Computer Vision and Machine Learning in Wearable Devices

Thesis submitted in partial fulfillment
of the requirements for the degree of

DEGREE
in
COURSE

by

Kumar Vishal
201250925
kumar.vishal@research.iiit.ac.in

Center for Visual Information Technology
International Institute of Information Technology
Hyderabad - 500 032, INDIA
December 2015

International Institute of Information Technology
Hyderabad, India

CERTIFICATE

It is certified that the work contained in this thesis, titled "Computer Vision and Machine Learning in Wearable Devices" by Kumar Vishal, has been carried out under my supervision and is not submitted elsewhere for a degree.

Date	Adviser: Dr. C. V. Jawahar

To my family

Acknowledgments

First of all, I would like to show my deepest gratitude towards my adviser, Dr. C. V. Jawahar, for his guidance, support, motivation and for keeping me from going off track. I would also like to wholeheartedly thank my family, my parents and my wife, for their love and unwavering support in all my ventures. I am thankful to my lab mates and friends at IIIT for all the fruitful(as well as not so fruitful) discussions Anand, Devendra, Udit, Praveen, Yashaswi, Nataraj, Mohak, Rajvi. A special thanks to all the faculty members at CVIT, my adviser Dr. Jawahar as well as Dr. P. J. Narayanan, Dr. Anoop Namboodiri, Dr. Jayanthi Sivaswamy for creating a wonderful environment in CVIT to do research, and CVIT administration Satya, Phani, Rajan and Nandini for helping me on numerous occasions.

Abstract

Sensor fusion is a process by which data from several different sensors are ”fused” to compute something more than could be determined by any one sensor alone. Simple embedded devices devices like smart phones, smart watches encompass wide variety of sensors like camera, GPS , accelerometer etc. In this thesis, we have shown how user experience of these devices can be greatly enhanced by providing them capability to process and analysis these data sensor data automatically. Fusing GPS and vision sensor we have designed an algorithm for localization in 3D and showed how these sensor complementary for each other. Moving ahead in a different project we showed how analysing and fusing the different sensor data along with machine learning algorithms can boost the performance of these devices. We tried to optimize the power consumption of batteries in smartphones by using the sensor data from accelerometer, touch screen, cpu usage etc.

Consider a wearable device for localization in 3D which answers the question ”Where Am I ?” for a given environment. Localization in 3D is an important problem with wide ranging applications from autonomous navigation in robotics to location specific services on wearable and mobile devices. GPS sensors are a commercially viable option for localization, and are ubiquitous in their use, especially in portable devices. With the proliferation of mobile cameras however, maturing localization algorithms based on computer vision are emerging as a viable alternative. Although both vision and GPS based localization algorithms have many limitations and inaccuracies, there are some interesting complimentarities in their success/failure scenarios that justify an investigation into their joint utilization. Such investigations are further justified considering that many of the modern wearable and mobile computing devices come with sensors for both GPS and vision. In this work, we investigate approaches to reinforce GPS localization with vision algorithms and vice versa. Specifically, we show how noisy GPS signals can be rectified by vision based localization of images captured in the vicinity. Alternatively, we also show how GPS readouts might be used to disambiguate images when they are visually similar looking but belong to different places. Finally, we empirically validate our solutions to show that fusing both these approaches can result in a more accurate and reliable localization of videos captured with a Contour action camera, over a 600 meter long path, over 10 different days.

There is a rapid growth in memory and processing power of mobile devices unfortunately the battery life is still limited in terms of size and capacity. This implies that managing the battery power is paramount in such devices. As long as the battery technology continues its slow pace of improvement, the only viable approach is to reduce the amount of energy required to provide specific services. Two of the most power consuming services on a smartphone are network and wireless data. Though internet connectivity is important, the nature of data transfer does not require uninterrupted service. We take advantage of this fact to cut off power to these modules when it is not required. The challenge is to predict, when the users require these services and when they do not. We log the sensor data and fuse them to make one single feature and using machine learning approaches to intelligently schedule Wi-Fi according to a users activity level. Several higher order features from the raw sensor data stream are used to classify the activity level. Two aspects of the problem are considered, namely the accuracy of estimating the activity level as well as the power required in sensing and estimation. Experimental results on Android based smartphones demonstrate that an active user can get up to 37% increase in battery life without significant effect on the user experience.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Sensor data fusion is an emerging research field whose aim is to combine information from multiple and diverse sources (e.g. different sensors thermal and visible cameras, laser, GPS , accerelometer etc.) to achieve inferences that cannot be obtained from a single sensor or source, or whose quality exceeds that of an inference drawn from any single source. Sensor data fusion is a multi-disciplinary subject that includes areas like statistical estimation, computer vision machine learning etc. In computer vision familiarly image sensor is fused with other sensors in order to develop an intelligent system. Now days cameras are cheap small and ubiquitous, and with the emergence of energy-efficient and powerful processors, it has become possible to incorporate practical computer vision and machine learning capabilities into embedded systems, mobile devices etc.

Localization is a fundamental problem associated with autonomous navigation. One of the simplest solution for the localization is with the help of Global Positioning Systems(GPS). In literature, computer vision methods have also been successfully used for predicting the location. A wide variety of modern gadgets (eg. smartphones) have both GPS and vision sensors, and such devices are becoming increasingly affordable for low cost robotic systems. Unfortunately, the accuracy of the popular GPS tracking devices are limited to only 10$\sim$20 meters, which is insufficient for many robotic tasks. Hence noise in the GPS signal becomes a significant issue if we want a reliable localization performance. Where as vision based localization methods suffers problems like occlusion, perceptual aliasing etc. The ambiguities in visual localization can be reduced with the help of GPS . Similarly, noise in the GPS signal can be reduced by looking at the visual consistency across multiple sessions. We empirically validate our solution to show that fusing both these approaches can result in a more accurate and reliable localization.

