SALIENCY AND FEATURE BASED VISUAL OBJECT CHARACTERIZATION FOR SEGMENTATION AND LOCALIZATION

PRERANA MUKHERJEE

DEPARTMENT OF ELECTRICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY DELHI

OCTOBER 2018

SALIENCY AND FEATURE BASED VISUAL OBJECT CHARACTERIZATION FOR SEGMENTATION AND LOCALIZATION

by

PRERANA MUKHERJEE

DEPARTMENT OF ELECTRICAL ENGINEERING

Submitted

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

to the

INDIAN INSTITUTE OF TECHNOLOGY DELHI

OCTOBER 2018

To my parents

Certificate

This is to certify that the thesis entitled "**SALIENCY AND FEATURE BASED VISUAL OBJECT CHARACTERIZATION FOR SEGMENTATION AND LOCALIZATION**" being submitted by **Ms. Prerana Mukherjee** to the Department of Electrical Engineering, Indian Institute of Technology Delhi, for the award of the degree of **Doctor of Philosophy** is the record of the bona-fide research work carried out by her under my supervision. In my opinion, the thesis has reached the standards fulfilling the requirements of the regulations relating to the degree.

The results contained in this thesis have not been submitted either in part or in full to any other university or institute for the award of any degree or diploma.

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

Date:
Place: New Delhi

Acknowledgments

I would like to express my sincere gratitude to my advisor **Dr. Brejesh Lall**. The genesis of this thesis would not have been possible without his guidance, critical review and encouragement. His intelligent ideas and comprehensive understanding helped me to establish the overall direction of the research, got me past many significant challenges and always motivated me to strive for excellence.

I thank the members of my thesis committee: Prof. S.D. Joshi, Prof. Prem Kalra and Dr. Sumantra Dutta Roy for their insightful suggestions which encouraged me to widen my research from various perspectives. I sincerely admire the contribution of all my lab mates (computer vision group) Mr. Siddharth Srivastava, Mr. Vinay Kaushik and Ms. Shiva Azimi for extending their unstinting support and cooperating attitude during the course of this study. My sincere thanks to our lab incharge and project staff Dilip Ji for all the timely logistic support.

At the end I acknowledge the most important contribution of my parents. Their infallible love and reassuring rock solid support has always been my strength. I owe it all to my father, Mr. Prasanta Kumar Mukherjee, who instilled in me the inspiration for excellence and hard work. He remains a life-model worthy of emulation.

Prerana Mukherjee

Abstract

Selective attentional processing involves suitable processing of visual stimuli to localize the key objects and hence is an important research area. Due to the sophisticated human visual capabilities, it is quite trivial to distinguish the salient object from any kind of complex background, or amongst a number of salient objects while it remains challenging in case of machines. Such a task of being able to recognize and segregate the salient objects is known as *salient object segmentation.* Although a multitude of prevailing salient object detection algorithms solve various aspects of object localization and segmentation wherein they require abundant training data and rely on utilization of sophisticated resource engines. It is still difficult to capture many interesting patterns such as convexity and smoothness of region boundaries locally. Further, saliency alone cannot serve as a single cue to efficient segmentation as various attentional models identify different regions as being salient. Mostly intelligent scribbles or markers are used to improve the segmentation accuracy, thus require manual intervention. We however design and develop segmentation and localization algorithms around the central idea of object characterization which is based on the premise that objects are characterized by a well defined boundary and unique contrasting regions. Unsupervised segmentation when implemented with great deal of precision and computational efficacy eliminates the need for extensive training overhead as observed in most of the deep learning algorithms. In this thesis, we present four works: saliency guided segmentation, object characterization, holistic segmenta-

tion framework using saliency and features, and ordered salient object proposals.

As a first contribution of the thesis, we congregate the features at different granularity levels: local, global and rarity cues to construct a saliency map as a prior for segmentation. Further, to increase the discriminability of the proposed saliency map, we apply a non-linear scale space preprocessor unit to the image. The technique is applied successfully to perform multi-object segmentation.

