Object Detection, Recognition and Re-identification in Video Footage

by

Martins E. Irhebhude

A Doctoral Thesis

Submitted in partial fulfilment
of the requirements for the award of

Doctor of Philosophy
of
Loughborough University

6th October 2015

Abstract

There has been a significant number of security concerns in recent times; as a result, security cameras have been installed to monitor activities and to prevent crimes in most public places.These analysis are done either through video analytic or forensic analysis operations on human observations. To this end, within the research context of this thesis, a proactive machine vision based military recognition system has been developed to help monitor activities in the military environment. The proposed object detection, recognition and re-identification systems have been presented in this thesis.

A novel technique for military personnel recognition is presented in this thesis. Initially the detected camouflaged personnel are segmented using a *grabcut* segmentation algorithm. Since in general a camouflaged personnel's uniform appears to be similar both at the top and the bottom of the body, an image patch is initially extracted from the segmented foreground image and used as the region of interest. Subsequently the colour and texture features are extracted from each patch and used for classification. A second approach for personnel recognition is proposed through the recognition of the badge on the cap of a military person. A feature matching metric based on the extracted Speed Up Robust Features (SURF) from the badge on a personnel's cap enabled the recognition of the personnel's arm of service.

A state-of-the-art technique for recognising vehicle types irrespective of their view angle is also presented in this thesis. Vehicles are initially detected and segmented using a Gaussian Mixture Model (GMM) based foreground/background segmentation algorithm. A Canny Edge Detection (CED) stage, followed by morphological operations are used as pre-processing stage to help enhance foreground vehicular object detection and segmentation. Subsequently, Region, Histogram Oriented Gradient (HOG) and Local Binary Pattern (LBP) features are extracted from the refined foreground vehicle object and used as features for vehicle type recognition. Two different datasets with variant views of front/rear and angle are used and combined for testing the proposed technique.

For night-time video analytics and forensics, the thesis presents a novel approach to pedestrian detection and vehicle type recognition. A novel feature ac-

quisition technique named, CENTROG, is proposed for pedestrian detection and vehicle type recognition in this thesis. Thermal images containing pedestrians and vehicular objects are used to analyse the performance of the proposed algorithms. The video is initially segmented using a GMM based foreground object segmentation algorithm. A CED based pre-processing step is used to enhance segmentation accuracy prior using Census Transforms for initial feature extraction. HOG features are then extracted from the Census transformed images and used for detection and recognition respectively of human and vehicular objects in thermal images.

Finally a novel technique for people re-identification is proposed in this thesis based on using low-level colour features and mid-level attributes. The low-level colour histogram bin values were normalised to 0 and 1. A publicly available dataset (VIPeR) and a self constructed dataset have been used in the experiments conducted with 7 clothing attributes and low-level colour histogram features. These 7 attributes are detected using features extracted from 5 different regions of a detected human object using an SVM classifier. The low-level colour features were extracted from the regions of a detected human object. These 5 regions are obtained by human object segmentation and subsequent body part sub-division. People are re-identified by computing the Euclidean distance between a probe and the gallery image sets. The experiments conducted using SVM classifier and Euclidean distance has proven that the proposed techniques attained all of the aforementioned goals. The colour and texture features proposed for camouflage military personnel recognition surpasses the state-of-the-art methods. Similarly, experiments prove that combining features performed best when recognising vehicles in different views subsequent to initial training based on multi-views. In the same vein, the proposed CENTROG technique performed better than the state-of-the-art CENTRIST technique for both pedestrian detection and vehicle type recognition at night-time using thermal images. Finally, we show that the proposed 7 mid-level attributes and the low-level features results in improved performance accuracy for people re-identification.

Acknowledgments

I would like to thank my supervisor, Prof Eran A. Edirisinghe, for the patient guidance, support, encouragement and advice throughout my time as a student. I have been extremely lucky to have a supervisor who cared so much about me and my work, his patience and support, when it seemed to be impossible, helped in the realisation of the research goal. I am particularly grateful to Dr Ana Salagean for her support in my study. My sincere thanks also go to Dr Helmut Bez and Dr Shaheen Fatima for their annual review on my research progress and advice given. I also appreciate the support and friendship provided by the academic and secretarial staff of the department of Computer Science, Loughborough University. To all I say thank you.

I must express my gratitude to Hyelni, my wife, Osose, Ese and Oseme, my daughters, for their support and encouragement. I was particularly amazed by Hyelni's willingness to proofread the pages of my thesis. I would also like to say a big thank you to my beloved mother, in-laws, brothers and sisters for their prayers and care before and through my study.

My heartfelt thanks go to my research colleagues of Computer Science Department, Loughborough University. I would also like to thank the staff members of Computer Science Department, Nigerian Defence Academy, Kaduna, Nigeria.

Finally, I would like to thank the Nigerian Defence Academy, Kaduna, Nigeria for the financial support provided through my research study through the Tertiary Education Trust Fund (TETFUND) intervention.

Martins E. Irhebhude, October 2015

Contents

Abstract ii

Acknowledgments iv

List of Abbreviations xiv

1 Introduction 1

1.1 Research Motivation . . . 3
1.2 Research Aim and Objectives . . . 4
1.3 Scholarly Contributions . . . 5
1.4 Thesis Layout . . . 7

2 Literature Review 8

2.1 Introduction . . . 8
2.2 Object Detection and People Identification . . . 9
2.3 Vehicle Detection and Recognition . . . 13
2.4 Pedestrian Detection and Vehicle Type Recognition in Night-time . 15
2.5 People Re-identification . . . 16

3 Background of Study 19

3.1 Red Green Blue (RGB) colour model . . . 19
3.1.1 Selected feature descriptors from RGB colour . . . 20
3.1.1.1 Normalised 2D Histogram . . . 20
3.1.1.2 Local shape features . . . 21
3.1.2 Converting from RGB to other colour models . . . 22
3.1.2.1 The Hue Saturation Intensity (HSI) colour model . 22
3.2 Canny Edge Detection . . . 24
3.3 Gray Level Co-occurrence Matrix (GLCM) . . . 25
3.4 Speed Up Robust Features (SURF) . . . 26
3.4.1 Interest point localisation . . . 27
3.4.2 Interest point description . . . 27
3.5 Histogram of Oriented Gradients (HOG) . . . 28

3.6 Local Binary Patterns . . . 30
3.7 Gaussian Mixture Model (GMM) . . . 31
3.8 Census Transform . . . 32
3.9 Correlation-based Feature Selection (CFS) . . . 33
3.10 Support Vector Machine (SVM) . . . 34
3.10.1 Binary classification . . . 34
3.10.2 Multi-class classification . . . 34
3.11 Receiver Operating Characteristics (ROC) curves . . . 35
3.12 Summary . . . 36

4 Recognition of Military Personnel 37
4.1 Introduction . . . 37
4.2 The Proposed System . . . 38
4.2.1 Pre-processing . . . 39
4.2.2 Feature extraction for camouflage type identification . . . 41
4.2.3 Feature selection . . . 42
4.2.4 Recognition of arm of service . . . 42
4.2.5 Recognition of arm of service of a plain cap . . . 43
4.3 Experimental Setup . . . 43
4.4 Experimental Result and Analysis . . . 44
4.4.1 Camouflaged type recognition army, navy, air force . . . 44
4.4.2 Cap type recognition camouflaged vs plain . . . 46
4.4.3 Plain cap type recognition using badges . . . 47
4.4.4 Effect on hue, saturation and texture on camouflage type recognition . . . 53
4.4.5 Comparison of proposed vs benchmark algorithms . . . 55
4.5 Conclusions . . . 62

5 Vehicle Type Recognition 64
5.1 Introduction . . . 64
5.2 Research Methodology . . . 66
5.2.1 Vehicular object segmentation . . . 68
5.2.2 Feature extraction . . . 69
5.2.2.1 Region descriptors/features . . . 70
5.2.2.2 HOG features . . . 71
5.2.2.3 LBP features . . . 71
5.2.2.4 Feature combination . . . 71
5.2.3 Feature selection . . . 71
5.3 Experimental Analysis . . . 72

5.3.1 Experiments on the front and rear view dataset 73
5.3.2 Experiments on angular view dataset 74
5.3.3 Analysis of results 75
5.4 Conclusion 84

6 Night-time Detection and Recognition 85
6.1 Introduction 85
6.2 CENTRIST and CENTROG Descriptors 86
6.3 Proposed System Description 88
6.3.1 Pedestrian detection 88
6.3.2 Vehicle type recognition 89
6.4 Experiments and Performance Analysis Results 92
6.4.1 Pedestrian detection experiments 92
6.4.2 Experiments on vehicle type recognition 95
6.4.3 Performance evaluation using ROC curves 98
6.5 Conclusion 101

7 People Re-identification by Low-Level Features and Mid-level Attributes 103
7.1 Introduction 103
7.2 Proposed Re-identification Framework 105
7.3 Human Body Part-based Feature Representation 106
7.3.1 Body region segmentation and sub-division 106
7.3.2 Low-level feature extraction and representation 106
7.4 Clothing Attribute Representation 107
7.4.1 Clothing attribute value determination 108
7.4.2 The combined feature vector 108
7.5 Experiments 108
7.5.1 Self-captured dataset 109
7.5.2 The VIPeR dataset 109
7.5.3 Evaluation and metrics used 110
7.6 Experimental Results and Analysis 111
7.6.1 Attributes detection 111
7.6.2 Matching performance analysis 112
7.7 Conclusion 116

8 Conclusions and Future Work 117
8.1 Conclusion 118
8.2 Future Work 120

References **122**

List of Figures

1.1 Object recognition paradigm 2

3.1 RGB Channels Separation: © Nevit Dilmen at Wikimedia (https://commons.wikimedia.org/wiki/File:RGB_channels_separation.png) 20
3.2 Shape patterns in [68] 21
3.3 The Hue Saturation Intensity (HSI) model 22
3.4 Pixel values with the GLCM representation 26
3.5 Angle and Distance between pixel 26
3.6 SURF box filter ©Copyright 2013, Alexander Mordvintsev Abid K. Last updated on Oct 31, 2014. 27
3.7 SURF orientation graph ©Copyright 2013, Alexander Mordvintsev Abid K. Last updated on Oct 31, 2014. 28
3.8 Cells and Overlapping Blocks 29
3.9 Cells and Overlapping Blocks in [27] 29

4.1 An overview of the proposed method for military personnel recognition 38
4.2 Detected people results 39
4.3 Segmented foreground images using grabcut 40
4.4 The two segmented body parts of a detected human 40
4.5 Examples of camouflage image patches 40
4.6 Examples of plain image patches 41
4.7 Classifiers and feature selection algorithms comparison model 43
4.8 Air Force cap badge matching 48
4.9 Army cap badge matching 49
4.10 Navy cap badge matching 50
4.11 Using SURF to recognise Air Force cap badge type 51
4.12 Using SURF to recognise Army cap badge type 52
4.13 Using SURF to recognise Navy cap badge type 53

