**Automated Surveillance using Background Subtraction and Shape Analysis**

Final Report for CS39440 Major Project

*Author*: Alexander Shaun O’Donnell ([alo14@aber.ac.uk](mailto:alo14@aber.ac.uk))

*Supervisor*: Bernie Tiddeman (bpt@aber.ac.uk)

16th April 2016

Version 1.0 (Draft)

This report is submitted as partial fulfilment of a BSc degree in  
Computer Science (Inc Integrated Industrial and Professional Training) (G401)

Department of Computer Science

Aberystwyth University

Aberystwyth

Ceredigion

SY23 3DB

Wales, UK

**Declaration of originality**

I confirm that:

* This submission is my own work, except where clearly indicated.
* I understand that there are severe penalties for Unacceptable Academic Practice, which can lead to loss of marks or even the withholding of a degree.
* I have read the regulations on Unacceptable Academic Practice from the University’s Academic Quality and Records Office (AQRO) and the relevant sections of the current Student Handbook of the Department of Computer Science.
* In submitting this work I understand and agree to abide by the University’s regulations governing these issues.

Name Alexander Shaun O’Donnell



Date 16/04/2017

**Consent to share this work**

By including my name below, I hereby agree to this dissertation being made available to other students and academic staff of the Aberystwyth Computer Science Department.

Name Alexander Shaun O’Donnell



Date 16/04/2017

**Acknowledgements**

I am grateful to…

I’d like to thank…

**Abstract**

Computer vision and image interpretation are evolving topics in the world of technology. Many institutions have invested in this field of research for security solutions, as technologies such as facial recognition, object placement and pedestrian detection can provide potential safety nets to the challenges they face. The rise of automation technologies is reducing human error, which is something a high-risk field such as security could do without. This project looks at pedestrian detection/tracking, the idea that a program could take up the role of monitoring live CCTV footage and keep a record of pedestrians and other movements. We can achieve this by using what is known as a classifier, an algorithm that can lead software to make categorical decisions. By providing examples of how pedestrians may appear in a video, the classifier can distinguish between what is and what isn’t a person. Assuming the cameras used in the videos are static, moving objects can be separated from a non-moving environment using background subtraction. Combining these tools, we can extract large moving objects from video footage and use the classifier to see whether the object is a person or not. Rather than having a police officer go through hours of CCTV footage, this program can view the video instead, recording any passenger or movement that occurred with a timestamp of when it happened.

**Contents**

1. Background, Analysis & Process 6

1.1. Background 6

1.1.1. Security and Automation 6

1.1.2. Detection through Vision 7

1.1.3. Pedestrian Detection 8

1.1.4. Motion Detection 9

1.2. Analysis 10

1.2.1. Objectives 10

1.2.2. Issues 10

1.2.3. Security 10

1.3. Process 11

2. Design 12

2.1. Overall Architecture 12

2.2. Detailed Design 12

2.2.1. Even More Detail 12

2.3. User Interface Design 12

2.4. Other Relevant Sections 12

3. Implementation 13

4. Testing 14

4.1. Overall Approach to Testing 14

4.2. Automated Testing 14

4.2.1. Unit Tests 14

4.2.2. User Interface Testing 14

4.2.3. Stress Testing 14

4.2.4. Other Types of Testing 14

4.3. Integration Testing 14

4.4. User Testing 14

5. Critical Evaluation 15

6. Appendices 16

A. Third-Party Code and Libraries 16

B. Ethics Submission 17

C. Code Samples 18

Annotated Bibliography 19

# Background, Analysis & Process

This section should discuss your preparation for the project, including background reading, your analysis of the problem and the process or method you have followed to help structure your work. It is likely that you will reuse part of your outline project specification, but at this point in the project you should have more to talk about.

**Notes:**

* **All of the sections and text in this example are for illustration purposes. The main Chapters are a good starting point, but the content and actual sections that you include are likely to be different.**
* **Look at the document on the Structure of the Final Report for additional guidance.**

## Background

### Security and Automation

As an interdisciplinary field, computer vision has provided many solutions to security. From facial recognition for matching perpetrators to crowd analysis for tackling gang culture and hooliganism, it synergises well with high risk environments. Originating in academic environments, computer vision made its way into the world of security as early as 2001. Police in Tampa, Florida used facial recognition software during the Super Bowl XXXV in January 2001 to detect known criminals and terrorists attempting to attend the event. The software, known as “FaceIt” was developed by Identix and managed to find 19 people with pending arrest warrants [1].