Next, we first characterize objects by developing a novel set of features: SIKA features which is a careful combination of Scale Invariant Feature Transform (SIFT) and KAZE features as opposed to just a compendium of interest points to represent the object characteristics. We also develop a salient keypoint selection strategy based on the discriminability, detectability and repeatability of keypoints and show its effectiveness in object representation.

Thereafter, we extend the performance analysis of the abovementioned paradigms in unsupervised segmentation. We provide an end-to-end architecture for unsupervised salient object segmentation by augmenting the saliency map with an edge-aware region prior map. We also propose a novel technique for selecting an appropriate threshold for binarization of the saliency map. To this end, we make use of salient keypoints to provide the foreground seeds for leveraging the strength in object representation.

Most of the existing generic object localization algorithms usually give the plausible object locations without taking into consideration the saliency ordering of the proposal set. We present a graph-based novel object proposal generation which ranks the key objects according to their saliency score in the proposal pool. It utilizes aggregated edge cues with relative spatial layout of the edge nodes thus employing multiple Gestault principles to facilitate structured prediction. Additionally, we demonstrate the applicability of the generated proposal set for content aware image retargeting.

सार

चुनिंदा ध्यान प्रसंस्करण में प्रमुख वस्तुओं को स्थानांतरित करने के लिए दृश्य उत्तेजना के उपयुक्त प्रसंस्करण शामिल है और इसलिए यह एक महत्वपूर्ण शोध क्षेत्र है। परिष्कृत मानव दृश्य क्षमताओं के कारण, किसी भी प्रकार की जटिल पृष्ठभूमि से, या मशीनों के मामले में चुनौतीपूर्ण बनाते समय, प्रमुख वस्तुओं में से किसी वस्तु से अलग वस्तु को अलग करना बहुत मुश्किल है। मुख्य वस्तुओं को पहचानने और अलग करने में सक्षम होने का ऐसा कार्य मुख्य वस्तु विभाजन के रूप में जाना जाता है। यद्यपि प्रचलित मुख्य वस्तु पहचान एल्गोरिदम की भीड़ ऑब्जेक्ट स्थानीयकरण और विभाजन के विभिन्न पहलुओं को हल करती है, जिसमें उन्हें प्रचुर मात्रा में प्रशिक्षण डेटा की आवश्यकता होती है और परिष्कृत संसाधन इंजनों के उपयोग पर भरोसा होता है। स्थानीय स्तर पर क्षेत्र सीमाओं के उत्थान और चिकनीता जैसे कई रोचक पैटर्न को पकड़ना अभी भी मुश्किल है। इसके अलावा, अकेले क्षमता कुशल विभाजन के लिए एक क्यू के रूप में काम नहीं कर सकती क्योंकि विभिन्न ध्यान मॉडल अलग-अलग क्षेत्रों को मुख्य मानते हैं। अधिकतर बुद्धिमान स्क्रिबल्स या मार्कर का उपयोग विभाजन सटीकता में सुधार के लिए किया जाता है, इस प्रकार मैन्युअल हस्तक्षेप की आवश्यकता होती है। हालांकि हम ऑब्जेक्ट कैरेक्टरेशन के केंद्रीय विचार के आस-पास विभाजन और स्थानीयकरण एल्गोरिदम को डिज़ाइन और विकसित करते हैं जो इस आधार पर आधारित है कि ऑब्जेक्ट्स को अच्छी तरह से परिभाषित सीमा और अद्वितीय विपरीत क्षेत्रों द्वारा विशेषता है। अप्रत्याशित विभाजन जब सटीक परिशुद्धता और कम्प्यूटेशनल प्रभावकारिता के साथ कार्यान्वित किया जाता है, तो अधिकांश गहरे शिक्षण एल्गोरिदम में देखे जाने वाले व्यापक प्रशिक्षण ओवरहेड की आवश्यकता को समाप्त कर देता है। इस थीसिस में, हम चार कार्य प्रस्तुत करते हैं: लचीला निर्देशित विभाजन, ऑब्जेक्ट विशेषता, समग्र विभाजन और सुविधाओं का उपयोग करके समग्र विभाजन फ्रेमवर्क, और ऑब्जेक्ट प्रमुख ऑब्जेक्ट प्रस्तावों का आदेश दिया।