4.14 Experimental Results for camouflage classification using different combinations of colour channels and GLCM texture Features. . . . 55
4.15 Recognition accuracies including area under curve (AUC) for various techniques using SVM classifier. 56
4.16 TP, FP rates, precision, recall, F-measure and ROC area plot of proposed technique vs methods in [58] and [68] 60
4.17 Proposed technique vs [58] and [68] techniques 62

5.1 Proposed methodology for vehicle classification 67
5.2 Diagram shows original vehicle, after edge detection, after removing extra edges and after dilate and fill operations respectively 68
5.3 Training samples some segmented vehicles from front/rear view dataset 69
5.4 Training samples some segmented vehicles from angular view dataset 69
5.5 Some examples of recognised vehicle from front/rear dataset 74
5.6 Some examples of recognised vehicle from angular dataset 75
5.7 Accuracy on both datasets 77
5.8 Speed of processing: whole vs selected. Notations used: C - combined view, FR - frontrear view, A - angular view 78
5.9 Weighted average plot of TP, FP rates, precision, recall and ROC area for RLH, HOG, Region and LBP feature sets 81
5.10 ROC curve of front/rear view datasets 82
5.11 ROC curve of angular view datasets 82
5.12 ROC curve of combined view datasets 83

6.1 Samples showing original image with processed images after CT, edge and HOG operations (a) Pedestrian sample, (b) Car sample, (c) Truck sample 87
6.2 Proposed night-time pedestrian detection technique 89
6.3 Proposed night-time vehicle classification technique 91
6.4 Some examples of extracted 20 × 40 pixel pedestrian and non-pedestrian images 93
6.5 Detected pedestrians using (a) CENTROG and (b) CENTRIST . . 94
6.6 Some examples of extracted segmented vehicles 96
6.7 Classified vehicles using CENTROG feature descriptor 96
6.8 Classified vehicles using CENTRIST feature descriptor 97
6.9 ROC curves showing the performance of CENTROG vs CENTRIST feature descriptors on the pedestrian detection experiment 98
6.10 ROC curves showing the performance of CENTROG vs CENTRIST feature descriptors on the vehicle type classification experiment . . 99

6.11 Performance analysis plot on pedestrian detection 100
6.12 Performance analysis plot on vehicle type recognition 100

7.1 Human re-identification process 105
7.2 Proposed system for person re-identification 105
7.3 Low-level feature concatenation 107
7.4 Definition of medium-level attributes 107
7.5 Total feature length 108
7.6 Samples from the self-captured data set 109
7.7 VIPeR data samples 109
7.8 Cumulative matching characteristic curves of proposed technique . . 113
7.9 Cumulative matching characteristic curves of proposed technique plotted within the narrow range of Rank-1 to Rank-20 113
7.10 Rank scores re-identification performance 115
7.11 Human re-identification on both datasets 115

List of Tables

3.1 RGB values of some basic colours in [32] 20
3.2 GLCM common angles 26

4.1 Classification Accuracy using four classifiers 44
4.2 SVM parameters 44
4.3 True, false positive rates, precision, recall, F-Measure and ROC area performance values 45
4.4 Confusion matrix for personnel recognition 46
4.5 Selected Features using CFS on the proposed feature sets 47
4.6 Experimental Results for camouflage classification using different combinations of colour channels and GLCM texture Features. . . . 54
4.7 Recognition accuracies for various techniques using SVM classifier . 56
4.8 True, false positive rates, precision, recall, F-Measure and ROC area performance values for whole feature sets when using different approaches. 57
4.9 True, false positive rates, precision, recall, F-Measure and ROC area performance values for selected feature sets when running different approaches. 59
4.10 Confusion matrix for personnel recognition using whole and selected feature sets on different experiments 61

5.1 Classification accuracy results with selected features 76
5.2 Confusion matrix for Angular view dataset using RLH feature . . . 77
5.3 Confusion matrix for F/R view dataset using RLH feature 77
5.4 Speed of processing using varying feature attributes 78
5.5 True, false positive rates, precision, recall, and ROC area performance values for angular view dataset 79
5.6 True, false positive rates, precision, recall, and ROC area performance values for F/R view dataset 79
5.7 True, false positive rates, precision, recall, and ROC area performance values for combined view dataset 80

5.8 Weighted average of true, false positive rates, precision, recall, and ROC area performance values on all datasets 81
5.9 F-measure recognition percentages 84

6.1 Pedestrian Camera parameter 92
6.2 True, false positive rates, precision, recall, F-measure and ROC area performance values for pedestrian detection 95
6.3 Confusion matrix for pedestrian detection 95
6.4 Vehicle Camera parameter 95
6.5 True, false positive rates, precision, recall, F-measure and ROC area performance values for vehicle type recognition 97
6.6 Confusion matrix for vehicle type recognition 98
6.7 Performance analysis on pedestrian detection 99
6.8 Performance analysis on vehicle classification 100
6.9 Processing time and accuracy rates after feature selection for both pedestrian detection and vehicle type recognition 101

7.1 Attributes description and values 108
7.2 Attributes classification accuracies based on VIPeR dataset 112
7.3 Person re-identification accuracy 114

List of Abbreviations

AOG - AND-OR Graph
AUC - Area Under Curve
BTF - Brightness Transfer Function
CED - Canny Edge Detection
CENTRIST - **CEN**sus **TR**ansformed h**IST**ogram
CENTROG - **CEN**sus **T**ransformed histog**R**am **O**riented **G**radient
CFS - Correlation-based Feature Selection
CMC - Cumulative Matching Characteristics
COV - Covariance Tensor Feature
CSD - Colour Structure Descriptor
CT - Census Transform
CV - Combined View DCD - Dominant Colour Descriptor
DRLBP - Discriminant Robust Local Binary Pattern
DRLTP - Discriminant Robust Local Ternary Pattern
DT - Dynamic Texture
ELTP - Extended Local Ternary Pattern
ETC - Electronic Toll Collection
FLD - Fisher Linear Discriminant
FP - False Positive
FPS - Frame Per Second
F/R - Front/Rear
GLCM - Gray Level Co-occurrence Matrix
GLDM - Gray Level Dependency Matrix
GMM - Gaussian Mixture Model
HD - High Definition
HDBN - Hybrid Dynamic Bayesian Network
HDR - High Dynamic Range
HFOF - High Frequency Optical Flow
HMMEM - Hidden Markov Model Expectation Maximisation
HOG - Histogram Oriented Gradient
HSI - Hue Saturation Intensity

HVS - Human Visual System
KNN - K-Nearest Neighbour
KPCA - Kernel Principal Component Analysis
LBP - Local Binary Pattern
LH - Local Binary Pattern Histogram Oriented Gradient
LTP - Local Ternary Pattern
MCM - Multiple Component Matching
MDA - Multiple Discriminate Analysis
MDMO - Mean Displacement Mean Orientation
MHOF - Multi-scale Histogram Optical Flow
MKL - Multiple Kernel Learning
MRCG - Mean Riemanian Covariance Grid
MRC - Mean Riemanian Covariance
NN - Neural Network
ODMO - χ^2-Optimal Displacement & Mean Orientation
ODOO - χ^2-Optimal Displacement & χ^2-Optimal Orientation
PCA - Principal Component Analysis
PCH - Probability Colour Histogram
PLS - Partial Least Square
RGB - Red Green Blue
RL - Region Local Binary Pattern
RLH - Region Local Binary Pattern Histogram Oriented Gradient
ROC - Receiver Operating Characteristics
ROI - Region Of Interest
SAR - Synthetic Aperture Radar
SCD - Shape Contest Descriptor
SDALF - Symmetry Driven Accumulation of Local Features
SIFT - Scale Invariant Feature Transform
SLBP - Simplified Local Binary Pattern
SQD - Sum of Quadratic Difference
SVM - Support Vector Machine
SURF - Speed Up Robust Features
TCS - Target Colour Structure
TP - True Positive
UELTP - Uniform Extended Local Ternary Pattern
ULBP - Uniform Local Binary Pattern
VIPeR - View Invariant Pedestrian Recognition
WSSIM - Weighted Structural Similarity

Chapter 1

Introduction

The integration of effective image capture, enhancement and processing techniques with computer vision and machine intelligence technologies have recently revolutionised intelligent systems that are able to mimic human behaviour, instinct and decision making power. Such integrated intelligent technologies have recently found their way through to many application domains ranging from smart phones, consumer electronic devices, electronic toys, automated vehicles, security and surveillance systems, entertainment and games industry, to name a few. In general, any application that can be served by the combined power and capability of the human eyes and brain, can be served with such technologies. The continuing challenge is to try and push the boundaries towards human like capabilities, meeting realistic sensitivity, fidelity and accuracy constraints.

Rightly serving the above advance of technology, many fundamental areas of digital imaging, computer vision and machine learning algorithms have shown rapid advances. Digital cameras now have the capability of recording in High Definition (HD) and High Dynamic Range (HDR) format, mimicking the Human Visual System (HVS), close to perfection. With the cost of cameras with such capability rapidly decreasing, it has become a de-facto assumption that the captured images are of perfect quality, noise free, super resolution and of perfect clarity, for example. In an intelligent imaging system once a perfect quality image is captured, it needs to be processed to enhance quality further and to remove any artefact that would be created by the captured devices. Image processing algorithms such as colour correction/constancy, distortion removal, white balancing, sharpening, de-blurring, filtering and morphological algorithms are able to further improve the quality of images, removing any artefacts that may be introduced due to any limitation of the image capture system. In the next stage, when the images are processed, one needs computer vision algorithms such as motion detection, foreground object extraction, object detection, anti-shake algorithms etc., to detect objects of importance in the scene. Once such objects are detected, recognising

their type and analysing their behaviour requires machine learning algorithms with embedded learning capability and built-in intelligence. Although significant advances have been made in computer vision and machine learning algorithms, either the limitations of the algorithms themselves or the complicated nature of their application within practical systems, still leave many areas within the above technology areas, open for further research and development. This thesis focuses in general on pushing the boundaries in object detection, object recognition, tracking and re-identification technologies. A number of original contributions in these areas have been made that are presented in this thesis. A particular emphasis is also given to night-time video analysis.