Computer Vision is an abstract topic, the idea of getting a system to understand what is being depicted in an image is particularly useful for security. Many institutions have researched this field as it can provide solutions to some of the challenges faced by the industry. For example, in 2015 the Ministry of Defence (MoD) cooperated with QinetiQ to develop middleware software architecture that would be used alongside sophisticated sensing equipment to perform detection and behaviour classification in a specified area [2].

A typical surveillance setup would involve a security guard or member of law enforcement observing several cameras. These cameras are usually static, where a single camera will maintain the same viewing angle for its duration. This results in institutions investing in many cameras to cover all angles of a premises, and can be straining on the employee tasked to observe them all. This is amplified in isolated or rural environments, where cameras may only observe a small number of pedestrians every hour leading to fatigue and thus an increase in human error. Previous studies have coined the term ‘boredom factor’, where an observer’s levels of concentration degrade over time from viewing the same monotonous feeds for long hour shifts [3].

Part of the job for police officers is to review pre-recorded CCTV footage recorded for potential evidence, with footage lengths ranging from minutes to days. In 2016 CCTV operators from the Suffolk Constabulary and the Metropolitan Police observed more than 1100 hours CCTV footage in the search for a missing person [4]. The hours spent on this particular case involved detection rather than recognition, as they do not include identifying the individuals found. The ability to automate this process would save time, resources and manpower that could be better spent in agencies such as these. It is vital that our law enforcement adapts and becomes more efficient at retrieving and storing information for the sake of the victims and our privacy. To automate the process of collecting timestamps of people appearing in videos, we can look no further than the field of pedestrian detection.

### Detection through Vision

Pedestrian detection is a sub topic of object detection, and has become a fundamental task in video surveillance. The ability to detect and track a human provides great support for reducing trespassers in restricted areas. Typical detection applications in computer vision rely on a classifying algorithm, and some form of machine learning with training/testing phases.

The training phase provides the program with examples of how the object it’s looking for may appear (a person in our case). These examples come in the form of a dataset, a collection of images representing the object to detect. Some datasets include what is known as negative examples, images which either represent something different or don’t include the object altogether. This opens up the classifier to different categories so it can overcome common appearances that interfere with its decision making. For example, including images of cars or animals would help the program differentiate pedestrians from other objects relevant to the environment.

There are a wide range of methods for detecting and classifying objects, many of which are based on the Viola-Jones object detection framework. Although the framework was initially intended for face detection, there have been adaptations to suit pedestrian detection. Viola-Jones is also referred to as a ‘Haar Cascade Algorithm’ since the framework relies on Haar-like feature selection and cascade architecture for classification.

Haar-like feature works by looking at the difference in the sum of the pixel intensities in specific regions of the image. These rectangular regions are positioned relative to a boundary box, this window is scanned across the image to find intensity changes that appear within the regions. Since faces share common appearances such as changes in colour between the eyes and cheek, it is worth positioning the rectangles in the upper areas of the bounding box. In the Fig 1, the rectangle in the middle compares the difference in colour between the eyes/eye brows and the nose/cheeks. The rectangle on the right compares the changes in colour for either side of the nose bridge, assuming that these areas will appear darker.



Fig 1. Rectangular regions used for detecting eyes during face detection. (Source: <http://docs.opencv.org/trunk/d7/d8b/tutorial_py_face_detection.html> )

By applying feature finding techniques to a training set, the program can get a better idea of how faces may appear, as well as a threshold for distinguishing faces from non-faces. There will always be a false positive risk with detection algorithms, for example pareidolic images without a face, that happen to have darker regions where the eyes would be could still possibly be classified as a face. Which is why many classifiers take in a threshold or a Gaussian of intensity values to differentiate faces from noisy images.

Viola-Jones relies on cascade classification, a combination of many classifiers in order of complexity. Each classifier is trained on the same sample, if one classifier rejects the image than the classification breaks out, reducing the processing time so it can search the next sample. This also enables the cascade to focus on more promising regions in the image that are more likely to contain the features, increasing the detection rates as the search space is reduced [5].