थीसिस के पहले योगदान के रूप में, हम अलग-अलग ग्रैन्युलरिटी स्तरों पर सुविधाओं को एकत्रित करते हैं: स्थानीय, वैश्विक और दुर्लभ संकेत विभाजन के लिए पहले एक लचीला मानचित्र बनाने के लिए संकेत देते हैं। इसके अलावा, भेदभाव बढ़ाने के लिए प्रस्तावित लचीलापन मानचित्र का, हम एक गैर-रैखिक स्केल स्पेस प्रीप्रोसेसर लागू करते हैं छवि के लिए इकाई। बहु-वस्तु विभाजन करने के लिए तकनीक सफलतापूर्वक लागू की जाती है।

अगला, हम पहले विशेषता है सुविधाओं का एक उपन्यास सेट विकसित करके वस्तुओं: एसआईकेए विशेषताएं जो स्केल इनवेरिएंट फ़ीचर का सावधानीपूर्वक संयोजन है ट्रांसफॉर्म (एसआईएफटी) और केएजेई फीचर्स के विपरीत ऑब्जेक्ट विशेषताओं का प्रतिनिधित्व करने के लिए ब्याज बिंदुओं का केवल एक सारांश। हम मुख्य बिंदुओं की भेदभाव, पहचान और दोहराने योग्यता के आधार पर एक प्रमुख कीपॉइंट चयन रणनीति भी विकसित करते हैं और ऑब्जेक्ट प्रस्तुति में इसकी प्रभावशीलता दिखाते हैं।

उसके बाद, हम उपरोक्त विभाजन में उपर्युक्त प्रतिमानों के प्रदर्शन विश्लेषण का विस्तार करते हैं। हम असुरक्षित मुख्य वस्तु के लिए एक अंत तक अंत वास्तुकला प्रदान करते हैं एक पूर्व-जागरूक क्षेत्र पूर्व मानचित्र के साथ लचीला मानचित्र को बढ़ाकर विभाजन। हम बिनराइजेशन के लिए उचित दहलीज चुनने के लिए एक उपन्यास तकनीक का भी प्रस्ताव देते हैं लचीला मानचित्र का। इस अंत तक, हम इसका उपयोग करते हैं ऑब्जेक्ट में ताकत का लाभ उठाने के लिए अग्रभूमि बीज प्रदान करने के लिए मुख्य कीपॉइंट्स प्रतिनिधित्व।

मौजूदा जेनेरिक ऑब्जेक्ट लोकलाइजेशन एल्गोरिदम आमतौर पर प्रस्तावित सेट के लचीले आदेश को ध्यान में रखे बिना व्यावहारिक ऑब्जेक्ट स्थानों को देते हैं। हम एक ग्राफ-आधारित उपन्यास ऑब्जेक्ट प्रस्ताव पीढ़ी प्रस्तुत करते हैं जो प्रस्ताव पूल में उनके लचीले स्कोर के अनुसार प्रमुख वस्तुओं को रैंक करता है। यह समेकित किनारे संकेतों का उपयोग किनारे नोड्स के सापेक्ष स्थानिक लेआउट के साथ करता है जिससे संरचित भविष्यवाणी की सुविधा के लिए कई गेस्टॉल्ट सिद्धांतों को नियोजित किया जाता है। इसके अतिरिक्त, हम सामग्री जागरूक छवि प्रतिस्थापन के लिए जेनरेट किए गए प्रस्ताव सेट की प्रयोज्यता का प्रदर्शन करते हैं।