In general an image processing pipeline that is used to recognise an object (see figure 1.1), goes through three main functional stages after image capture and preprocessing, namely, feature extraction, feature selection and object classification. Prior to recognising the object, one needs to detect the object. However prior to detecting the object one has to capture features of the object that can be compared against the features of a known object of a particular type. Object recognition on the other hand is a step beyond object detection within a hierarchy of an object recognition task. It involved using a learning algorithm that can be taught to recognise objects based on a training phase during which known features of each identifiable object is given with an indication of their origin. In literature more specific contributions to the open research problems in the above pipeline's sub tasks, have been made.

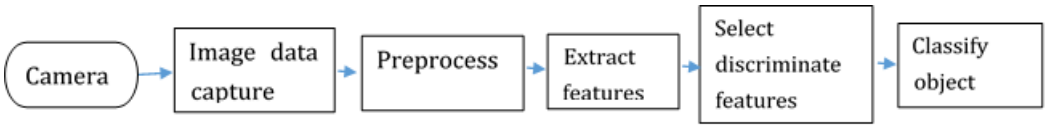


Figure 1.1: Object recognition paradigm

In an attempt to solve the human detection and recognition problem, [102] proposed using Contour Cues, Cascade Classifier (C4) and the **CEN**sus **TR**ansformed h**IST**ogram (CENTRIST) descriptor for human object detection. In [68], to recognise people and estimate their pose, four different feature based techniques were proposed. In [58] colour and texture features were used to train an incremental SVM classifier, for on-line human recognition. Similarly, to recognise vehicle types, [45] investigated the use of a Hybrid Dynamic Bayesian Network (HDBN) classifier. In this approach different features extracted from the tail light and vehicle dimensions with respect to the license plate, were used for vehicle classification. In analysing thermal images, Riaz et.al. in [77], detected pedestrians by using CENTRIST features. Beyond object detection and recognition, algorithms have been proposed for object re-identification. Object re-identification leads to the

identification of different instances of the same object that may be visible within multiple camera views. For example, in [39], re-identification was achieved by a combination of hue, saturation histogram and Saliency Maps features extracted from selected body parts of a human.

A detailed literature review carried out as a part of the research presented in this thesis, presented in chapter-2, revealed that there are still research gaps and potential areas for improvement within a selected set of practical application domains of interest, namely; recognition of military personnel, people detection & vehicle type recognition at both day and night times and people re-identification. Identifying the need to close the existing research gaps in the above application domains, a number of novel algorithms and approaches are proposed in this thesis. They are presented in chapters 4-7.

1.1 Research Motivation

With military organisation facing different security threats on a daily basis; such as personnel disguise, unauthorised restricted information access and leakage, identity stealing and/or impersonation to mention a few, the threat to global security is on the increase. While several efforts have been made to resist with these challenges, crime perpetrators continuously seek new avenues to compromise the efforts of the military, in particular through insider attacks, so as to cause additional challenges from within the organisations. Newer and stronger attacks are being launched to render the security of the military and police forces at large, useless. Automated video surveillance systems that are able to identify and recognise military/security personnel, within the perimeters of a secured premises, can largely help in the timely identification on insider attacks and strengths.

A number of other application domains associated with general public spaces provides motivation for further research into the application of computer vision technologies for video surveillance. One such application domain is where access to vehicles needs to be monitored automatically to manage and control their movement to and from secure sites, in motorways and across international borders. Similarly, the increase of road traffic over the years has caused serious concerns about the level of pollution caused by vehicular traffic. In such situations the automated computer aided detection of vehicles, recognising their type and their make and model, provides motivation for further research.

An additional common concern within that above application domain is that techniques that are designed for day video analysis typically fail when applied in their original form on night-time footage. This challenge provide motivation for further research in developing algorithms that are either adoptable to lighting

changes or new algorithms that works efficiently on night vision videos.

A further application that has recently attracted much attention is people re-identification in surveillance videos. The idea is to try and recognise a person in video footage captured at different times, locations and perhaps by different cameras. In attempts to apply computer vision and pattern recognition algorithms to resolve this practical problem, challenges in illumination changes, view dependency, object scaling, different camera characteristic, noise levels etc have to be met. Algorithms and systems that meet some or all of the above challenges can contribute effectively towards an ultimate solution.

Traditionally in many of the above application domain direct human observation or CCTV operator based surveillance have been used to monitor and analyse scenes/content observed. Advancements in camera technology supported by the significant improvements in computer vision, pattern recognition and machine learning algorithms have enabled some or most of the above manual operations to be automated exceeding accuracy levels obtainable by human observers. Further the recent advancement in computing technologies such as multi-core technologies, distributed computing clusters and cloud technology have significantly increased the available computing resources, allowing the scaling of such applications to levels that it could even replace teams of human observers working real-time over months.

All of the above factors have provided the motivation behind the research conducted and presented in this thesis.

1.2 Research Aim and Objectives

It is important for the military organisations to be proactive by setting up effective military surveillance systems that will help improve the security within their premises/environments. Such systems for example should monitor people and vehicle movement in and out of military camps so as to keep a count of the number of people and vehicles within an environment. Consequently, when less or more than expected number of people and vehicles are present, relevant military authorities can be alerted of unusual activities, so that a check within the environment can be done for malicious and suspicious elements, which in turn will lead to an early warning signal. Further the detailed analysis of the appearance and movement patterns of people and vehicles will provide useful information to further strengthen security levels. The ability to do the analysis above regardless of whether its day-time or night-time will further improve the level of security provided by such systems.

Given the above observations the aim of the research presented in this thesis is

to develop robust and efficient techniques that can detect people in both day-time and night-time, recognise and re-identify people in day-time, detect and recognise vehicle types in both day and night-times.

The above aim will be met by following the objectives listed below:

- Carry out a detailed literature review in the areas of people/vehicle detection and recognition, military personnel recognition and people re-identification including analysis of night-time CCTV footage.
- Design, and implement a robust and efficient algorithm for military personnel recognition;
- Design, and implement a scalable and robust algorithm for recognition of vehicles in day-time and night-time videos;
- Design, and implement an efficient algorithm for pedestrian detection at night-time;
- Design, and implement an efficient algorithm for people re-identification.
- Evaluate the performance of all algorithms implemented using standard and specially captured databases and suggest possible future enhancements and improvements.

1.3 Scholarly Contributions

The research conducted within the context of this thesis has led to the following original contributions:

1. A novel technique to classify the arm of service of a camouflaged military personnel using a combination of Gray Level Co-Occurrence Matrix (GLCM) texture and Hue colour histogram bin features. This research demonstrated the fact that texture alone cannot discriminate between the various camouflage classes, without the integration of colour.
2. A novel technique to classify the arm of service of a uniformed military personnel, using the classification of their cap type, based on feature recognition of badges, using Speed Up Robust Features (SURF) feature matching.
3. A robust and efficient technique to recognise vehicle type (e.g. car, van, bus etc) irrespective of their angle of view. The use of a combination of Region, Histogram Oriented Gradient (HOG) and Local Binary Pattern (LBP) histogram features that are scale and rotation invariant, was proposed.

4. A novel algorithm that detects pedestrians and recognises vehicle types at night-time using CENTROG features, was proposed.

5. An efficient and simple algorithm to re-identify people in non-overlapping cameras, even in the presence of occlusion was proposed based on a normalised 3D 8 bin colour histogram and seven described attributes of a person.

Apart from the above novel algorithms and approaches, the research conducted within the context of this thesis also contributed to the general subject area of object detection, analysis and recognition by developing transferable and re-usable algorithms in other application domains, outside those considered for the purpose of the research presented.

The above original contributions have resulted in the following conference and journal paper contributions.

Refereed Journal Publication

1. Martins E. Irhebhude and E.A. Edirisinghe, *"Personnel recognition in the military using multiple features"*, in: International Journal of Computer Vision and Signal Processing (IJCVSP), September 2015 5(1), pp. 23-30.

Refereed Conference Proceeding

2. Martins E., Irhebhude, and Eran A. Edirisinghe. "Military personnel recognition system using texture, colour, and SURF features." In SPIE Defense+ Security, pp. 90900Q-90900Q. International Society for Optics and Photonics, 2014.

3. Martins E., Irhebhude, Mohammad Athar Ali, and Eran A. Edirisinghe. "Pedestrian detection and vehicle type recognition using CENTROG features for nighttime thermal images." In Intelligent Computer Communication and Processing (ICCP), 2015 IEEE International Conference on, pp. 407-412. IEEE, 2015.

4. Quang A. Nguyen, Martins E., Irhebhude, Mohammad Athar Ali and E.A. Edirisinghe, *"Vehicle Type Recognition in Video using Multiple-Feature Combinations"*, accepted for presentation at the IS & T International Symposium on Electronic Imaging 2016 in Video Surveillance and Transportation Imaging Applications Conference, February 14-18 2016.

1.4 Thesis Layout

For clarity of presentation, the thesis is structured as follows: chapter 2 focuses on the study of literature in people detection, recognition & re-identification, vehicle detection & type recognition, pedestrian detection & vehicle type recognition at night-time; chapter 3 provides the theoretical and conceptual background on colour, texture and shape models and the popular Support Vector Machine (SVM) classifier. The chapter also present feature selection techniques and experimental performance models; chapter 4 presents a novel algorithm for the recognition of arm of service of a military personnel; chapter 5 presents a novel algorithm for vehicle type recognition; chapter 6 presents a novel approach to pedestrian detection and vehicle type recognition, in night-time thermal images; chapter 7 presents a novel approach to people re-identification; and finally, chapter 8 concludes with a view to further improvements of the proposed algorithms and suggestions for future research.

Chapter 2

Literature Review

2.1 Introduction

This section defines a number of important terminology commonly used in image processing related research which are vital to prevent reader confusion. In particular the terms computer vision, object detection, object recognition and object re-identification are defined.

From the web definition [15]:

> ***Computer vision*** *is a field that includes methods for acquiring, processing, analysing, and understanding images and, in general, high-dimensional data from the real world in order to produce numerical or symbolic information, e.g., in the forms of decisions.*

In computer vision terminology, the term ***object detection*** generally refers to the classification of a perceived object into any human identifiable type, for e.g., a car, human, animal, house, tree etc.

In contrast, in computer vision terminology, ***object recognition*** refers to the classification of a perceived object into a particular group of the human identifiable types, for e.g., a car, human, animal, house, tree etc. Object recognition requires a more detailed analysis of features for the purpose of classification into one of the sub groups of identifiable objects. Importantly, object recognition algorithms needs to be robust to changing situations like different camera viewpoints and orientations, varying light conditions, pose variability and clothing appearance [15], which means that the computer should recognise or identify objects irrespective of illumination, background, and pose of the object relative to the camera.

Following the above definitions of object detection and object recognition the following statements can be made. Object detection refers to identifying that a perceived object belongs to either of a number of human recognisable objects.

