Detection and Classification of moving objects with Raspberry-Pi for the Future Cities Project

Fernando Nuno Rodrigues Fernandes da Costa

PREPARATION OF THE MSc DISSERTATION

Master in Electrical and Computers Engineering
Major Telecommunications, Electronics and Computers

Supervisor: Prof. Daniel Cardoso de Moura, PhD (FEUP)
Co-supervisor: Eduardo Marques, Eng (INESC TEC)

February 13, 2013
Abstract

This dissertation proposal aims for the development of a novel method to perform detection and classification of moving objects using a low resources computer module, the Raspberry-Pi platform.

This document summarizes the work done for the course unit “Preparation of the MSc Dissertation” in Electrical and Computer Engineering. It presents the research done and literature reviewed as well as the technologies and tools needed for this project. Additionally, it defines the future work planned for the second semester of the current academic year for the course unit “Dissertation”.
# Contents

Chapter 1..............................................................................................................................................1  
Introduction ........................................................................................................................................1  
  1.1 - Motivation .................................................................................................................................1  
  1.2 - Context ......................................................................................................................................2  
  1.3 - Objectives and Contributions ....................................................................................................2  
  1.4 - Outline ......................................................................................................................................3  
Chapter 2..............................................................................................................................................4  
State of the Art......................................................................................................................................4  
  2.1 - Detection based on image features ............................................................................................4  
  2.2 - Detection based on HOG descriptors ..........................................................................................6  
  2.3 - Detection based on contextual relations ...................................................................................6  
Chapter 3..............................................................................................................................................7  
Work Plan............................................................................................................................................7  
  3.2 - Development Stages ...................................................................................................................8  
Chapter 4..............................................................................................................................................10  
Conclusion ..........................................................................................................................................10  
References .........................................................................................................................................11
List of Figures

Figure 1 - From left to right: HOG descriptor and using HOG to encode a local region of a human sample........................................................................................................................................................................................................................................... 6
Figure 2 - A detailed look into the places of the different components on the RPi .......... 8
Figure 3 - Gantt chart................................................................................................................................................................................................................................................................. 9
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPU</td>
<td>Graphics Processing Unit</td>
</tr>
<tr>
<td>HOG</td>
<td>Histogram of Oriented Gradients</td>
</tr>
<tr>
<td>MP</td>
<td>MegaPixels</td>
</tr>
<tr>
<td>RPi</td>
<td>Raspberry-Pi</td>
</tr>
<tr>
<td>SIFT</td>
<td>Scale-Invariant Feature Transform</td>
</tr>
<tr>
<td>SoC</td>
<td>System on a Chip</td>
</tr>
<tr>
<td>SVM</td>
<td>Support Vector Machine</td>
</tr>
<tr>
<td>P4</td>
<td>Pentium 4</td>
</tr>
<tr>
<td>PC</td>
<td>Personal Computer</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

The detection of moving objects and the ability to classify them into semantically meaningful categories are the cornerstones of a video surveillance system. With the rapid development of video capture technologies, video has become a cheap yet important mean for information record. With a variety of quality and performance options, a video surveillance system can be adapted to different situations and needs. Typical applications include traffic monitoring, crime prevention, trajectory extraction, and so on.

This document presents the dissertation proposal whose overall goal is to develop and implement algorithms capable of detecting and classifying moving objects, specifically people and vehicles, under low computing resources constrains. These algorithms will be developed for a Raspberry-Pi platform, a single-board computer with limited resources, so the processing power and memory available of this platform should be taken into account. Later on, there will be an opportunity to test and evaluate the performance of these algorithms through public image datasets and cameras placed in urban settings in the city of Porto.

This chapter provides a brief background of the Future Cities project, which embraces this dissertation proposal, describes the motivation, the main objectives and the expected contributions of this dissertation. It also presents an outline detailing the organization and content of each remaining chapter of this document.

1.1 - Motivation

The detection and recognition of objects categories is one of the most challenging problems in computer vision, especially in urban areas where the presence of clutter, different viewing angles and lightning may compromise this task. Understanding video objects is of utmost interest due to its fundamental role for the design of an autonomous and robust surveillance system.

With the ongoing improvements in video capturing technologies different approaches have been developed to address this problem. Taking into account quality and performance requirements, several algorithms have been designed with success.