### Pedestrian Detection

Pedestrian detection can be achieved using Haar-like features, changes in pixel intensity in simple environments can be seen on people as well as faces. The rectangles can be rearranged within the search window, looking for intensity changes around the head, hands and feet. This assumes the appearance of the person stands out from the environment or that an appropriate threshold is used.

Local Haar features can be combined to build a pedestrian model. Recently, pedestrian detection was achieved by constructing square body shape models from gradient magnitudes and colour channels (Zhang et al, 2014) [6]. Their statistical model takes in images and computes an average edge map of the sample. The resulting edge map is divided into cells, which construct the square regions that make up the body. Their body model is divided into three sections, the head, torso and legs. This is useful as it reduces the search space for features, since it would be safe to assume the feet features always appear below the torso. The core of their project uses multi-channel descriptors for the cells for more information at a local level, and AdaBoost classification, similar to that of Viola-Jones.

Pedestrian detection has seen progress outside of Haar Cascade in recent decades, the increase in publicly available datasets and frameworks has made computer vision more accessible. 2005 saw Dalal & Triggs introduce the histogram of gradients (HOG) feature descriptor which became a landmark for human detection. The descriptor computes the gradients of the input image, and then dividing it into cells. For each cell, a histogram is made of the edge orientations, with the most common orientation representing the direction of the edge within that cell. Linking these orientations together forms a representation of the shape, and exceeds at handling rotations and different human poses [7].

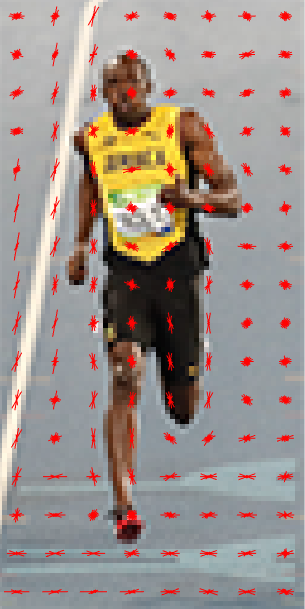


Fig 2. Visual representation of cell directions derived from the HOG feature descriptor. (Source: <http://www.learnopencv.com/histogram-of-oriented-gradients/> )

Detectors such as these begin to use more information from the given images to handle variation in clothing and lighting. These methods demonstrate how effective it is to reduce feature complexity by summarising them into simpler regions. Ultimately making pedestrian detection more adjusted to variations in a person’s appearance. These pedestrian detectors rely on deriving shapes from edges, colour and orientation, we can also derive shape from motion.

### Motion Detection

Points of interest in an image or video are usually moving, with CCTV cameras for example interesting frames usually involve people entering and leaving the perspective. It worth considering the options that motion detection algorithms provide regarding pedestrian detection. As Haar has shown, reducing the search space to regions of interest lead to more computationally efficient programs. Reducing the search space of our pedestrian detector to moving objects could increase its optimisation, as well as reduce the amount of background interference.

One method for detecting motion is background subtraction(BGS), which involves taking the parts of a video that aren’t moving and ignoring them. The moving/non-moving parts in the video are separated, the moving parts are extracted to what is known as the foreground mask. BGS requires a static camera, since it would fail to distinguish anything from a frame where everything is moving. Fortunately, many surveillance systems use on static cameras, and rely on using many cameras to cover key areas of the perimeter. A simple BGS formula would simply subtract the change in pixel intensities between the current frame and the previous frame.



Fig 3. BGS being run on the CAVIAR dataset.

You also need to consider lighting and animated background features. An environment that has a constant shift in lighting effects may return a poor foreground mask. A BGS camera setup of a road with street lamps would struggle with lighting changes as well as shadows emanating from objects at certain angles. Meanwhile a coastal area would have light reflecting of the waves further distorting the mask. Considering not all moving feature of an image are of interest, animated background areas may cause issues. Things such as waves, flags, vehicles and small animals would all appear on the foreground mask if no filtering process was present. Many of these issues can be filtered or reduced through parameters and noise cancelling.

To determine whether a pixel is counted as moving, more advanced methods rely on a ‘history’ parameter. This parameter is usually arbitrary, if something moves in that pixel, the history represents the number of time/frames that pass before the pixel is counted as static. Reducing the history value could tackle the lighting issue, depending how slow the change in lighting is. If the foreground mask is only registering quicker movements, than slower changes in lighting could be ignored.