In contrast object recognition refers to the identification of the perceived object as belonging to a particular group of human recognisable objects, for e.g. a car. Following the same terminology detection of a human refers to detecting that it's any human. Recognising a human means that the human is a particular known individual. In the case of a vehicle one could detect a vehicle (i.e. it belongs to any type of vehicle, car, bus, van etc.), recognise that it's a bus (i.e. a particular type only), or recognise that its BMW, 3 series manufactured between 2000-2002 (i.e. a particular model of a particular type, i.e. a car).

Extending object detection/recognition towards more specific requirements, ***object re-identification*** aims to correctly identify all instances belonging to the same visual object at any time or location [18]; which means choosing the most probable object among sets of possible matches of consecutive observations of the same target at different camera views [5].

To detect, recognise and re-identify an object, the appearance of such object needs to be studied as this is vital for the classification of the object into any of the identifiable types, in line with the above definitions. Appearance-based techniques rely on visual information [58] (i.e. visible parts) for object classification. Similarly, appearance-based methods rely on visual or perceptual principles to extract features for object classification [23]. Colour, shape and texture are common ways in which an appearance can be studied. Texture contains structural arrangement of a surface and its relation with the environments' information [38]. Colour on the other hand, is the perceptual property of red, green, blue etc perceived by humans/machines. Similarly, shape means the form an object assume. Hence, tools that help capture colour, shape and texture information will be the background techniques that will be exploited in this research for the purpose of object detection, recognition and re-identification.

In the following sections a comprehensive review of reported algorithms in literature for object detection, object recognition, vehicle type recognition, and people re-identification for images captured both during day and night times are reviewed.

2.2 Object Detection and People Identification

Many algorithms have been proposed to detect objects and identify the detected objects as humans using attributes of a human appearance. The appearance of a person is the visible foreground image after background subtraction [58]. Appearance-based methods rely on clothes, visual parts or perceptual principles (colour, texture and shapes) to extract features for object recognition [58, 23]. Therefore, colour, texture and shapes information can be considered as features

for object classification. The so-called Gray Level Co-occurrence Matrix (GLCM), which is described as a descriptive texture feature can be used to provide a feature descriptor for classification and is popularly used in literature due to its popularity and simplicity [38, 85, 86, 14, 90]. In general colour, texture and shape features can be used as descriptors for appearance based object recognition [1, 84, 59]. In [38], a novel procedure for extracting textural features from image blocks was proposed for image classification. Gray-tone spatial-dependence probability distribution matrices or GLCM was computed on a given image to form a matrix from which statistical features were extracted and used for classification. Four directions were exploited: $0^0, 45^0, 90^0, 135^0$ using linear discriminant function, Min-Max decision rule and piecewise linear discriminant function classifiers for the respective datasets. Similarly, the effectiveness of using GLCM features was studied in detail in [90]. In this work the use of quantization, displacement and orientation parameter values in the discrimination of sea ice synthetic aperture radar (SAR) imagery datasets was investigated. In the evaluation conducted using a Bayesian classifier, experiments with three different feature sets were conducted; 10 textural features extracted using the following matrices, Mean Displacement & Mean Orientation (MDMO) matrix, χ^2-optimal Displacement & Mean Orientation (ODMO) matrix and χ^2-optimal Displacement & χ^2-optimal Orientation (ODOO) matrix. In [14] the effect of quantization levels and classification accuracy was studied. Five quantization levels of 8, 16, 32, 64, and 256 at 0, 45, 90, 135 orientations within a distance of 1 was studied to extract 8 shift invariant texture features. Correlation analysis was used as a feature selector to improve classification using a Fisher Linear Discriminant (FLD) classifier.

In military applications camouflages are generally useful to confuse an enemy's surveillance system. In [85] using a dendrogram as a classifier and mean of the GLCMs the feature, camouflaged objects were recognised in a defence environment. The use of quantised colour histogram and fuzzy c-means with morphological operations was proposed in [8] for camouflage pattern recognition. The use of spectral texture was proposed in [91] as a discriminating feature for differentiating between green colour camouflage and background that consists of green vegetation. Further the work presented in [54] proposed the use of weighted structural similarity (WSSIM) and nature image features for the evaluation of camouflage texture designs, with evaluation done by the perceived differences between camouflage textures and background image features. Four GLCM texture features fused with a non-singleton dimension was used to recognised a camouflaged objects in [86]. In [59] edge features and differential image detection techniques was used to recognise targets.

Partial Least Squares was used to help reduce dimensions and improve recog-

nition that was based on colour, texture and edge information [84] for human recognition. In [66], the use of semantic and fourier Local Binary Pattern (LBP) features was proposed for human object detection. Experiments were performed to compare LBPs performance with Histogram Oriented Gradient (HOG) and covariance tensor feature (COV) descriptors; results show that LBP outperforms both of the alternative feature techniques. In [82], use of two different sets of edge-texture features, i.e. Discriminative Robust LBP (DRLBP) and Discriminative Robust Ternary Patterns (DRLTP), was proposed for object recognition. Investigations show the limitations of LBP and its variants; hence, the relative advantage of using the new feature sets of DRLBP and DRLTP. In solving the partial occlusion problem, [100] proposed the combination of HOG and LBP feature sets for human object detection. In [102], contour cues, cascade classifier (C4) was proposed for human object detection using **CEN**sus **TR**ansformed h**IST**ogram (CENTRIST) descriptor. Authors claimed that C4 is extremely fast for human detection compared to HOG and LBP. In order to eliminate the false alarm associated with human recognition [60] proposed a background modeling algorithm using fussy logic for accurate foreground segmentation. In [79] the role of face familiarity and motion was examined. It was found that both roles promote recognition in difficult situations. In the paper [68], 4 different feature based techniques were used to recognise and estimate the pose of full body of a person. Similarly, [58] used the incremental SVM as a classifier on colour and thirteen Haralick texture features from RGB image of segmented body parts (head, top, bottom) of foreground image for an on-line human recognition system.

The use of Gray-Level Dependency Matrix (GLDM) within a texture-based object detection technique was proposed in [29] to detect and localise a crowd. The crowd was categorised as follows: 0-1 person (no crowd), 2-4 persons (low), 5-9 persons (med) and 10 and more people (high). In [108], a novel technique based on a combination of High Frequency Optical Flow (HFOF), Multi-scale Histogram Optical Flow (MHOF) and Dynamic Textures (DT) was proposed to detect anomalies in a crowded environment. HFOF captures dynamics of motion behaviour, MHOF captures motion direction and energy information and DT captures dynamic appearance properties. These features help to effectively classify behavioural activities in a crowd image using Multiple Kernel Learning (MKL) as a classification tool. The use of Extended Local Ternary Patterns (ELTP) and Uniform ELTP (UELTP) was proposed by [53] for noise resistivity and classification respectively. The use of a spectral clustering technique [53] and the observation of patterns, helped reduce dimensions in the feature. Experiments conducted in [53] using the Support Vector Machine (SVM) as the classifier include features; LPB, Uniform LBP (ULBP), ELTP with 64 dimension, ELTP with 128 dimension,

UELTP with 58 dimension, UELTP with 128 dimension. In [21], c-means clustering on a contourlet sub-bands was used to extract cluster features which were combined with variance and norm-2 energy features and used for supervised classification. Relative-L1 distance metric is computed on each sub-band to eliminate discrepancy in feature vectors with k-nearest neighbour classifier. The standard feature vector was rearranged using random assignment to form a matrix; recomputed using LBP, local ternary pattern (LTP) and wavelet coiflets to form a descriptor used for classification [69]. Using SVM classifier on 50 different random assignments, results show that the proposed descriptors outperformed the regular vector descriptor and that LTP descriptor can be combined with the regular descriptor to improve overall performance of the classifier. In [44] a technique that allows to search volumes of video data to find candidate person using attributes information was proposed.

In [52] kernel-Principal Component Analysis (kPCA) was used as a dimensionality and noise reduction tool on sets of colour and texture features. The features were extracted using a colour histogram and ULBP for colour and texture, respectively. In separate experiments carried out, results show that the fused and kPCA techniques performed better with the later recording less computation time. However, in [4] PCA helped reduce dimensionality and improve recognition in all cases of experiment. In [1], colour and texture features were captured in the wavelet domain on the YCbCr colour space and using histogram of both features to capture the signature; a k-means algorithm was used to improve the recognition and reduce dimensionality. In [84] colour, texture and edge information were used to capture the signature of the appearance for object recognition. Recognising the enormity of the dimension, the author chose Partial Least Square (PLS) technique as against the popular PCA technique as classification results from both techniques show better performance when using PLS. Further PLS is a class aware dimensionality reduction tool i.e. it provides information regarding the importance of features as a function of location. In [36] two feature selection techniques were compared to determine which one improves machine learning and computation time. From the two approaches considered, namely the wrapper approach and filter approaches, the author recommended a filter based approach, i.e., Correlation based Feature Selection (CFS); results shows that CFS performed better in comparison with the wrapper method.

The literature reviewed above show that significant amount of work has been conducted in object detection and human object identification. Our detailed review of literature showed that no work exist on detection and recognition of military personnel which is a key application focus of the research presented within this thesis. We will use the appearance of a camouflage military person to recog-

nise the arm of service of the personnel. Apart from recognising a service of a military personnel using appearance based features, we will also exploit the use of badge of the military cap to recognise the service of a personnel. This means we will develop a system which can be integrated into a face recognition system that can be used to recognise a particular personnel and determine the arm of service to which the military personnel belongs. This system can do military personnel arm of service persons count, which will help check if a particular personnel is present or absent within a particular arm; hence, check if more service personnel are present at a time in an environment or see if a service personnel is on AWOL.

