NOVEL DEEP LEARNING BASED FRAMEWORK FOR IMAGE AND VIDEO ENHANCEMENT

MANOJ SHARMA

DEPARTMENT OF ELECTRICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY DELHI

JANUARY 2020

NOVEL DEEP LEARNING BASED FRAMEWORK FOR IMAGE AND VIDEO ENHANCEMENT

by

MANOJ SHARMA

DEPARTMENT OF ELECTRICAL ENGINEERING

Submitted

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

to the

INDIAN INSTITUTE OF TECHNOLOGY DELHI

JANUARY 2020

To my family

Certificate

This is to certify that the thesis titled **Novel Deep Learning based Framework for Image and Video Enhancement** being submitted by **Mr. Manoj Sharma** 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 his 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

I would like to express my sincere gratitude to my advisors **Prof. Santanu Chaudhury** and **Prof. Brejesh Lall**. This thesis would not have been possible without their guidance, critical review and encouragement. Their 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. Prem Kalra, Dr. Sumantra Dutta Roy and Dr. Sumeet Agarwal for their insightful suggestions which encouraged me to widen my research from various perspectives. I sincerely admire the contribution of my lab mates Ms. Swati Bhugra, Ms. Ritu Chaudhury, Ms. Anupama Ray, Mr. Raunak Gupta and Ms. Nidhi Chahal for extending their unstinting support and sympathetic attitude during the course of this study. I am grateful to my wife for supporting me during my Ph.D. My thanks to Ushaji, Manrika Mam and Amit Bhaiya for all the timely logistic support.

I acknowledge the significant contribution of my parents. Their patience and support will always be my inspiration. I would like to express my heartiest gratitude to my friends, Satyam and Sudhakar, who stood by me through all my efforts and encouraged me to pursue my Ph.D.

Manoj Sharma

Abstract

Image enhancement is a technique to improve the interpretability or information perception in images for human understanding. The enhanced image can also be used as an input to other image processing applications. However, image enhancement is extremely challenging due to lack of sufficient prior information and is still an open problem. Enhancement encapsulates various task such as image super-resolution, denoising, contrast enhancement, noise-free super-resolution to name a few. These enhancement techniques serve as a pre-processing step in various computer vision applications such as optical character recognition (OCR), bio-medical etc. The advancement in image enhancement techniques has resulted in the requirement of efficient frameworks that can evaluate the quality of image/video data automatically. Objective quality estimation approaches can be applied to quality control systems for monitoring the image and video quality, parameter setting and optimization of benchmark algorithms for image and video processing system.

The main objective of the thesis is image and video enhancement. Among various enhancement techniques, image and video super-resolution (SR) has gained a lot of attention because of increasing availability of high-resolution (HR) displays. This is a challenging problem due to lack of prior information and image statistics. For images, we focus on spatial resolution enhancement. In addition, noise-robust resolution enhancement of noisy low-resolution (LR) images is even more challenging as both image de-noising and super-resolution tasks are ill-posed inverse problems. If an LR image contains noise or distortions, a simple resolution enhancement task gives rise to additional incremental noise or distortions. Moreover, performing de-noising followed by super-resolution does not provide HR image with fine details. Thus,

there is a need for joint optimization of image de-noising and super-resolution while preserving fine details. In case of video, resolution enhancement means enhancing resolution in both space and time. For video enhancement, our work focuses on finding a solution in order to get space-time resolution enhancement of the LR video sequences.

Image in-painting is a restoration technique that focuses on removing distorted parts or filling out missing regions to enhance images. However, in-painting of larger missing region with local details has not yet been solved. In this, we present a solution for filling larger missing regions by using super-resolution based in-painting. As previously mentioned, the solutions of the aforementioned enhancement tasks depend on image/video quality assessment. For quality assessment, past efforts have largely focused on specific kind of distortion. However, it is not practically feasible to define different type of distortions and their causes. Thus our work focuses on a technique which is independent of the type of error present and concentrates on image quality rather than on the type of specific error.

Optical character recognition (OCR) systems face severe challenges in recognizing text from document image that lack resolution. Most OCR models trained on high resolution (HR) text images fail in case of low resolution (LR), noisy or noisy-LR text images. Enhancement of noisy-LR text images, such that existing OCR engine can recognize the text from noisy-LR images is of interest to the OCR community. In the thesis, we have demonstrated an application of noise-robust resolution enhancement technique to boost existing OCR systems.

सार

इमेज एन्हांसमेंट व्याख्या या सूचना की धारणा को बेहतर बनाने की तकनीक है मानव समझ के लिए चित्र। बढ़ी हुई छवि को अन्य के लिए एक इनपुट के रूप में भी इस्तेमाल किया जा सकता है छवि प्रसंस्करण अनुप्रयोगों। हालाँकि, छवि वृद्धि बेहद चुनौतीपूर्ण है पर्याप्त पूर्व सूचना की कमी और अभी भी एक खुली समस्या है। एन्हांसमेंट बढ़ जाता है विभिन्न कार्य जैसे कि इमेज सुपर-रिज़ॉल्यूशन, डीनोइजिंग, कंट्रास्ट एन्हांसमेंट, नॉइज़-फ्री सुपरसेलिंग कुछ नाम है। ये एन्हांसमेंट तकनीक प्री-प्रोसेसिंग कदम के रूप में काम करती है विभिन्न कंप्यूटर दृष्टि अनुप्रयोग जैसे कि ऑप्टिकल चरित्र पहचान (ओसीआर), जैव-चिकित्सा आदि छवि वृद्धि तकनीकों में प्रगति की आवश्यकता के परिणामस्वरूप हुई है कुशल चौखटे जो स्वचालित रूप से छवि / वीडियो डेटा की गुणवत्ता का मूल्यांकन कर सकते हैं। उद्देश्य निगरानी के लिए गुणवत्ता नियंत्रण प्रणालियों के लिए गुणवत्ता अनुमान दृष्टिकोण लागू किया जा सकता है छवि और वीडियो की गुणवत्ता, पैरामीटर सेटिंग और छवि के लिए बेंचमार्क एल्गोरिदम का अनुकूलन और वीडियो प्रसंस्करण प्रणाली।