For the non-interest animated background features, a threshold can be introduced. This threshold would define how much the pixel intensity needs to change between frames before the pixel can be seen as ‘moving’. Through thresholding the foreground mask, smaller movements within the picture such as shadows can be ignored. Modern BGS functions can adapt this threshold to deal with long term changes within the background, such as objects being placed in the scene and day/night cycles.

Another way of detecting motion is using optical flow. Similar to the visualisation in Fig 2, regions within an image are said to have a direction. An optical flow method works out the direction of travel for each cell of an image. Optical flow methods are useful as they can show the strength of the direction as well as it’s orientation, which gives information on how fast the motion is. Optical flow is more suited to dynamic environments, since it can quickly adapt if the camera moves.

One of the biggest issues in computer vision is tackling noise. Image noise is unwanted side effects that occur when visualising a computer vision algorithm, and has a wide range of forms and causes. Examples of noise include speckles, film grain and static, which all add extraneous information. The most common cause of noise is the use of low resolution cameras, a lower number of pixels means that noisy areas will have a greater effect on the overall image quality. It is important to consider noise reduction techniques as clearer images which produce more obvious features and shapes, making easier for algorithms to find them.

Erosion morphology can be used to reduce the amount of speckle noise in BGS. Assume it is being applied to binary foreground mask, where pixel greyscale values are either 0 (non-moving) or 255 (moving). Erosion uses a structuring element to determine which of the current pixel’s neighbours must be 255, for its value to still be 255 after processing. Erosion examines the neighbouring pixels of a selected pixel, if the current pixel has at least one neighbour with a value of 0, the pixel’s value becomes 0.

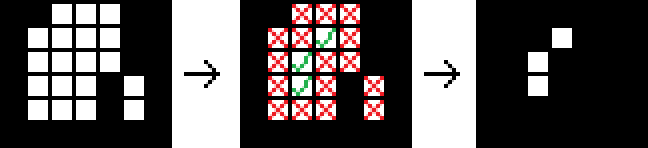


Fig 4. Applying a 3x3 structuring element across an image, a 255 pixel (white) must have the pixels adjacent to it be 255 to survive erosion.

## Analysis

### Objectives

The aim of this project is to automate the process of detecting and tracking pedestrians within a video. From loading the project, the user will be able to select a video to analyse, gaining live feedback of the programs interpretation of what is going on. The program must be able to somehow document the pedestrians/features it has witnessed throughout the video. Rather than having security officers go through hours and hours of footage looking for people, the program will have a full list of the pedestrians that appeared with timestamps. It must have some sort of interpretation or knowledge of what it’s looking at, so it can distinguish pedestrians from noise. This would require the code to implement some sort of basic training stage so it can construct its own idea of what a pedestrian looks like. It should return accurate cut outs of pedestrians that appear in the video, rather than printing out every single entire frame as this would show it has poor reliability.

With the amount of freely available datasets and open source frameworks, it would be ideal to use proven computer vision functions for this project. OpenCV provides a wide range of implementations for feature selection methods such as Haar, as well as basic image processing/video reading functions that would be useful for the project.

BGS, ground truth shape analysis, Feature selection (tested haar, no more hog, so try my own methods combined with BGS, BGS helps reduce false positives), classifier. Consider splitting choices of combining BGS/feature selection, basic classification. OpenCVs BGS equation

### Datasets and Frameworks

### Motion Detection Choice

### Feature Selection and Classification

### Issues

BIG EFFORT TO REDUCE FALSE POSITIVES, false negatives not so bad but cause loss of trust. Racial profiling, Difficulties such as orientation. Filtering smaller motion.

### Security

Taking into account the problem and what you learned from the background work, what was your analysis of the problem? How did your analysis help to decompose the problem into the main tasks that you would undertake? Were there alternative approaches? Why did you choose one approach compared to the alternatives?

There should be a clear statement of the objectives of the work, which you will evaluate at the end of the work.

In most cases, the agreed objectives or requirements will be the result of a compromise between what would ideally have been produced and what was felt to be possible in the time available. A discussion of the process of arriving at the final list is usually appropriate.