2.3 Vehicle Detection and Recognition

Existing literature in vehicle detection, counting and type recognition proposes a number of different approaches. The authors of [51] showed that even in a congested road traffic condition an AND-OR graph (AOG) using bottom-up inference can be used to represent and detect vehicle objects based on both front and rear views. In a similar environment, [65] proposed the use of strong shadows as a feature to detect the presence of vehicles in a congested environment. In [104], vehicles were partitioned into three parts; road, head and body, using a tripwire technique. Subsequently Haar wavelet features extracted from each part with PCA performed on features calculated to form 3 category PCA-subspaces. Further, Multiple Discriminate Analysis (MDA) is performed on each PCA-subspace to extract features, which are subsequently trained to identify vehicles using the Hidden Markov Model-Expectation Maximisation (HMMEM) algorithm. In another experiment, a camera calibration tool was used on detected and track vehicle objects so as to extract object parameters, which were then used for the classification of the vehicle into classes of cars and non-cars [34]. In [16] vehicle objects were detected and counted using a frame differencing technique with morphological operators: dilation and erosion. In [98], using maximum likelihood Bayes decision rule classifier on normalised local features (roof, two tail-lights and headlights of rear and front view) vehicle or non-vehicle objects were detected. Further to handle the unevenness in the road surfaces, the author added simulated images and applied PCA on each sub-region to reduce feature sets, computation time and hence speed-up the processing cycle. In another classification task in [75], segmentation through image differencing was used to obtain foreground object, thereafter, sobel edge were computed on each foreground image. Furthermore, the foreground image size feature was extracted with two levels dilation and fill morphological operations; and classified into small, medium and large categories. In [70], an alternative to expensive Electronic Toll Collection (ETC) full-scale

multi-lane free flow traffic system was proposed; the technique used Scale-Invariant Feature Transform (SIFT), the Canny edge detector, k-means clustering with Euclidean matching distance metric for inter and intra class vehicle classification. In [72], a technique for traffic estimation and vehicle classification using region features with a neural network (NN) classifier was proposed. A technique for rear view based vehicle classification was proposed in [45] with investigation of Hybrid Dynamic Bayesian Network (HDBN) in vehicle classification. Tail light and vehicle dimensions with respect to the dimensions of the license plate were the feature sets used for classification. The width, distance from license plate and the angle between the tail light and the license plates formed the eleven features used for classification. The experiment was performed in two phases; known vs unknown classes and four known classes using HDBN. HDBN was compared with three other classifiers. The performance evaluation result using a ROC curve shows that HDBN is the best classifier for rear view vehicle classification.

In observing the vehicle detection and recognition techniques proposed in summarised paragraph above, it can be concluded that vehicles are recognised and classified at different angles under different conditions using different feature sets, classification techniques and hence algorithms. In other words a change of camera angle may require a change of features that need to be extracted for classification. The classification technique that performs best will also change. Further, most techniques have been tested either on rear or front views only. In practice once a camera is installed in an outdoor environment with the hope of capturing video footage for vehicle type recognition, it is likely that due to wind or neglect in installation, the camera could turn in due course. If the vehicle type recognition system was dependent significantly on the angle of view, the system would thus fail to operate accurately. Further at the point of installation practical problems may be such that the camera position and orientation will have to be changed as compared to the fixed angular view that it has originally being designed for. This will either require the system to be re-redesigned using different feature sets, classifiers and algorithms or the system having to go through a camera calibration processes, which is typically non-trivial and time consuming. It would be ideal if at the new orientation the captured content could still be used for classification.

Given the above observations we propose a novel algorithm for vehicle type recognition and subsequent counting, which is independent of the camera view angle. We adopt a strategy that uses multiple features that are scale and rotation invariant, leading to the accurate classification of vehicles independent of the camera angle.

2.4 Pedestrian Detection and Vehicle Type Recognition in Night-time

This section reviews a number of reported techniques that can detect pedestrians and recognize vehicle types within night-time thermal images. In a situation where colour and texture information are missing in a dataset, shape information can help in recognising objects. Therefore, for night-time datasets, several techniques are reviewed to understand how pedestrians can be detected and vehicles can be recognised at night-time. Benezeth et.al. [7] proposed the use of a Gaussian-based segmentation method with Haar-like features using a cascade of boosted classifiers to detect humans in a room. Two contributions were made by Yun et. al. in [106]; segmentation based on histogram cluster analysis using k-means and a feature extraction technique based on histogram of maximal oriented energy map using log-Gabor wavelets for selecting orientation. An evaluation of the efficiency of a night-time mid-range infrared sensor and its application in human detection and recognition was done by Bourlai et. al. in [10]. The local oriented shape context feature was used by Li et.al. in [50] to detect pedestrians in a night-time scenario by adding orientation information to shape context feature, thereby capturing appearance and shape information. In [57], Liu et. al. proposed a technique based on entropy weighted HOG as a feature detector and SVM as a classifier. The authors sped up the classification phase by reducing the number of support vectors and filtered false alarms by introducing a validation phase that examined the gray-level intensity of pedestrians heads. For thermal images, Chang et. al. [13] used HOG features and Adaboost to detect and classify pedestrians. Their feature extraction method included image segmentation and Region-of-Interest (ROI) generation. In [77], Riaz et.al. detected pedestrians within thermal images by using CENTRIST features and compared the performance of their technique with the popular HOG based techniques. Both above-reported techniques proved that CENTRIST-based approaches exhibit better detection accuracy with lesser computation time when compared to other methods. In [55], a feature combination of HOG and contour was proposed for pedestrian detection. The authors also proposed a foreground segmentation technique for smart region detection. In [99], Wang et.al. proposed a shape context descriptor (SCD) based on the Adaboost cascade classifier framework. The technique was applied to thermal images and the results were compared with the rectangle-based detection feature. The authors claimed that their technique outperformed the rectangle-based features in terms of detection accuracy but suffers in terms of higher computation intensity.

A look at the state-of-the-art reveals a lack of techniques to recognize vehicle types in night-time thermal image sets, though quite a few techniques have been

reported for visible images. For instance in [43], Iwasaki et. al. reported a vehicle detection mechanism within thermal images using the well known Viola Jones detector. The technique involved detecting the thermal energy reflection area of tires as a feature.

2.5 People Re-identification

As mentioned before, the goal of object re-identification is to correctly identify all instances of the same visual object at any time or location [18]; meaning, choosing the most probable object among sets of possible matches of consecutive observations of the same target at different camera views [5]. In [23] three features were accumulated; entire colour content, colour regions, texture characteristics of recurrent region to form Symmetry Driven Accumulation of Local Features (SDALF) and used on three datasets to give a novel state-of-art performance in object recognition and re-identification. In [81], authors combined [23]'s SDALF technique with mid-level semantic feature attribute to identify candidate objects. Further the importance of attributes and how relevant attribute features can be selected for object re-identification task was also demonstrated. Random forest technique was used by [56] to determine the importance of individual feature attributes under different circumstances of various roles for object classification. A framework, Multiple Component Matching (MCM) was proposed in [83] for object re-identification. MCM was explained as an ordered set of sequences containing several components with simulated parts generated to cater for illumination variation. Authors however established that simulated components increased the computation complexity. To correct the computation complexity issue authors vectorised and clustered the MCM to form a prototype. The matching were done in the dissimilarity space with text information used as a query for image retrieval. Mean Riemannian Covariance Grid (MRCG) in [3], modeled clothing information to describe the human object for recognition. Covariance matrix was used to describe images of fixed sizes with equal grid structures and averaged to get the Mean Riemannian Covariance (MRC) that describes the object for re-identification. In [2] HOG features were trained to detect body parts; top, torso, leg, left arm, right arm. Covariance's of colour gradient and orientation was computed on each region including the full body to get discriminative signature used for people re-identification. In [22] the standard LBP was modified by setting dimensionality at 16 to form the Simplified LBP (SLBP) to detect people's head and face. In order to re-identify people; authors used [3]'s MRCG technique to model the detected head and face so as to capture a discriminative signature. An optimised Speed Up Robust Feature (SURF) named Camellia Key

Point was used in [37] to describe grayscale (to eliminate variation in colours) candidate objects and used for re-identification on CAVIAR datasets with the threshold set at 15. In [18] colour samples were modelled using fuzzy K-Nearest Neighbour (KNN) algorithm to segment candidate objects into eleven culture colours. Probability Colour Histogram (PCH) plot were used to identify an object at a set threshold after comparing two targets in intra and inter camera scenarios. People in a crowded environment can be identified by integrating appearance features: selective upper body patch and candidate position and direction of travel using a landmark-based model [78]. Analysis showed that the proposed technique performed better than the full body based integration. In [20], SURF features was proposed for interest point extraction using Sum of Quadratic Difference (SQD) as a point correlation tool for object identification in a distributed camera network. In a similar scenario, [87] proposed an unsupervised iterative brightness transfer function (BTF), a technique to handle the variability that occurs in illumination conditions. BTP helps to map brightness values between intra camera views while cumulative BTF helps to adapt colours in inter camera views for people re-identification. In a low quality camera network; [5] used a Colour Structure Descriptor (CSD) by extracting dominant colours from regions of interest (shirt and pant); derived CSD by evaluating the differences of dominant colours between the two targets and proposed a so-called Target Colour Structure (TCS) for people re-identification. A two feature approach was proposed in [88] for object recognition, i.e., Haar and Dominant Colour Descriptor (DCD) features. Haar features of the foreground mask recognised an object in the first technique while DCD works by partitioning the detected foreground object into two, then using the dominant colours of both regions as descriptors for object recognition. In [9] two techniques were proposed; Red Green Blue (RGB) colours were used alongside the height feature histogram and transformed normalised RGB colours plus the height feature histogram techniques to identify objects using histogram matching. Instead of recognising objects using a distance measure, [12] proposed Ensemble RankSVM for ranking image sets with the correct match having the highest ranking score. A comparison between rank and distance measure techniques for object re-identification was conducted. Ensemble RankSVM was however recommended because of the scalability of the technique. In [46], ULBP and Hue Saturation Value (HSV) histogram were used as features extracted from body segmented into 3 parts of a detected target to capture local texture and colour features. These features alongside direction of view captured different identifiers for 3 views; front, back and side that helped in person re-identification. In [103], a model which is a function of pose was developed to capture human appearance. With the rectified pose prior image specific person's feature of colour and textures were extracted; re-

identification and identification of targets became more robust to viewpoint on the trained dataset. In [61], persons were re-identified by accumulating local weight map histogram features from 3 areas of a segmented human body. The local weighted histograms were trained for optimal weight map. These local weight map histograms were integrated to form a feature vector used for identification. In [49], the use of middle-level clothing attribute information was described to assist in person re-identification. Re-identification performance was improved by treating clothing attributes as real value variables. In their pre-processing steps, a body part-based representation approach was proposed by extracting HSV colour histogram and HOG as features. A further contribution was the generation of a large-scale dataset that contains more samples and camera views than the currently available public datasets. In [39], people were re-identified by a combination of features; hue, saturation histogram and Saliency Maps from selected body parts. In [28], a technique that identifies human action and appearance based on colour and optical flow models was proposed. The mean features from two regions of a detected candidate identified a person's action and appearance. The colour features were extracted from 8 colour spaces; R, G, B, H, S, Y, Cb, Cr channels respectively. In [93], instead of solving people identification problem using ranking and distance measures, Takač et. al. used an appearance based learning algorithm such as SVM and the Naive Bayes classifiers to identify people. Finally, [47] proposed a mid-level identification approach called the Optimised Attribute Re-identification. 21 attributes were proposed and detected.

Chapter 3

Background of Study

This chapter presents the theoretical and conceptual background on which the novel and contributory work of this thesis presented in chapters 4-7 are based. It presents the reader to colour models, Canny edge detection, SURF features, Census transforms, HOG features, GLCM matrix, LBP descriptor, Gaussian Mixture Models (GMM) and Support Vector Machine (SVM) classifiers, Correlation based Feature Selection (CFS) and performance analysis based on Receiver Operating Characteristics (ROC) curves.