Contents

List of Figures

List of Tables

Acronyms

HVS	**H**uman **V**isual **S**ystem
RF	**R**eceptive **F**ield
ILSVRC	**I**magenet **L**arge **S**cale **V**isual **R**ecognition **C**hallenge
SIFT	**S**cale **I**nvariant **F**eature **T**ransform
LGN	**L**ateral **G**eniculate **N**ucleus
SVM	**S**upport **V**ector **M**achine
ScSPM	**S**parse coding **S**patial **P**yramidal **M**atching
VR	**V**ariable **R**esolution
ROI	**R**egon **O**f **I**nterest
CRF	**C**onditional **R**andom **F**ield
PQFT	**P**hase **Q**uaternion **F**ourier **T**ransform
SR	**S**pectral **R**esidue
PFT	**P**hase spectrum of **F**ourier **T**ransform
QFT	**Q**uaternion **F**ourier **T**ransform
LoG	**L**aplacian of **G**aussian
DoG	**D**ifference of **G**aussian
AOS	**A**dditive **O**perator **S**plitting
BoVW	**B**ag of **V**isual **W**ords
FV	**F**isher **V**ector
GMM	**G**aussian **M**ixture **M**odel
HoG	**H**istogram **O**f **G**radient
MSRA	**M**icrosoft **R**esearch **A**sia
EM-GMMbic	**E**xpectation **M**aximization **G**aussian **M**ixture **M**odel using bayesian information criteria

GMM-HMRF	**G**aussian **M**ixture **M**odel based **H**idden **M**arkov **R**andom **F**ield
BSD	**B**erkeley **S**egmentation **D**ataset
MAP	**M**aximum **AP**Aposteriori
MCM	**M**inimal **C**omplexity **M**achine
GSS	**G**aussian **S**cale **S**pace
VC	**V**apnik **C**hervonenkis
LPP	**L**inear **P**rogramming **P**roblem
ED	**E**uclidean **D**istance
PCA	**P**rincipal **C**omponents **A**nalysis
mAP	**m**ean **A**verage **P**recision
KOS	**K**eypoint **O**verlap **S**core
MKOS	**M**ean **K**eypoint **O**verlap **S**core
DRFI	**D**iscrimintive **R**egional **F**eature **I**ntegration
HS	**H**ierarchical **S**aliency
MAE	**M**ean **A**verage **E**rror
PRI	**P**robabilistic **R**and **I**ndex
VoI	**V**ariation of **I**nformation
GCE	**G**lobal **C**onsistency **E**rror
BDE	**B**oundary **D**isplacement **E**rror
EFI	**E**dge **F**oci **I**nterest
EBR	**E**dge **B**ased **R**egion
CAKE	**C**ontext **A**ware **K**eypoint **E**xtractor
IRQA	**I**mage **R**etargeting **Q**uality **A**ssessment
OEF	**O**riented **E**dge **F**orets
NMS	**N**on **M**aximal **S**uppression
LBP	**L**ocal **B**inary **P**attern
LTP	**L**ocal **T**ernary **P**attern
IoU	**I**ntersection over **U**nion
CASC	**C**ontent **A**ware **S**eam **C**arving
GBVS	**G**raph **B**ased **V**isual **S**aliency
EMD	**E**arth **M**over **D**istance
ARS	**A**spect **R**atio **S**imilarity
BSDIL	**BD**idirectional **S**alient **I**nformation **L**oss
FSIL	**F**orward **S**aliency **I**nformation **L**oss
BSIL	**B**ackward **S**aliency **I**nformation **L**oss
SGSD	**S**alient **G**lobal **S**tructure **D**istortion
GOP	**G**eodesic **O**bject **P**roposal
MCG	**M**ultiscale **C**ombinatorial **G**rouping