As mentioned in the lectures, think about possible security issues for the project topic. Whilst these might not be relevant for all projects, do consider if there are relevant for your project. Where there are relevant security issues, discuss how they will this affect the work that you are doing. Carry forward this discussion into relevant areas for design, implementation and testing. ALTERNATIVE FEATURE SELECTION, SOME CLASSIFICATION. ASSUMPTIONS, compare problems of surveillance to pedestrian detection solutions, motion doesn’t interpret what it’s looking at like Haar/HOG does (simply a tool for getting information) False positives a breach of rights?. MOTION. Difficulties such as orientation. Filtering smaller motion. Frameworks/dataset assistance SECURITY. OpenCVs BGS equation

## Process

You need to describe briefly the life cycle model or research method that you used. You do not need to write about all of the different process models that you are aware of. Focus on the process model that you have used. It is possible that you needed to adapt an existing process model to suit your project; clearly identify what you used and how you adapted it for your needs. FDD GITHUB DIARY LOG. C++ VISUAL STUDIO.

# Design

You should concentrate on the more important aspects of the design. It is essential that an overview is presented before going into detail. As well as describing the design adopted it must also explain what other designs were considered and why they were rejected.

The design should describe what you expected to do, and might also explain areas that you had to revise after some investigation.

Typically, for an object-oriented design, the discussion will focus on the choice of objects and classes and the allocation of methods to classes. The use made of reusable components should be described and their source referenced. Particularly important decisions concerning data structures usually affect the architecture of a system and so should be described here.

How much material you include on detailed design and implementation will depend very much on the nature of the project. It should not be padded out. Think about the significant aspects of your system. For example, describe the design of the user interface if it is a critical aspect of your system, or provide detail about methods and data structures that are not trivial. Do not spend time on long lists of trivial items and repetitive descriptions. If in doubt about what is appropriate, speak to your supervisor.

You should also identify any support tools that you used. You should discuss your choice of implementation tools - programming language, compilers, database management system, program development environment, etc.

Some example sub-sections may be as follows, but the specific sections are for you to define.

## Overall Architecture

## Detailed Design

### Even More Detail

## User Interface Design

## Other Relevant Sections

# Implementation

The implementation should look at any issues you encountered as you tried to implement your design. During the work, you might have found that elements of your design were unnecessary or overly complex; perhaps third party libraries were available that simplified some of the functions that you intended to implement. If things were easier in some areas, then how did you adapt your project to take account of your findings?

It is more likely that things were more complex than you first thought. In particular, were there any problems or difficulties that you found during implementation that you had to address? Did such problems simply delay you or were they more significant?

You can conclude this section by reviewing the end of the implementation stage against the planned requirements.

# Testing

Detailed descriptions of every test case are definitely not what is required here. What is important is to show that you adopted a sensible strategy that was, in principle, capable of testing the system adequately even if you did not have the time to test the system fully.

Provide information in the body of your report and the appendix to explain the testing that has been performed. How does this testing address the requirements and design for the project?

How comprehensive is the testing within the constraints of the project? Are you testing the normal working behaviour? Are you testing the exceptional behaviour, e.g. error conditions? Are you testing security issues if they are relevant for your project?

Have you tested your system on “real users”? For example, if your system is supposed to solve a problem for a business, then it would be appropriate to present your approach to involve the users in the testing process and to record the results that you obtained. Depending on the level of detail, it is likely that you would put any detailed results in an appendix.

The following sections indicate some areas you might include. Other sections may be more appropriate to your project.

## Overall Approach to Testing

## Automated Testing

### Unit Tests

### User Interface Testing

### Stress Testing

### Other Types of Testing

## Integration Testing

## User Testing

# Critical Evaluation

Examiners expect to find in your dissertation a section addressing such questions as:

* Were the requirements correctly identified?
* Were the design decisions correct?
* Could a more suitable set of tools have been chosen?
* How well did the software meet the needs of those who were expecting to use it?
* How well were any other project aims achieved?
* If you were starting again, what would you do differently?

Other questions can be addressed as appropriate for a project.

Such material is regarded as an important part of the dissertation; it should demonstrate that you are capable not only of carrying out a piece of work but also of thinking critically about how you did it and how you might have done it better. This is seen as an important part of an honours degree.

There will be good things and room for improvement with any project. As you write this section, identify and discuss the parts of the work that went well and also consider ways in which the work could be improved.

In the latter stages of the module, we will discuss the evaluation. That will probably be around week 9, although that differs each year.

# Appendices

The appendices are for additional content that is useful to support the discussion in the report. It is material that is not necessarily needed in the body of the report, but its inclusion in the appendices makes it easy to access.

