Annual Progress Seminar - II Report

on

Investigation of Techniques for Latent Space Representation of Social Networks

Submitted

in partial fulfillment of

the requirements of the degree of

Doctor of Philosophy (Computer Engineering)

by

Mr. Subodh Wandile

(171080074)

LOGO.png LOGO.png

Department of Computer Engineering & Information Technology

Veermata Jijabai Technological Institute, Mumbai - 400019

(Autonomous Institute Affiliated to University of Mumbai)

2017 - 2018

CERTIFICATE

This is to certify that, **subodh wandile (ID- 171080074** , a student of **Bachelor Of Technology (Information Technology)** has completed the lab report held in Jan-Feb 2019 for working on COMPUTER VISION to our satisfaction.

Mr.Pranav Nerurkar	Dr. V. B. Nikam
Supervisor	HOD of CE & IT

Declaration of the Student

I declare that this written submission represents my ideas in my own words and where others' ideas or words have been included, I have adequately cited and referenced the original sources.

I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea / data / fact / source in my submission.

I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

subodh wandile
(Reg no : 171080074)   Date: ________________

Acknowledgements

My sincere thanks goes to Dr. Dhiren Patel, Director of VeermataJijabai Technological Institute, Mumbai who provided me an opportu-nity to work as a research scholar and gave me access to the laboratoryand research facilities which helped me to conduct this research.I would like to express my sincere gratitude to my co-supervisor Dr.M. M. Chandane and supervisor Dr. S. G. Bhirud for the continuoussupport of my Ph.D study and related research, for their patience,motivation, and immense knowledge. Their guidance helped me in allthe time of research and writing of this report.I thank our Head of the Department Dr. V. B. Nikam, for extendinghis invaluable support. I would also thank my institute and the de-partment's faculty members without whom this research work wouldhave been a distant reality.I also extend my heartfelt thanks to my family and well wishers

PREFACE

Camera phones have penetrated every corner of society and have become a fo-cal point for communications. In our research we extend the traditional use of suchdevices to help bridge the gap between physical and digital worlds. Their combinedimage acquisition, processing, storage, and communication capabilities in a compact,portable device make them an ideal platform for embedding computer vision andimage processing capabilities in the pursuit of new mobile applications. This disser-tation is presented as a series of computer vision and image processing techniquestogether with their applications on the mobile device. We have developed a set oftechniques for ego-motion estimation, enhancement, feature extraction, perspectivecorrection, object detection, and document retrieval that serve as a basis for suchapplications. Our applications include a dynamic video barcode that can transfersignificant amounts of information visually, a document retrieval system that canretrieve documents from low resolution snapshots, and a series of applications for the users with visual disabilities such as a currency reader. Solutions for mobile de-vices require a fundamentally different approach than traditional vision techniquesthat run on traditional computers, so we consider user-device interaction and thefact that these algorithms must execute in a resource constrained environment. Foreach problem we perform both theoretical and empirical analysis in an attempt tooptimize performance and usability. The thesis makes contributions related to effi-cient implementation of image processing and computer vision techniques, analysisof information theory, feature extraction and analysis of low quality images, anddevice usability.

10th April 2019

Contents

Chapter 1

Introduction

Integrated image capture, processing, and communication power on a compact,portable, hand-held device is attracting increased interest from computer vision re-searchers with a goal of applying a diverse collection of vision tasks on the smallhand-held device. In this dissertation, we survey the state-of-the-art computer vi-sion/pattern recognition technologies that have been developed on camera phones.We identify potential problems and difficulties with the camera phone and providesolutions to these problems. Furthermore we foresee the trend of vision applicationson camera phones and propose solutions to the major technical challenges involvingmobile vision and pattern recognition. On March 10, 1876, Alexander Graham Bell first transmitted speech electron-ically, setting in motion the development of telephone systems which have evolved-into the devices we enjoy today. Throughout this revolution, voice communicationhas remained as the primary function of phones despite the introduction of wirelesscapabilities and, more recently, the view of mobile devices as extending the desktopcomputer to be used anytime and anywhere