3.1 Red Green Blue (RGB) colour model

The Red, Green and Blue (RGB) colour model represent all visible colours based on a combination of three primary colours red, green and blue [64, 63]. When red, green and blue colour components are combined in equal proportions of 1, white colour is produced. Other weightings produced all other colours with equal proportions of 0, creating the colour black (see table 3.1). An RGB colour image of size $M \times N$ is represented as a $M \times N \times 3$ array of colour pixels, with each pixel triplet corresponding to RGB components at specific pixel location [32, 63]. An RGB colour image can also be viewed as a stack of three gray-scale images, when fed into a colour monitor, produces a colour image on the screen (see figure 3.1).

Table 3.1: RGB values of some basic colours in [32]

Name	RGB Values
Black	[0 0 0]
Blue	[0 0 1]
Green	[0 1 0]
Cyan	[0 1 1]
Red	[1 0 0]
Magenta	[1 0 1]
Yellow	[1 1 0]
White	[1 1 1]

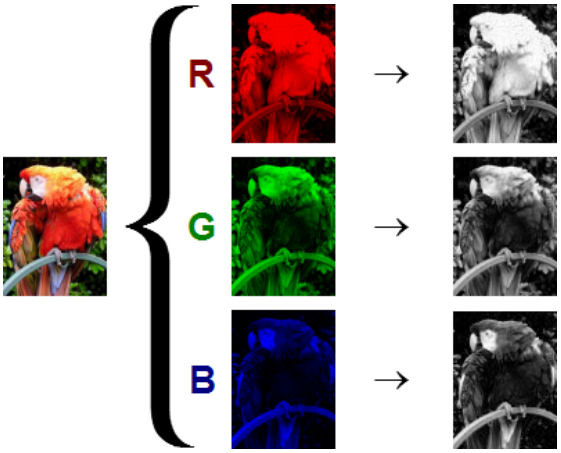


Figure 3.1: RGB Channels Separation: © Nevit Dilmen at Wikimedia (https://commons.wikimedia.org/wiki/File:RGB_channels_separation.png)

3.1.1 Selected feature descriptors from RGB colour

We will briefly describe two selected features which can be extracted from RGB colour channels. Normalised 2D histogram and local shape features.

3.1.1.1 Normalised 2D Histogram

According to [68], two dimensional (2D) normalised color histograms can be calculated as:

$$\begin{aligned} r &= \frac{R}{(R+G+B)}, \\ g &= \frac{G}{(R+G+B)} \end{aligned} \tag{3.1}$$

Normalised 2D Histogram is therefore the histogram of the r and g channels respectively.

3.1.1.2 Local shape features

According to [68], the local features of an image are obtained by convolving the local shape patterns shown in figure 3.2. These patterns were used to extract features for position invariant person detection.

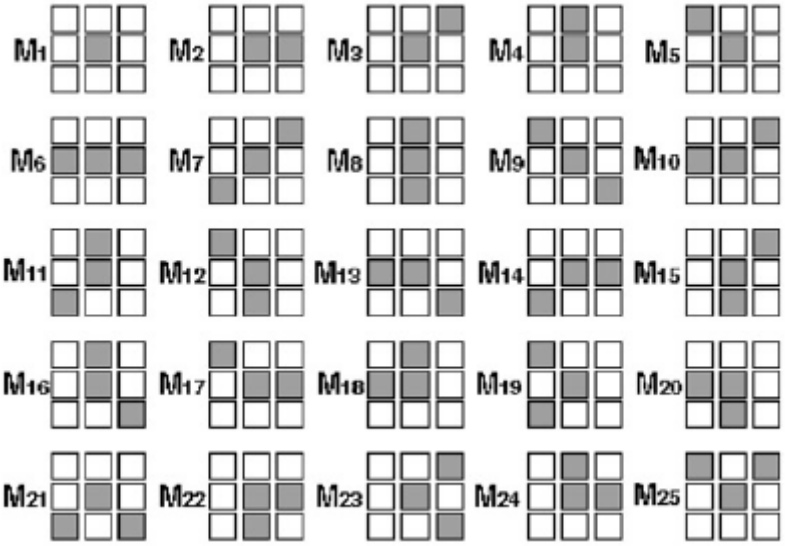


Figure 3.2: Shape patterns in [68]

Convolution operations considered are linear and non-linear operations:
The linear operation is given by;

$$\sum_k V_k.M^i$$

where $M^i, i = 1, 2, \cdots, 25$, are the patterns in figure 3.2 and V_k is a 3×3 patch at pixel k in an image and the sum is on the image pixels.

Non-linear operation is given by;

$$F_i = \sum_k C_{(k,i)}$$

where

$$C_{(k,i)} = \begin{cases} V_k.M^i \ if \ V_k.M^i = max_j(V_k.M^j) \\ 0 \ otherwise \end{cases}$$

The local shape feature implementation uses the simple linear convolution from the patterns 1 to 5 and the non-linear convolution from the patterns 6 to 25.

3.1.2 Converting from RGB to other colour models

A number of other colour representation schemes exist [32], of which the HSI/HSV (Hue, Saturation and Intensity/Value) colour scheme being the representation scheme most widely used in image processing and manipulation.

3.1.2.1 The Hue Saturation Intensity (HSI) colour model

In [32, 94] Hue Saturation Intensity (HSI) is described as a model that helps computers see colours in a way similar to human understanding. Hue is described as pure colour or pigment, saturation measures rate of white light dilution on the colour while intensity describes gray levels or brightness rates of the perceived object. HSI is very useful for comparing two colours and changing from one colour say cyan to another, yellow. It is also important to note that HSI is very useful for measuring colour characteristics in objects [62].

The HSI model is specified using a three-dimensional color tree [48] (see figure 3.3). Saturation of a colour increase as a function of distance from intensity axis [32].

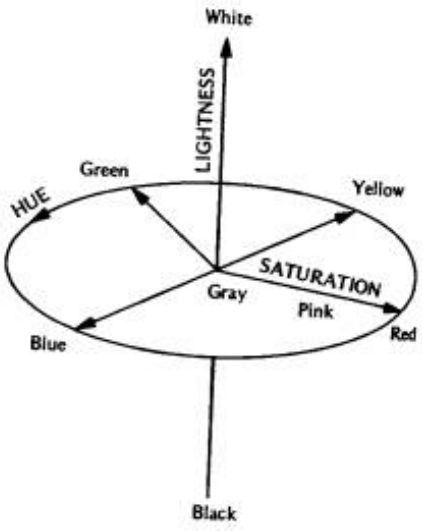


Figure 3.3: The Hue Saturation Intensity (HSI) model

Given and RGB image, HSI components can be obtained using equation 3.2.

$$\begin{cases} H = \begin{cases} \theta, & if\ \mathrm{B} \leq \mathrm{G} \\ 360 - \theta, & if\ \mathrm{B} > \mathrm{G} \end{cases} \\ with\ \theta = \cos^{-1}\left\{ \frac{0.5[(R-G)+(R-B)]}{[(R-G)^2+(R-B)(G-B)]^{\frac{1}{2}}} \right\} \\ S = 1 - \frac{3}{R+G+B}[min(R,G,B)] \\ I = \frac{1}{3}(R+G+B) \end{cases} \tag{3.2}$$

where I and S are in the range of $[0, 1]$ and H $[0, 360]$. In conversion from HSI to RGB there are areas of interest depending on the values of H corresponding to the 120^0 intervals between the primaries. When H is between $0^0 and 120^0$ conversion procedure is as shown in equation 3.3.

$$\begin{aligned} R &= I\left[1 + \frac{ScosH}{cos(60^0 - H}\right], \\ G &= 3I - (R+B), \\ B &= I(1-S) \end{aligned} \tag{3.3}$$

When H is between $120^0 and 240^0$ conversion procedure is as shown in equation 3.4.

$$\begin{aligned} H &= H - 120^0 \\ R &= I(1-S) \\ G &= I\left[1 + \frac{S\cos H}{\cos(60^0 - H)}\right] \\ B &= 3I - (R+G) \end{aligned} \tag{3.4}$$

When H is between 240^0 *and* 360^0 conversion procedure is as shown in equation 3.5.

$$\begin{aligned} H &= H - 240^0 \\ R &= 3I - (G+B) \\ G &= I(1-S) \\ B &= I\left[1 + \frac{S\cos H}{\cos(60^0 - H)}\right] \end{aligned} \tag{3.5}$$

Advantages of HSI colour model

The advantages of the HSI colour model over other colour models are listed below:

- HSI components correlates better with human perception of color.
- Perfect for image processing applications.

- The hue component can be used for segmentation process rather than the three original components.
- Separates the saturation and intensity values from colour.

3.2 Canny Edge Detection

According to [73] an edge detector is an operator that is sensitive to grey level change in an image. Detecting these changes in intensity can be accomplished using first or second-order derivatives [31]. Finding edge strength and direction at location (x, y) of an image, I, is accomplished using the gradient, denoted by ∇I; defined by the vector [31]:

$$\nabla I \equiv grad(I) \equiv \begin{bmatrix} g_x \\ g_y \end{bmatrix} = \begin{bmatrix} \frac{\partial I}{\partial x}, \frac{\partial I}{\partial y} \end{bmatrix} \tag{3.6}$$

Equation 3.6 has an important geometrical property that it points in the direction of the greatest rate of change of I at location (x, y).

The direction measured with respect to the $x - axis$ and the value of the rate of change in the direction of the gradient vector is denoted as [31]:

$$M(x, y) = mag(\nabla I) = \sqrt{g_x^2 + g_y^2} \tag{3.7}$$

and

$$\alpha(x, y) = tan^{-1} \left[\frac{g_y}{g_x} \right] \tag{3.8}$$

Canny defined a set of goals for an edge detector and described an optimal method for achieving them [73]. The goals are,

- Error rate: a detector should respond to edges only and not miss any.
- Localisation: the distance between pixels found and the actual edge should be as small as possible.
- Response: should not identify multiple edges where only one edge is present.

Canny assumed a step edge subject to white Gaussian noise [73]. The edge detector was assumed to be a convolution filter f, which would smooth the noise and locate the edge.

The steps to implementing a canny edge detector are [33]:

- first smooth the image to eliminate the noise;

- find the image gradient to highlight regions with high spatial derivatives;
- algorithm track along these regions and suppress any pixel that is not at the maximum;
- the gradient array is now further reduced by hysteresis;
- hysteresis is used to track along the remaining pixels that have not been suppressed;
- hysteresis uses two thresholds and if the magnitude is below the first threshold, it is set to zero (made a non-edge);
- if the magnitude is above the high threshold, it is made an edge;
- if the magnitude is between the two thresholds, it is set to zero unless there is a path from this pixel to a pixel with a gradient above high threshold.

