	Comparison with other approaches in terms of Precision, Recall and F-measure for ECSSD dataset	78
3.16	Feature based comparison in terms of Precision, Recall and F-measure for Berkley dataset	79
3.17	Final saliency results: (a)Original image (b)Color map (c)Information map (d)Combined Saliency Cut (e)Ground Truth	80
3.18	Final saliency results: (a)Original image (b)Color map (c)Information map (d)Combined Saliency Cut (e)Ground Truth	81
4.1	First frame of each segment (FFS) based processing	85
4.2	Membership function for fuzzy segment distance (μ_d is the mean of the inter-segment pixel difference of the entire video)	88
4.3	Summarized videos of a)Diagonal strip chosen b)Artificial image created concatenating diagonals from each FFS	90
4.4	Possible FFSs selected as salient frames from diagonal based artificial image for cricket YouTube video (Total number of frames: 1304, Number of Key frames selected: 36)	91
4.5	Summarized videos of a)Horizontal strip chosen b)Artificial image created concatenating horizontal strips from each FFS	92
4.6	Summarized videos of a)Vertical strip chosen b)Artificial image created concatenating vertical strips from each FFS	92
4.7	Strip representation in each direction	93
4.8	Key frames of BMW-1 video for a)Sony's method b)Lee's method c)Ground Truth	98
4.9	Key frames of BMW-1 video for a)FRM b)SM	99
4.10	Key frames of cricket video for a)Ground Truth b)SM c)FRM	100
5.1	Basic architecture	106
5.2	Examples of Markov Logic Network for unusual activity detection	107
5.3	Examples of depth based frames	110
5.4	Examples of pseudocolored frames	111

5.5	Example of grid for frame	112
5.6	Example of Markov Logic Network for activity1	114
5.7	Examples of actions in THUMOS challenge	123
6.1	Object Discovery in Videos	130
6.2	Salient object discovery: Number of frames in Video=160, Number of frames extracted=9, Number of frames with object detected=9	131
6.3	Salient object discovery: Number of frames in Video=180, Number of frames extracted=9, Number of frames with object detected=9	132
6.4	Perception based compression results (a)Original Image (size=36.7KB) (b) Salient area marked by rectangle (c) Compression without using Saliency (size=22KB) (d) Compression with saliency(size=24.6KB)	133
6.5	Perception based compression results (a)Original Image (size=22KB) (b) Salient area marked by rectangle (c) Compression without using Saliency (size=16KB) (d) Compression with saliency (size=18.7KB)	133
6.6	Example 1: Perception based image compression results (a)Original Image, (b) DSR saliency cut [86](c) SUN saliency cut [17] (d) Our Saliency result (e) Compression without using saliency (f) Compression using saliency of DSR (g)Compression using saliency of SUN (h)Compression using our saliency . . .	134
6.7	Example 2: Perception based image compression results (a)Original Image, (b) DSR saliency cut [86](c) SUN saliency cut [17] (d) Our Saliency result (e) Compression without using saliency (f) Compression using saliency of DSR (g)Compression using saliency of SUN (h)Compression using our saliency . . .	135
6.8	Example 3: Perception based image compression results (a)Original Image, (b) DSR saliency cut [86](c) SUN saliency cut [17] (d) Our Saliency result (e) Compression without using saliency (f) Compression using saliency of DSR (g)Compression using saliency of SUN (h)Compression using our saliency . . .	136
6.9	Perception based compression results (a)Original Image, (b) Half Scale Com- pression, (c) Quarter Scale Compression, (d) Eighth Scale Compression	137
6.10	Image retargeting result using Jiang’s saliency method [5] (a) Original image (b) Demarcation of salient region (c) image retargeting at 80% width (d) image retargeting at 120% width	137

6.11 Image retargeting results using Liu’s saliency method [1] (a) Original image (b) Demarcation of salient region (c) image retargeting at 80% width (d) image retargeting at 120% width	138
6.12 Image retargeting result using our saliency method (a) Original image (b) Demarcation of salient region (c) image retargeting at 80% width (d) image retargeting at 120% width	138
6.13 A example of saliency based non-photorealistic rendering (a) Original image with salient region marked (b) Quantization of image (c) Enhancement of image (d) Lighter edges (e) Bolder edges (f) Edges present only at salient region (g) Edges not present at salient region (h) Lighter edges outside, bolder edges inside salient region (i) Bolder edges outside, lighter edges inside salient region	140
6.14 Non-photorealistic rendering comparison (a) Original image (b) Our salient region detection(c) Our results (d) Results based on Decarlo’s[54] method	141
6.15 Non-photorealistic rendering comparison (a) Original image (b) Our salient object detection(c) Our results(lighter edges) (d) Results based on Decarlo’s[54] method	141
6.16 Non-photorealistic rendering example 1 (a) Original image with yellow rectangle around detected salient region (b) Color Quantization of image(c) Enhancement of image (d) Our results using lighter edges (e) Our results using bolder edges	142
6.17 Non-photorealistic rendering example 2 (a) Original image with yellow rectangle around detected salient region (b) Color Quantization of image(c) Enhancement of image (d) Our results using lighter edges (e) Our results using bolder edges	143
6.18 Non-photorealistic rendering example 3 (a) Original image with yellow rectangle around detected salient region (b) Color Quantization of image(c) Enhancement of image (d) Our results using lighter edges (e) Our results using bolder edges	144
6.19 Non-photorealistic rendering example 4 (a) Original image with yellow rectangle around detected salient region (b) Color Quantization of image(c) Enhancement of image (d) Our results using lighter edges (e) Our results using bolder edges	145

List of Tables

1.1	Summary of saliency methods in images	9
1.2	Summary of saliency methods in videos	17
2.1	Components of chromosome labelgenes	40
4.1	Fidelity based comparison for single shot videos from ucf action dataset	96
4.2	Compression Ratio based comparison for single shot videos for ucf action dataset	96
4.3	Fidelity based comparison for multiple shot videos	96
4.4	Compression Ratio based comparison on multiple shot videos	96
4.5	Case Study for BMW-1 video [106]	97
5.1	Example of activity list	105
5.2	Example of activity list	105
5.3	Markov Logic Network Rules based on activity1	113
5.4	Some training instances	117
5.5	Some testing instances	118
5.6	Number of Instances for training	118
5.7	Number of Instances for testing	118
5.8	Confusion Matrix	119
5.9	Confusion Matrix	120
5.10	MaP based comparison for Task 1	125
5.11	MaP based comparison for Task 2	125