In a smart phones data from different sensors can be used to improve the performance of the device. We propose a novel approach to schedule services like Wi-Fi and 3G on smartphones. Using Wi-Fi as an example, we show that intelligent scheduling based on a users activity level leads to lower power consumption without adversely affecting the user experience. Services like Wi-Fi and 3G continue to consume significant amount of battery on smartphones. Hence, reducing the power consumption by these processes could help in huge power savings. We model and monitor the user activity level to decide

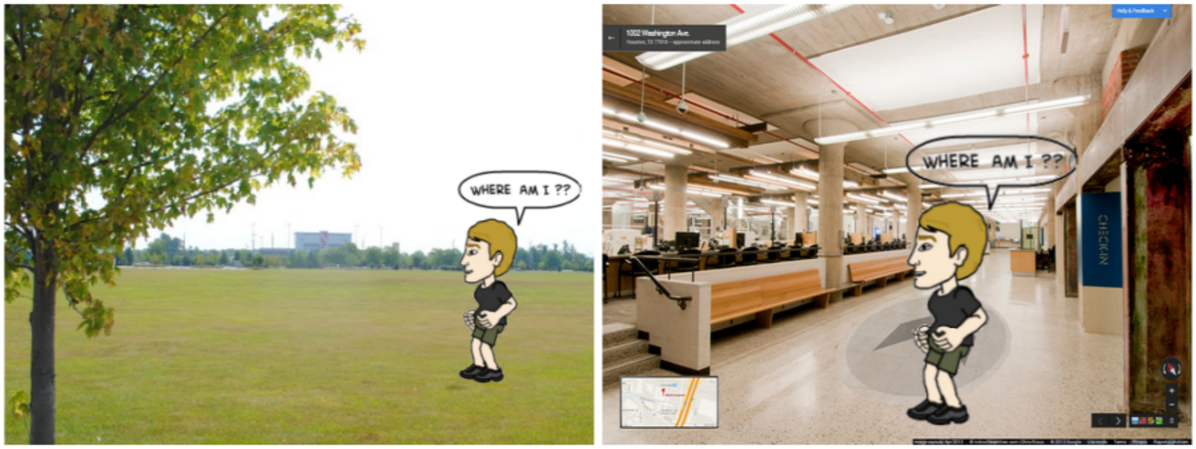

Figure 1.1 Localization is a process that determine the position of a robot/human pedestrians in a given environment.

when the wireless data module should be turned off for maximal energy ssaving without compromising the user experience.

But first, in section 1.1, we describe the computer vision tasks tackled in this thesis in detail. In section 1.2, we discuss the novel scenario challenges which need to be addressed while designing strategies for the computer vision tasks. In section 1.3, we present few machine learning approaches which could be used for tackling the novel scenarios in the computer vision tasks. In Section 1.4, we discuss our problem statement. The primary technical contributions of this thesis have been highlighted in Section 1.5

1.1 First Section

Text of section 1 goes here...

1.2 Second Section

Text of section 2 goes here...

Few suggestions

1.2.1 Mathematics

Please number all of your sections and displayed equations. It is important for readers to be able to refer to any particular equation. Just because you didn't refer to it in the text doesn't mean some future reader might not need to refer to it. It is cumbersome to have to use circumlocutions like "the equation second from the top of page 3 column 1". (Note that the ruler will not be present in the final copy, so is not an alternative to equation numbers). All authors will benefit from reading Mermin's description of how to write mathematics (see math.pdf).

1.2.2 Footnotes

Please use footnotes[1] sparingly. Indeed, try to avoid footnotes altogether and include necessary peripheral observations in the text (within parentheses, if you prefer, as in this sentence). If you wish to use a footnote, place it at the bottom of the column on the page on which it is referenced. Use Times 8-point type, single-spaced.

1.2.3 References

List and number all bibliographical references in 9-point Times, single-spaced, at the end of your paper. When referenced in the text, enclose the citation number in square brackets, for example [1]. Where appropriate, include the name(s) of editors of referenced books.

1.2.4 Illustrations, graphs, and photographs

All graphics should be centered. Please ensure that any point you wish to make is resolvable in a printed copy of the paper. Resize fonts in figures to match the font in the body text, and choose line widths which render effectively in print. Many readers (and reviewers), even of an electronic copy, will choose to print your paper in order to read it. You cannot insist that they do otherwise, and therefore must not assume that they can zoom in to see tiny details on a graphic.

1.2.5 Color

Color is valuable, and will be visible to readers of the electronic copy. However ensure that, when printed on a monochrome printer, no important information is lost by the conversion to grayscale.

For more suggestions to improve your document, see preparationGuide.pdf

[1]This is what a footnote looks like. It often distracts the reader from the main flow of the argument.

Chapter 2

Chapter Name

Chapter 2 goes here ...

Chapter 3

Chapter Name

Chapter 3 goes here ...

Chapter 4

Chapter Name

Chapter 4 goes here ...

Chapter 5

Conclusions

Conclusion goes here ....

Related Publications

Bibliography

[1] Authors. The frobnicatable foo filter. 2006. ECCV06 submission ID 324. Supplied as additional material `eccv06.pdf`.