Chapter 2

Technical Contributions and Related Research

In this chapter we present a series of computer vision and image processingtechniques that we have developed for mobile applications. These techniques repre-sent the basic elements required to pursue mobile applications that involve vision andrecognition. The techniques being described relate to vision and recognition fromthe images of planar objects such as documents, signs and symbologies, but theymay also be applied to other objects. These techniques span the areas of patternrecognition, geometric vision (e.g., perspective correction), egomotion estimation,and pixel level image processing (e.g., contrast enhancement and interpolation). Inone common feature shared by all the techniques introduced in this chapter, they-all have been implemented, tested, and optimized for mobile devices. They serve asbuilding blocks to develop the vision system and applications throughout this thesisand will be referred to in the following chapters. We also survey related researchand discuss how each contributes to the field. Pattern recognition may serve as an important capability in the pursuit ofnew mobile applications and will lead us to the goal of making the physical world"clickable". For example, a recognized logo may lead the user to digital couponsor online ads, or a recognized face may serve as the key to open the mobile device.Pattern recognition can also help people with visual disability to identify objectssuch as currencies in daily life. In this section, we describe efficient technology for"target based" recognition.Classic pattern recognition algorithms usually include feature extraction andfeature classification as two core components. Popular features such as SIFT[37] orSIFT-like[38] have high repeatability. SVM [39] and neural networks [40] can betrained to achieve high accuracy given enough time and space allowance. However,these classic pattern recognition approaches cannot be ported directly to mobiledevices. As we noted in the introduction, implementing pattern recognition onmobile devices has three major challenges. 1) The limited processing power ofthe device, 2) the fact that the captured scene could contain complex backgroundresulting in false positive that must be eliminated, and 3) the expectation of the userwho typically expects instant feedback and requires online (real time) recognition.

Chapter 3

VCode - Imaging as an Alternative Mobile Data Channel

We introduce a novel usage of the camera as a visual communication channel,which is especially appropriate for mobile applications. Data is encoded as a series ofmachine readable dynamic frames, captured by a mobile camera, and reconstructedon the device. We address the problem of information uncertainty in the camerachannel because errors may exist within a given frame and frames can be missedduring detection. We introduce a novel spatial-temporal error correction method toencapsulate data being sent through this highly vulnerable channel with both errorand data loss. To maximize data capacity per pixel, we use color to represent bits soeach pixel can carry multiple bits of information. We introduce a color calibrationpattern to adapt for different cameras and displays and simultaneously learn thechannel property when decoding the captured frames. As a theoretical contribution,we estimate the capacity of the camera channel using mutual information theory anddemonstrate that a proper selection of colors for encoding can optimize throughput. V-Code can allow a camera to operate as an alternative data channel. Thebasic concept of the camera channel is illustrated in Figure 3.2. The data is encodedinto a sequence of images, which are animated on a flat panel display, acquired by thecamera, and decoded by software on the device. The camera channel is free, passive,pervasive, and especially attractive to advertisers because it can be integrated intothe physical world. Flat panel displays have been installed widely in public areas, sothe camera channel has a serious potential for delivering multimedia content (ringtones, pictures, Java games, etc.) to mobile users.The development of V-Code is motivated by three primary factors. First,the current mobile data access is still sub-optimal. GPRS/CDMA data plans arecostly and may not work indoors or underground due to poor signal coverage. Blue-tooth and Infrared require extra hardware (adapter) and driver installation

3.1 MobileEye Tools - for Persons with Visual Disabilities

Visual impairments affect a large percentage of population in various ways,including color-deficiency, myopia, low vision, and other more severe visual disabil-ities [118]. Current estimates suggest approximately ten million blind or visually-impaired individuals live in the United States alone. Visual impairment significantlyaffects quality of life within these populations and most have no effective cure. De-vices that provide visual enhancement are in large demand, but are often expensive,bulky, and dedicated to a single task. For example, an optical image enhancer suchas a magnifier can provide basic zoom function but cannot enhance the contrast,brightness, color, and other details of an image. Electronic image enhancers (e.g.,"Acrobat LCD" and "Sense View Duo") are powered by digital image processingtechnology and these programmable devices can be customized for various require-ments of the visually impaired users. This device, like others, is a special purposepiece of hardware.