However, a typical surveillance system raises some pertinent issues. Video recording or streaming in public areas can lead to privacy issues, and even when those are overcome, video streaming requires a significant bandwidth or high cost infrastructures in analog systems. A way to avoid these problems is to perform all the information processing locally, so that video
recording or streaming are no longer necessary. Additionally, there is no need of a dedicated server processing the video extracted from several video cameras spread across an area. Furthermore, when trying to monitor greater areas, several systems need to be deployed so that a proper scan can be obtained. This makes low cost technology a necessity.

The system proposed in this dissertation using the Raspberry-Pi [1] platform, a low cost single board computer, meets all these requirements. However, this approach brings up a new challenge, since the RPi has limited resources, it is necessary to study the viability of using a computer with low processing power and limited memory to run detection and classification of moving objects algorithms.

This dissertation focus on overcoming these challenges by developing a low resource, low cost and privacy aware system capable of detecting and classifying moving objects in different urban scenarios.

1.2 - Context

This dissertation proposal arises within the Future Cities Project [2] funded by the European Union under the seventh Framework Programme for Research (FP7) [3]. It takes a multidisciplinary approach, involving research groups from several faculties such as Engineering (FEUP), Psychology (FPCEUP) and Sciences (FCUP). Its key goal is to turn Porto into a smart city, an “urban-scale living lab”, where researchers and companies can develop and test different technologies, exploring areas such as sustainable mobility, urban-scale sensing, safety and privacy, with the common purpose of improving the quality of life for citizens. By providing the city with a wide range of sensors and communication equipment, it hopes to create the necessary conditions for future research and development using advanced technologies for data collection through mobile platforms, wireless communication and large-scale information processing.

Under this project, over the coming months, several video cameras will be placed in the city. One purpose of these cameras will be to detect and identify different moving objects, specifically people and vehicles, so that statistical data can be gathered and trajectories evaluated. The goal of this dissertation will be the development and test of algorithms capable of doing exactly that.

1.3 - Objectives and Contributions

The following objectives are expected to be accomplished at the end of this dissertation:

- To carry out research on object detection and classification techniques;
- To evaluate the possibility of using the RPi as viable technology to do local processing of detection and classification of moving objects;
- To design an algorithm for detection and classification of moving objects, especially people and vehicles, taking into account the computational limitations of the RPi;
- To implement the developed algorithm on a prototype system for detection and classification of moving objects in urban environments;
- To evaluate the performance of the developed system based on the obtained results.

These objectives will lead to the following contributions:

- Further research on the viability of using low resources technology for detection and classification of moving objects with vision-based algorithms;
- Creation of a solution capable of doing all the processing locally without any video recording or streaming;
• Design and implementation of an algorithm on the RPi capable of properly detect and classify moving objects in urban environments;
• Collaborative study and deployment of a low resources and low cost, privacy preserver surveillance system that will be part of the Future Cities project.

1.4 - Outline

The remaining of this dissertation proposal is organized as follows:

Chapter 2 - State of the Art: Presents a detailed overview on different object detection and classification techniques and their algorithms.
Chapter 3 - Work Plan: Describes the main technologies that will take part in this project, and schedules the future work plan.
Chapter 4 - Conclusion: Presents a brief summary of this document and some final considerations about the project.
Chapter 2

State of the Art

The recognition of moving objects is a challenging task, due to the following three aspects. First, the objects have diverse visual appearances and they may vary significantly due to different viewing angles and lighting. This may result in large intra-class variations. Second, the size of the moving objects may change with the distance to the camera. Third, the performance of video object recognition should be real-time so that the system has time to respond to the ongoing events. Thus, to build a robust object detection system all these factors must be taken into account.

There are two main classes of approaches used for localization and categorization of candidate objects. One approach is to directly detect the object in single frames without prior segmentation. The other is to perform object classification on detected moving objects or tracked object sequences. Moving objects can be separated from a static background reasonably by background subtraction. In this technique each video frame is compared against a reference or background model. Pixels in the current frame that deviate significantly from the background are considered to be moving objects. These “foreground” pixels are then further processed for object classification through feature extraction.

The next sections present different methods following these approaches highlighting some of its particular aspects.

2.1 - Detection based on image features

2.1.1 - Shape features

Since humans are capable of recognising a wide variety of objects based on 2-D outline sketches alone, counter-based features have been extensively used for object detection techniques.

Mikolajczyk et al [4] developed a recognising system for poorly textured objects. They introduced a new edge-based local feature detector, an extension of the SIFT descriptor, that is invariant to similarity transformations. These features are localized on edges and a neighborhood is estimated in a scale invariant manner.