For example, if you have developed a Design Specification document as part of a plan-driven approach for the project, then it would be appropriate to include that document as an appendix. In the body of your report you would highlight the most interesting aspects of the design, referring your reader to the full specification for further detail.

If you have taken an agile approach to developing the project, then you may be less likely to have developed a full requirements specification. Perhaps you use stories to keep track of the functionality and the ’future conversations’. It might not be relevant to include all of those in the body of your report. Instead, you might include those in an appendix.

There is a balance to be struck between what is relevant to include in the body of your report and whether additional supporting evidence is appropriate in the appendices. Speak to your supervisor or the module coordinator if you have questions about this.

* 1. Third-Party Code and Libraries

If you have made use of any third party code or software libraries, i.e. any code that you have not designed and written yourself, then you must include this appendix.

As has been said in lectures, it is acceptable and likely that you will make use of third-party code and software libraries. If third party code or libraries are used, your work will build on that to produce notable new work. The key requirement is that we understand what is your original work and what work is based on that of other people.

Therefore, you need to clearly state what you have used and where the original material can be found. Also, if you have made any changes to the original versions, you must explain what you have changed.

As an example, you might include a definition such as:

**Apache POI library** – The project has been used to read and write Microsoft Excel files (XLS) as part of the interaction with the client’s existing system for processing data. Version 3.10-FINAL was used. The library is open source and it is available from the Apache Software Foundation [5]. The library is released using the Apache License [6]. This library was used without modification.

* 1. Ethics Submission

This appendix includes a copy of the ethics submission for the project. After you have completed your Ethics submission, you will receive a PDF with a summary of the comments. That document should be embedded in this report, either as images, an embedded PDF or as copied text. The content should also include the Ethics Application Number that you receive.

* 1. Code Samples

This is an example appendix. Include as many appendices as you need. The appendices do not count towards the overall word count for the report.

For some projects, it might be relevant to include some code extracts in an appendix. You are not expected to put all of your code here - the correct place for all of your code is in the technical submission that is made in addition to the Final Report. However, if there are some notable aspects of the code that you discuss, including that in an appendix might be useful to make it easier for your readers to access.

As a general guide, if you are discussing short extracts of code then you are advised to include such code in the body of the report. If there is a longer extract that is relevant, then you might include it as shown in the following section.

Only include code in the appendix if that code is discussed and referred to in the body of the report.

Random Number Generator

The Bayes Durham Shuffle ensures that the pseudo random numbers used in the simulation are further shuffled, ensuring minimal correlation between subsequent random outputs.

// Some example code here…

# Annotated Bibliography

This final section should list all relevant resources that you have consulted in researching your project. Each reference should also include a brief annotation.

1. John Perdikaris, “*Physical Security and Environmental Protection*”, January 2014, Chapter 5.3 Surveillance and Counter Surveillance, page 147.
2. UK Government, “*Plug and Play Autonomous Sensors*”, 2015. Available: <https://www.gov.uk/government/news/plug-and-play-autonomous-sensors>. [Accessed 06 February 2017].

A press release discussing the development of new concepts in modular autonomous sensing. Provides details on some of the technologies used.

1. Gavin J.D Smith, “*Behind the Screens: Examining Constructions of Deviance and Informal Practices among CCTV Control Room Operators in the UK*”, pages 388-389, 2004. Available: <http://surveillance-and-society.org/articles2(2)/screens.pdf>. [Accessed 18 April 2017].
2. BBC, "*Missing Corrie Mckeague: Suffolk Police ‘search lacks resources’”,* November 2016. Available: <http://www.bbc.co.uk/news/uk-england-suffolk-38052508>. [Accessed 18 April 2017]
3. Paul Viola & Michael J. Jones, “*Robust Real Time Face Detection”,* received September 2001, pages 138-139. Available: <http://www.vision.caltech.edu/html-files/EE148-2005-Spring/pprs/viola04ijcv.pdf>. [Accessed 19 April 2017]
4. Shanshan Zhang, Christian Bauckhage, Armin B. Cremers, “*Informed Haar-like Features Improve Pedestrian Detection*”, 2014 IEEE Conference on Computer Vision and Pattern Recognition, pages 949-950.
5. Navneet Dalal & Bill Triggs, “*Histograms of Oriented Gradients for Human Detection*”, 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05), Chapters 3 & 6.4.
6. CREDIT CAVIAR/DAIMLER OPENCV