3.3 Gray Level Co-occurrence Matrix (GLCM)

Gray Level Co-occurrence Matrix (GLCM) is a tabulation or matrix that considers the relationship between neighbouring pixels in an image by calculating how often a pixel intensity value occurs to another pixel value [38]. The GLCM is a tabulation of how often different combinations of pixel brightness values (grey levels) occur in an image. The GLCM is a second order function which measures the angular relationship and distance between neighbouring pixels in an image [90]. Table 3.2 specify common angles, given the distance D. GLCM probability measure as defined by [14] is shown in equation 3.9.

$$Pr(x) = \{C_{ij}|(\delta, \theta)\};\ C_{ij} = \frac{P_{ij}}{\sum_{i,j=1}^{G} P_{ij}} \tag{3.9}$$

where P_{ij} is the number of occurrences of gray levels i and j within the given window, given a certain (δ, θ) pair, (inter-pixel distance (δ) and orientation (θ)); while G is the quantised number of gray levels.

Some pixel values and their GLCM representations are illustrated in fig (3.4).

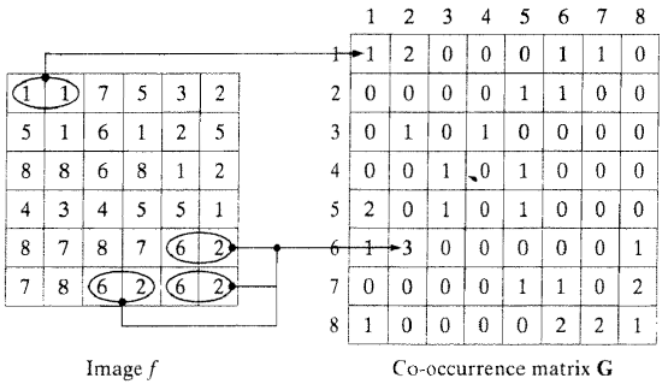


Figure 3.4: Pixel values with the GLCM representation

GLCM works by filling a cell with the number of times the combinations occur, for instance, for top left cell, how many times did 0,0 occur in the image i.e. how many times within the image area a pixel with grey level 0 (neighbour pixel) falls to the right of another pixel with grey level 0 (reference pixel).

Table 3.2: GLCM common angles

Angle	**Offset**
0	[0 D]
45	[-D D]
90	[-D 0]
135	[-D -D]

where offset is the distance between pixel of interest and its neighbour with respect to the direction.

When D is 1, the common angles are as illustrated in fig 3.5.

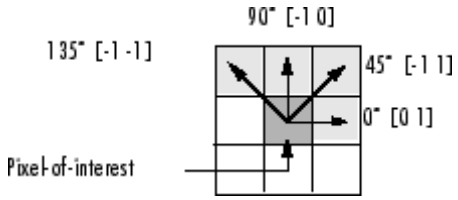


Figure 3.5: Angle and Distance between pixel

3.4 Speed Up Robust Features (SURF)

Inspired by the popular Scale Invariant Feature Transforms (SIFT) [101], SURF is known as a fast, robust local feature detector that is based on the Hessian Matrix. SURF can be used to extract features of an object of interest and thus can be used for object recognition [67]. The SURF descriptor is extracted by positioning

a square region around a point of interest to gather reproducible orientation information [6]. A brief summary of the process of obtaining SURF features is as follows [67]:

3.4.1 Interest point localisation

Given a point $x = [x, y]$ in an image I, the Hessian matrix $H(x, \sigma)$ in x at scale σ is defined as follows:

$$H(x, \sigma) = \begin{bmatrix} L_{xx} & L_{xy} \\ L_{xy} & L_{yy} \end{bmatrix} \tag{3.10}$$

where $L_{xx}(x, \sigma)$ is the convolution of the Gaussian second order derivative $\frac{\partial^2}{\partial_x^2} g(\sigma)$ with the image I in point x, and similarly for $L_{xy}(x, \sigma)$ and $L_{yy}(x, \sigma)$.

Location and scale of interest points are selected by relying on the determinant of the Hessian. Further a non-maximum suppression is applied on a $3 \times 3 \times 3$ neighbourhood of a localised interest point in the image scale and space.

3.4.2 Interest point description

SURF constructs a circular region around the detected interest points in order to assign a unique orientation and gain invariance to image rotation. The orientation is computed using Haar wavelet responses in both x and y directions. The Haar wavelets can be computed using integral images, similar to Gaussian second order approximated box filters (see Fig (3.6)). The dominant orientation is estimated and included in the interest points information.

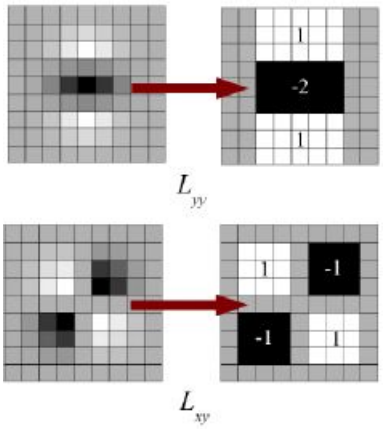


Figure 3.6: SURF box filter ©Copyright 2013, Alexander Mordvintsev Abid K. Last updated on Oct 31, 2014.

SURF descriptors are constructed by extracting square regions around the interest points. The windows are split into sub-regions to retain spatial information.

Haar-wavelets are extracted at regularly spaced sample points. The wavelet responses in horizontal and vertical directions (dx and dy) are summed up over each sub-region (see Fig (3.7)). Absolute values $|d_x|$ and $|d_y|$ are further summed in order to obtain information about the polarity of the image intensity changes as described by:

$V = [\sum d_x, \sum d_y, \sum |d_x|, \sum |d_y|]$, which is the SURF descriptor.

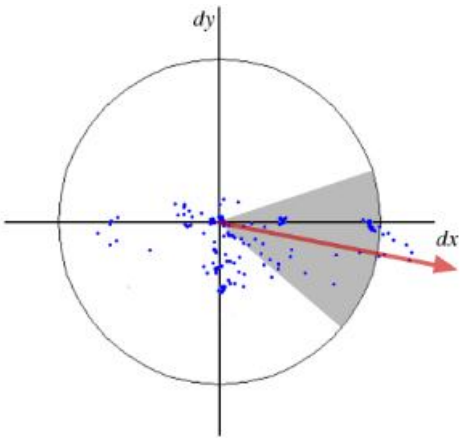


Figure 3.7: SURF orientation graph ©Copyright 2013, Alexander Mordvintsev Abid K. Last updated on Oct 31, 2014.

3.5 Histogram of Oriented Gradients (HOG)

According to [17] HOG is described as a concept that the local appearance and shape of an object can be characterized well by the distribution of local intensity gradients or edge direction, without knowledge of edge positions. It is usually implemented by dividing an image window into small regions named cells and accumulating each local cell's $1-D$ histogram of gradient directions or edge orientations over the pixels. The technique of HOG works by counting the occurrences of gradient orientation in localized portions of an image. The feature HOG captures local object appearance and shape which can often be characterized rather well by the distribution of local intensity gradients or edge directions as reported in [77].

The combined entries form a representation that is contrast normalised to ensure invariance to illumination. This normalisation is extended to all cells in the block to form the HOG descriptor. Dalal and Triggs [11] explored different methods for block normalization. Gradient is computed by applying $[-1, 0, 1]$ and $[-1, 0, 1]^T$ in horizontal and vertical directions of image [77]. Gradient information is collected from local cells into histograms using tri-linear interpolation. On the overlapping blocks composed of neighboring cells, as shown in fig (3.8), normalization is performed.

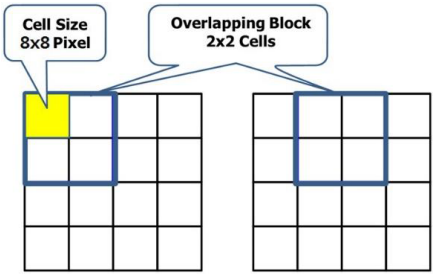


Figure 3.8: Cells and Overlapping Blocks

An example can be illustrated as in fig (3.9) below:

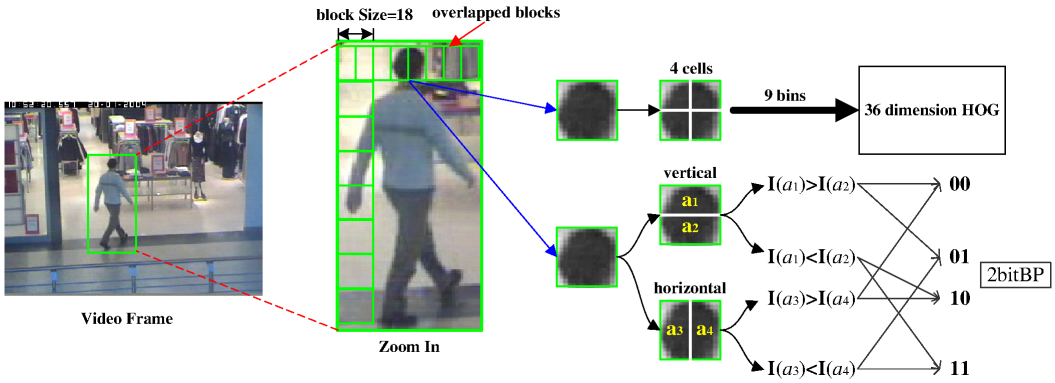


Figure 3.9: Cells and Overlapping Blocks in [27]

Then the normalization factor can be one of the following:

$$L2-norm: f = \frac{v}{\sqrt{\|v\|_2^2 + e^2}} \tag{3.11}$$

$L2-hys: L2-norm$ followed by clipping (limiting the maximum values of v to 0.2) and renormalizing,

$$L1-norm: f = \frac{v}{(\|v\|_1 + e)} \tag{3.12}$$

$$L1-sqrt: f = \sqrt{\frac{v}{(\|v\|_1 + e)}} \tag{3.13}$$

Let v be the non-normalized vector containing all histograms in a given block, $\|v\|_k$ be its $k-norm$ for $k = 1, 2$ and e be a small constant.

In their experiments, Dalal and Triggs found that the $L2-Hys$, $L2-norm$, and $L1-sqrt$ schemes provide similar performance, while the $L1-norm$ provides a slightly less reliable performance; however, all four methods showed significant improvement over the non-normalized data.

Extracted HOG features are robust to changes in lighting conditions and small variations in pose.