Their system begins by clustering on a local transformation to reduce ambiguity, and then estimating a global (affine) transformation to detect the object in an image.
Many recognition systems, like the one stated above, match the image against whole object templates, either for particular rigid objects or articulated ones. These techniques often require a full 3-D model of the object or require a large set of templates to represent all joint object configurations. Alternative approaches use only fragments of contour, guarantying significant advantages since they explicitly account for shape variations and for the randomness in the presence/absence of features due to occlusion and detector errors.

Blake et al [5] followed this solution. Their object detector is built on local contour-based features, local fragments of contour spatially arranged in a star configuration, matched with an oriented chamfer measure, which in boosted combination form a location-specific classifier that is used for detection.

2.1.2 - Texture features

Object shapes in video may change drastically under different camera viewing angles. In addition, the detected shapes may be noised by shadows and other factors. Also, shape-based approaches often require that the scene and camera view for test are very similar to those used for training. Such assumptions are inadequate in real applications. Therefore, new features with other discriminative characteristics, intensity and texture, became an important tool for object detection systems.

Li et al [6] present an appearance-based method to achieve real-time and robust objects classification in diverse camera viewing angles. They propose a new descriptor, the Multi-block Local Binary Pattern (MB-LBP), an extension to the original LBP operator, to capture the large-scale structures in object appearances. Their method begins with a simple background subtraction algorithm, the Gaussian Mixture Model, to detect the moving objects, calculate their bounding boxes and select the detected foreground patches. Each of these patches is normalized to a unified pixel scale and converted to gray-scale, building the training set. The MB-LBP feature is then used to represent the objects’ appearance features. By applying the AdaBoost learning algorithm [7], the most efficient MB-LBP features are finally selected, and a decision function is learned from training data. During this initial training stage, the ECOC-based classifier [8] is used to solve the multi-class problem.

In the test phase, the foreground patches including the moving objects are first detected and normalized to the same unified scale as that for training. Then only the selected features via the AdaBoost learning method are calculated and supplied to the decision function to obtain a class classification.

This system achieved considerable success with precision rates around 85% on the classification of people and different kinds of vehicles. Furthermore, the processing time of the classification method for a 320x240 image resolution was less than 0.1s/frame on a P4 3.0GHz PC.

2.1.3 - Color features

Murphy et al state that any object recognition technique that ignores color features, focusing only on the knowledge gained from the spatial arrangements of the shape features, has significant limitations[9]. They advocate that the representation scheme should carry the color information and its pattern of appearance on the object surface.

In most cases, wherever there is a shape or structural information in the object, the corresponding patterns in the image possess discontinuities in colors. As a result, extraction of information regarding patterns of colors automatically leads to extracting shape and structural information of the object.

They proposed a scheme to describe an object in such a way that the description contains the color information as well as the patterns of colors on the object surface.
2.2 - Detection based on HOG descriptors

Histogram of Oriented Gradient (HOG) descriptors were first introduced in 2005 by Navneet Dalal and Bill Triggs[10]. The essential idea behind this method is that a local object appearance and shape within an image can be characterized by the distribution of local intensity gradients or edge directions. The implementation of these descriptors is done by dividing the image into spatial regions, called cells, which will contain a local 1-D histogram of gradient directions or edge orientations over the pixels of the cell. The combination of these histograms represents the image descriptor (see Figure 1).

In order to reduce the negative effects of illumination and shadowing variance, the local histograms should be contrast-normalize. This can be achieved by measuring the intensity across a larger region of the image, called a block, and using this value to normalize all cells within the block.

In their work, Dalal and Triggs, to perform human detection, tiled the detection window with a dense grid of HOG descriptors and used the combined feature vector in a conventional SVM based window classifier.

This method proved to be so much successful, especially in human detection techniques, that it was the main inspiration for several algorithms later designed.

2.3 - Detection based on contextual relations

Understanding a scene arguably requires parsing the image into its constituent objects. In urban scenarios composed of many different objects, the spatial configuration of one object can facilitate recognition of related ones, and quite often ambiguities in recognition cannot be resolved without looking beyond the spatial extent of the object in question. Thus, algorithms which jointly recognize many objects at once by taking account of contextual relationships have received quite attention.