थीसिस का मुख्य उद्देश्य छवि और वीडियो वृद्धि है। विभिन्न वृद्धि के बीच तकनीक, छवि और वीडियो सुपर-रिज़ॉल्यूशन (एसआर) ने बहुत अधिक ध्यान आकर्षित किया है उच्च-रिज़ॉल्यूशन (एचआर) डिस्प्ले की बढ़ती उपलब्धता। यह एक चुनौतीपूर्ण समस्या है पूर्व सूचना और छवि आँकड़ों की कमी। छवियों के लिए, हम स्थानिक संकल्प पर ध्यान केंद्रित करते हैं वृद्धि। इसके अलावा, शोर कम संकल्प (LR) शोर-मजबूत संकल्प वृद्धि छवियां और भी चुनौतीपूर्ण हैं क्योंकि दोनों चित्र डी-नॉइज़िंग और सुपर-रिज़ॉल्यूशन कार्यों को अनप्लग कर रहे हैं उलटा समस्या। यदि LR छवि में शोर या विकृतियां होती हैं, तो एक सरल रिज़ॉल्यूशन वृद्धि कार्य अतिरिक्त वृद्धिशील शोर या विकृतियों को जन्म देता है। इसके अलावा, प्रदर्शन सुपर रिज़ॉल्यूशन के बाद डी-नॉइज़ ठीक विवरण के साथ एचआर छवि प्रदान नहीं करता है। इस प्रकार, संरक्षण करते समय छवि डी-शोर और सुपर-रिज़ॉल्यूशन के संयुक्त अनुकूलन की आवश्यकता है बारीक विवरण। वीडियो के मामले में, रिज़ॉल्यूशन एन्हांसमेंट का मतलब दोनों में रिज़ॉल्यूशन को बढ़ाना है

स्थान और समय। वीडियो एन्हांसमेंट के लिए, हमारा काम पाने के लिए एक समाधान खोजने पर ध्यान केंद्रित करता है LR वीडियो दृश्यों का स्पेस-टाइम रिज़ॉल्यूशन एन्हांसमेंट। छवि-पेंटिंग एक पुनर्स्थापना तकनीक है जो विकृत भागों को हटाने पर केंद्रित है या छवियों को बढ़ाने के लिए लापता क्षेत्रों को भरना। हालांकि, बड़े लापता क्षेत्र की पेंटिंग स्थानीय विवरण के साथ अभी तक हल नहीं किया गया है। इसमें, हम बड़े लापता को भरने के लिए एक समाधान प्रस्तुत करते हैं इन-पेंटिंग आधारित सुपर-रिज़ॉल्यूशन का उपयोग करके क्षेत्र। जैसा कि पहले बताया गया है, के समाधान उपरोक्त संवर्द्धन कार्य छवि / वीडियो गुणवत्ता मूल्यांकन पर निर्भर करते हैं। गुणवत्ता के लिए मूल्यांकन, पिछले प्रयासों ने काफी हद तक विशिष्ट प्रकार की विकृति पर ध्यान केंद्रित किया है। हालांकि यह है विभिन्न प्रकार की विकृतियों और उनके कारणों को परिभाषित करने के लिए व्यावहारिक रूप से संभव नहीं है। इस प्रकार हमारा काम एक ऐसी तकनीक पर ध्यान केंद्रित करता है जो वर्तमान में मौजूद त्रुटि के प्रकार से स्वतंत्र होती है और उस पर ध्यान केंद्रित करती है विशिष्ट त्रुटि के प्रकार के बजाय छवि गुणवत्ता।

ऑप्टिकल कैरेक्टर रिकग्निशन (OCR) सिस्टम टेक्स्ट को पहचानने में गंभीर चुनौतियों का सामना करते हैं दस्तावेज़ छवि से जिसमें संकल्प की कमी है। उच्च रिज़ॉल्यूशन (एचआर) पर प्रशिक्षित अधिकांश ओसीआर मॉडल पाठ चित्र कम रिज़ॉल्यूशन (LR), शोर या शोर-LR पाठ छवियों के मामले में विफल होते हैं। वृद्धि शोर-एलआर पाठ छवियों, जैसे कि मौजूदा ओसीआर इंजन शोर-एलआर से पाठ को पहचान सकता है चित्र OCR समुदाय के लिए रुचि रखते हैं। थीसिस में, हमने एक एप्लिकेशन का प्रदर्शन किया है मौजूदा ओसीआर प्रणालियों को बढ़ावा देने के लिए शोर-मजबूत संकल्प वृद्धि तकनीक।

Contents

List of Figures

List of Tables