3.2 Existing System and Design Principles

The MobileEye system is designed for persons with visual impairments, it isimpractical to require sophisticated operations, which are already difficult for userswith normal vision because of the small keypad input [19]. We follow the overalldesign principle of minimizing user operation, especially with respect to the keypad hits e.g. the pattern recognizer. A typical mobile pattern recognition system (such asthe Kurzweil K1000) requests the user to take a snapshot, then the system attemptsto recognize the result. If the image is imperfect, the recognition may fail and theuser will have to repeat the process. However, we cannot expect a visually impaired-user to perform such tasks, and it is impractical to mandate high quality pictures forrecognition. We choose to process the image in real time instead of an "acquisition-process-repeat" loop, which provides a much smoother user experience. Our patternrecognizer operates on the device and processes approximately 10 frames per second,so, the user receives an instant response as the camera acquires new content. Thisintroduces the challenge that we must process the video stream from the cameraquickly. We address this problem using a boosted object detector with both highefficiency and accuracy.

3.3 Abstract

Camera phones have penetrated every corner of society and have become a fo-cal point for communications. In our research we extend the traditional use of suchdevices to help bridge the gap between physical and digital worlds. Their combinedimage acquisition, processing, storage, and communication capabilities in a compact,portable device make them an ideal platform for embedding computer vision andimage processing capabilities in the pursuit of new mobile applications. This disser-tation is presented as a series of computer vision and image processing techniquestogether with their applications on the mobile device. We have developed a set

oftechniques for ego-motion estimation, enhancement, feature extraction, perspectivecorrection, object detection, and document retrieval that serve as a basis for suchapplications.

3.4 Display Frame Rate

The display frame rate generally depends on the speed with which a frame canbe captured and processed by camera phones, which is device dependent. A framecannot be displayed too quickly as camera phones need to have enough time to perform geometric correction, decoding, and error correction. If what is displayed tooslowly, however, the camera phone will have to process the same frame repeatedly.Although the duplicate data will be identified and removed, re-decoding decreases. the overall bit rate. The ideal situation occurs when camera phones process everyframe exactly once. If a frame is dropped, it can be recovered by error correctionor be recaptured in the following round because the VCode is displayed in a loop.We tested four different display frame rates using a NOKIA 6680 camera phone asthe capture device. The data file selected was a 4KB MIDI ring tone encoded as aVCode containing 60 frames. The VCode was displayed at frame rate of 20, 10, 6.6,4 frames/second respectively on a 15-inch flat panel computer monitor. For eachframe rate we let three users download the file onto the camera phone. The timetused for download is recorded for each run, and the throughput is calculated as40968/tbps.

3.5 Overall Downloading Bit Rate

After analyzing specific factors affecting the download speed, we evaluate theoverall throughput in a more comprehensive data set. We selected three data files,including a MIDI ring tone, a Java game, and a 3GP video, as our test set. Table 3.3lists the sizes of these files. We allowed the same three users download these files andrecorded the time spent on downloading when the final download is complete. Thebit rate is defined as the quotient of a file size over the time spent on downloading.The average bit rates for downloading are also shown in Table 3.3, demonstratingthat the bit rate decreases as the file size increases. For comparison, we put thephone on a dock on a desk so both the phone and monitor are static, a configurationwe call “dock” mode. In dock mode, the download bit rate is stable, independent ofthe file size, since no users’ factors are involved and the bit rate is higher (around3.3 Kbps) than in hand-held mode.