Desai et al [11] introduced a new method for multi-class object recognition that casts the problem as a structured prediction task. Rather than predicting a single label for each image window independently, their model learns statistics that capture the spatial arrangements of various object classes.
Chapter 3

Work Plan

This chapter presents an overview on how the project development will take place. It describes the main tools that will be used and details the future work plan with the support of a Gantt chart.

3.1 - Tools

This project will be developed on a Raspberry-Pi computer running a Linux operating system using an appropriate Integrated Development Environment (IDE). The programming language will be C/C++ with the addition of the OpenCV library.

3.1.1 - Raspberry-Pi

The Raspberry-Pi is a credit-card sized single-board computer developed in the UK with the intention of promoting the teaching of basic computer science in schools. Being so versatile, it soon became one important tool which can be used in different electronic projects.

The Raspberry-Pi has a Broadcom BCM2835 SoC, which includes an ARM1176JZF-S 700 MHz processor and a Videocore 4 GPU. The more recent model (B) has 512 MB of RAM, 2 USB ports and an Ethernet port. It does not have a built-in hard disk, so it relies on an SD card for booting and persistent storage. It also includes a camera board that connects to the CSI-2 camera port on the Raspberry Pi using a short ribbon cable. It provides connectivity for a camera capable of capturing still images or video recordings. The camera connects to the Image System Pipeline (ISP) in the Raspberry Pi's SoC, where the incoming camera data is processed and eventually converted to an image or video on the SD card. This camera module is capable of taking photos up to 5 MP (2592x1944 pixels) and can record video at resolutions up to 1080p30 (1920x1080x30fps). A detailed diagram denoting the place of the different components on the RPi can be seen below.
3.1.2 - OpenCV

OpenCV (Open Source Computer Vision Library) is an open-source BSD-licensed library for computer vision. It was designed for computational efficiency and focused mainly on real-time image processing, providing several algorithms capable of implementing all kind of image and video manipulation operations. Its main programming language is C++ and it runs on Linux and other operating systems. OpenCV also includes a statistical machine learning library that contains some of the most relevant classifiers to date: Boosting, Support vector machine, among others.

OpenCV has been applied in numerous areas like facial recognition systems, object detection, motion tracking, augmented reality and even interactive art performances.

3.2 - Development Stages

Figure 3 shows the estimated duration of each stage of the development of this dissertation. Although there might be some minor changes to the proposed planning, no significant deviation is expected.

Stage 1. Familiarization with Raspberry-Pi platform: practical experimentation of the Raspberry-Pi computer.

Stage 2. Experimentation with the chosen image processing libraries: several tests will be conducted with the OpenCV library.

Stage 3. Solution specification: a proposed algorithm will be designed aimed to meet the requirements of this project.

Stage 4. Implementation of the previous specified algorithm: deployment of the proposed solution as functional prototype.

Stage 5. Algorithm testing and rectification/optimization with public datasets: the proposed solution will be submitted to testing with public datasets images so that it can be improved.

Stage 6. Algorithm testing and rectification/optimization in a controlled environment: the proposed solution will be submitted to testing within a controlled environment so that it can be improved.

Stage 7. Algorithm testing and evaluation in an urban environment: the proposed solution will be submitted to several tests in urban scenarios, and its performance evaluated.
Stage 8. Dissertation writing and preparation of the scientific paper: after completion of the dissertation writing, a scientific article will be written to demonstrate the results obtained and the contributions of this dissertation.

A Gantt-chart scheduling these stages is presented below.

![Gantt chart](image.png)

*Figure 3 - Gantt chart*
Chapter 4

Conclusion

The detection and classification of moving objects is still one of the most challenging aspects in the field of computer vision. It is a topic in constant evolution not only due to the technology advances that enable the development of increasingly more complex methods, but also due to the necessity of developing more discriminative features and efficient classifiers with lower computational requirements.

The state of the art presented in this document shows some of the most relevant work done in this area. By focusing on these approaches, a new solution is expected to be developed targeting the requisites of this project. However, none of these methods were evaluated on low resources platforms, like the RPi, this should be carefully taken into account when designing the solution.

This dissertation aims for the deployment of an object detection and classification method capable of operating at a low resources computer, the Raspberry-Pi platform. In the first weeks of this project, several approaches will be implemented and tested until a fitting solution that meets the requirements is obtained. Later on, the proposed solution will be submitted to several tests and evaluations under different environments so that it becomes fully functional and able to be an important tool for the Future Cities project.
References