3.6 Color VCode

Naturally the addition of color to the VCode has the potential to increase theVCode’s bit capacity significantly by representing more than one bit per pixel. maximize VCode’s capacity, it is desirable to represent as many bits as possible perpixel. This leads to a fundamental question of image information theory, how manybits can be read from a color pixel, subject to camera dependent distortion andenvironmental noise. We will try to answer this question using mutual informationtheory. Moreover, we will show that the capacity of the camera channel can bemaximized using a proper selection of colors. Despite these theoretical challenges,we will also discuss practical issues, such as how to locate the V-Code accurately inreal time and how to guarantee data integrity when a frame drop occurs. Finally,the overall bit rate of the colored V-Code will be tested by downloading multimediacontent to the camera phone via the camera channel.

Chapter 4

Conclusion

4.1 Summary of Contributions

In this dissertation we have presented a series of computer vision and im-age processing techniques designed to operate on mobile devices with low qualityimaging. Using these technologies, we have built applications that bridge the gapbetween the physical and digital worlds. We have developed fast algorithms toperform perspective correction, ego-motion estimation, contrast enhancement, andbinary image interpolation. We have eliminated floating point operations in perspec-tive correction and simplified it to the computation of seven cross products, and usea look-up tables to speed interpolation. Our coarser-to-fine image ego-motion es-timation runs in real time and can be used to control the cursor and browsing onmobile devices. Besides these techniques, we have answered fundamental questionsin two areas: visual communications and document retrieval. We have built a modelfor color uncertainty of the camera channel. Using mutual information theory, wehave estimated the capacity per color pixel from a camera captured frame. We haveaddressed the problem of rejecting a query that is not in the repository for documentimage retrieval. We also quantitatively estimated the minimum area required in aquery on a 100K page document database.

4.2 Summary

We have introduced our MobileEye system, which aids individuals with vi-sual impairments to better "see" and understand their surroundings. Our system requires little effort to operate and can be deployed to a wide varieties of cameraenabled mobile devices. We do face one major challenge in reaching the end user.Most handsets together with their software applications are deployed by the wirelessservice providers. We have put our Color Mapper subsystem online and received85 downloads and some positive feedback. Our end user with visual disabilities,however, may not have the knowledge and skill to download and install the soft-ware themselves. It may require the cooperation of service providers and probablegovernment support to promote the MobileEye system to a larger number of users. generated the .eps. Alternatively GSView13 can compute the optimal bounding box14 or the user can directly open the .eps le with a text editor and modify the values dening the bounding box, which are usually in the rst few lines. The details on how to use gures with PDF LATEX

4.3 Future Working

our research, we utilize the phone's camera as an alternative input devicefor bridging the gap between the physical and digital worlds. We can interact withthe physical world using the camera. However, increasingly we will be driven tomove from the physical world to the mobile world and wish to do so in a seamlessway. In our future research we will unify recognition and retrieval on the mobiledevice. With the ultimate goal of making the physical world "clickable," no matterthe source is a barcode, a document, or a TV screen, we will link them to thecorresponding digital content. At that time the physical world will have multiplelayers of multimedia representation. The camera enabled mobile device will serve asthe key to access these layers and numerous novel applications can be built on themobile devices. We will extend the "MobileRetriever" to non-Latin languages and then to general scenes. The computation is not necessarily hosted on the device, with the help of "cloud computing," we can achieve a transparent access to multimedia representation of the world.

Chapter 5

REFERENCES

[1]C. Milanesi, A. Liang, N. Mitsuyama, H. Vergne, T. Nguyen,

[2]R. Ballagas, J. Borchers, M. Rohs, and J.G. Sheridan. The

[3]M. Davis, M. Smith, F. Stentiford, A. Bamidele, J. Canny,

[4]T. Yeh and T. Darrell. Multimodal question answering for m

[5]Y. Ijiri, M. Sakuragi, and S. Lao. Security management for

[6]X. Chen, J. Yang, J. Zhang, and A. Waibel. Automatic detec

[7]A pda-based sign translator. InICMI '02: Proceedings of th

[8]J. Zhang, X. Chen, J. Yang, and A. Waibel. A PDA-based Sig

[9]C. Seifert, L. Paletta, A. Jeitler, E. Hodl, J.P. Andreu,