HOG feature length is computed as:

$$\begin{aligned} \vec{x} &= (\vec{I}./\vec{z} - \vec{y})./(\vec{y} - \vec{ol}) + 1 \\ length &= \prod[\vec{x}, \vec{y}, k] \end{aligned} \tag{3.14}$$

where $\vec{y}$ is a two element vector block size, k is constant 9 (number of bins), $\vec{z}$ is a two element vector cell size, $\vec{I}$ is the image size, and $\vec{ol}$ is block overlap.

3.6 Local Binary Patterns

The Local Binary Pattern (LBP) operator labels the pixels of an image with decimal numbers that encode the local structure around each pixel of an image [40]. Each pixel (i.e. $g^1, g^2, \cdots, g^8$) is compared with its eight neighbours (see equation 3.15) by subtracting the center pixel value; the results; if negative, are encoded as 0, and the otherwise 1 (see equation 3.16). For each given pixel, a binary number is obtained by concatenating all these binary values (referred to as LBPs, see equation 3.17) in a clockwise direction, which starts from the one of its top-left neighbour. The corresponding decimal value of the generated binary number is then used for labeling the given pixel.

LBP can be described as follows:

Pixel neighbourhood:

$$\begin{pmatrix} g_8 & g_1 & g_2 \\ g_7 & g_c & g_3 \\ g_6 & g_5 & g_4 \end{pmatrix} \tag{3.15}$$

thresholding:

$$\begin{pmatrix} s(g_8 - g_c) & s(g_1 - g_c) & s(g_2 - g_c) \\ s(g_7 - g_c) & & s(g_3 - g_c) \\ s(g_6 - g_c) & s(g_5 - g_c) & s(g_4 - g_c) \end{pmatrix} s(x) = \begin{cases} 1, x \geq 0 \\ 0, x < 0 \end{cases} \tag{3.16}$$

LBP for pixel:

$$LBP = \sum_{p=0}^{P-1} s(g_p - g_c) 2^P$$

Example

$$\begin{pmatrix} 56 & 58 & 95 \\ 20 & 80 & 98 \\ 22 & 79 & 80 \end{pmatrix}$$

$$\begin{pmatrix} s(56-80) & s(58-80) & s(95-80) \\ s(20-80) & & s(98-80) \\ s(22-80) & s(79-80) & s(80-80) \end{pmatrix} \Rightarrow \begin{pmatrix} 0 & 0 & 1 \\ 0 & & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\begin{pmatrix} 0 \times 2^7 & 0 \times 2^0 & 1 \times 2^1 \\ 0 \times 2^6 & & 1 \times 2^2 \\ 0 \times 2^5 & 0 \times 2^4 & 1 \times 2^3 \end{pmatrix} \Rightarrow 00001110_2 \tag{3.17}$$

$$00001110_2$$
$$\Rightarrow 14$$

Other variants are; circular LBP, rotation invariant LBP, uniform LBP, multi-scale LBP, and multi-dimensional LBP.

3.7 Gaussian Mixture Model (GMM)

According to [76, 109] a GMM is a parametric probability density function that is represented as a weighted sum of Gaussian distributions. The GMM technique uses a method to model each background pixel by a mixture of k Gaussian distributions [71]. The weight of the mixture represents the time proportion for which the pixel values stay unchanged in a scene. Probable background colours stay longer and are more static than the foreground colours.

In [92], the recent history of each pixel, $X_1, ..., X_t$, is modelled by a mixture of K Gaussian distributions. The probability of observing the current pixel value is defined as:

$$P(X_t) = \sum_{i=1}^{K} \omega_{i,t} * \eta(X_t, \mu_{i,t}, \sum_{i,t}) \tag{3.18}$$

where K is the number of distributions, $\omega_{i,t}$ is an estimate of the weight (what portion of the data is accounted for by this Gaussian) of the i^{th} Gaussian in the mixture at time t, $\mu_{i,t}$ is the mean value of the i^{th} Gaussian in the mixture at time t, $\sum_{i,t}$ is the covariance matrix of the i^{th} Gaussian in the mixture at time t, and η is a Gaussian probability density function of the form:

$$\eta(X_t, \mu, \sum) = \frac{1}{(2\pi)^{\frac{n}{2}} |\sum|^{\frac{1}{2}}} e^{-\frac{1}{2}(X_t - \mu_t)^T \sum^{-1} (X_t - \mu_t)} \tag{3.19}$$

The covariance matrix is of the form:

$$\sum_{k,t} = \sigma_k^2 I \tag{3.20}$$

3.8 Census Transform

According to [107], census transform (CT) is a non-parametric local transforms which can be described as follows:

Let P be a pixel, $I(P)$ its intensity (usually an 8-bit integer), and $N(P)$ the set of pixels in some square neighborhood of diameter d surrounding P. All non-parametric transforms depend upon the comparative intensities of P versus the pixels in the neighborhood $N(P)$.

Define $\xi(P, P')$ to be 1 if $I(P') < I(P)$ and 0 otherwise. The non-parametric local transforms depend solely on the set of pixel comparisons, which is the set of ordered pairs

$$\Xi(P) = \bigcup_{P' \in N(P)} (P', \xi(P, P')) \tag{3.21}$$

The census transform $R\tau(P)$ maps local neighbourhood surrounding a pixel P to a bit representing the set of neighbouring pixels whose intensity is less than that of P. Therefore, CT compares the intensity value of a pixel with its eight surrounding neighbours; in other words, CT is a summary of local spatial structure given by equation (3.22) [107]:

$$R\tau(P) = \bigotimes_{[i,j] \in D} \xi(P, P + [i, j]) \tag{3.22}$$

where D is a set of displacements, and $\bigotimes$ be concatenation.

To illustrate the manner in which these transforms tolerate factionalism, consider a 3×3 region of an image whose intensities are:

127	127	129
126	**128**	129
127	131	A

for some value $0 \leq A \leq 256$.

All the elements of Ξ except one will remain fixed as A changes. Ξ will be

1	1	0
1		0
1	0	a

where a is 1 if $A < 128$, and otherwise 0. The CT simply results in the bits of Ξ in some canonical ordering such as 1, 1, 0, 1 0, 1, 0, a.

Example:

$$\begin{array}{|c|c|c|} \hline 26 & 75 & 65 \\ \hline 26 & \mathbf{46} & 22 \\ \hline 26 & 40 & 65 \\ \hline \end{array} \Rightarrow \begin{array}{ccc} 1 & 0 & 0 \\ 1 & & 1 \\ 1 & 1 & 0 \end{array} \Rightarrow (10011110)_2 \Rightarrow CT = 158$$

From example above, it can be seen that if the pixel under consideration is larger than (or equal) to one of its eight neighbours, a bit 1 is set in the corresponding location; else a bit 0 is set. The eight bits generated from intensity comparisons can be put together in order and converted to a base-10 value. This is the computed CT value for the pixel under consideration.

3.9 Correlation-based Feature Selection (CFS)

According to [36] as reported by [58], CFS is a filtering algorithm that evaluates subsets of features based on the predicting power of the individual features of a class label. In [58] CFS is defined as:

$$Merit_{S_k} = \frac{k\overline{r_{cf}}}{\sqrt{k + k(k-1)\overline{r_{ff}}}}.$$

Here, S_k is the number of features selected in the current subset,
$\overline{r_{cf}}$ is the average value of all feature-classification correlations, and
$\overline{r_{ff}}$ is the average value of all feature-feature correlations.

It begins with an empty set of features adds one feature at a time that holds best discriminative value.

The CFS criterion is defined as follows:

$$\text{CFS} = \max_{S_k} \left[\frac{r_{cf_1} + r_{cf_2} + \cdots + r_{cf_k}}{\sqrt{k + 2(r_{f_1f_2} + \cdots + r_{f_if_j} + \cdots + r_{f_kf_1})}} \right]. \tag{3.23}$$

The r_{cf_i} and $r_{f_if_j}$ variables are referred to as correlations.

According to [105], CFS involves the following two aspects:

1. how to decide whether a feature is relevant to the class or not; and
2. how to decide whether such a relevant feature is redundant or not when considering it with other relevant features.

A feature is significant if it is the main predicting power in a class, and feature selection for classification is a process that identifies all these principal features to the class concept and removes the rest. If two features are found to be redundant to each other, CFS removes one of them that is less relevant to the class concept; and hence, keeps more information to predict the class.

3.10 Support Vector Machine (SVM)

According to [68] SVM is a technique used to train classifiers, regressors and probability densities that is well-founded in statistical learning theory. SVM can be used for binary and multi-classification tasks.

3.10.1 Binary classification

SVM perform pattern recognition for two-class problems by determining the separating hyperplane with maximum distance to the closest points of the training set. In this approach, optimal classification of a separable two-class problem is achieved by maximising the width of the margin between the two classes [63]. The margin is the distance between the discrimination hyper-surface in n-dimensional feature space and the closest training patterns called support vectors. If the data is not linearly separable in the input space, a non-linear transformation $\Phi(.)$ can be applied, which maps the data points $x \in \mathbb{R}$ into a high dimensional space H, which is called a feature space. The data is then separated as described above. The original support vector machine classifier was designed for linear separation of two classes; however, to solve the problem of separating more than two classes, the multi-class support vector machine was developed.

3.10.2 Multi-class classification

SVM was designed to solve binary classification problems. In real world classification problems however, we can have more than two classes. In the attempt to solve $qclass$ problems with SVMs; training q SVMs was involved, each of which separates a single class from all remaining classes, or training q^2 machines, each of which separates a pair of classes. Multi-class classification allows non-linearly separable classes by combining multiple $2-class$ classifiers. $N-class$ classification is accomplished by combining N, $2-class$ classifiers, each discriminating between a specific class and the rest of the training set [63]. During the classification stage, a pattern is assigned to the class with the largest positive distance between the classified pattern and the individual separating hyperplane for the N binary classifiers. One of the two classes in such multi-class sets of binary classification problems will contain a substantially smaller number of patterns than the other class [63].

SVM classifier was chosen because of its popularity and speed of processing.

3.11 Receiver Operating Characteristics (ROC) curves

The Receiver Operating Characteristic (ROC) curve helps to visualise classification performance in detail. In a ROC curve the True Positive Rate (sensitivity or recall is the fraction of relevant instances that are retrieved) is plotted as a function of the False Positive Rate (false alarm rate refers to the probability of falsely rejecting the null hypothesis for a particular test) for different cut-off points or threshold of a parameter [24]. On the other hand, precision is the fraction of retrieved documents that are relevant. Given by:

True Positive Rate (Recall) = $\frac{tp}{(tp+fn)}$;
False Positive Rate = $\frac{fp}{(fn+tn)}$;
Precision = $\frac{tp}{(tp+fp)}$;

where, tp denotes the number of true positives (an instance that is positive and classified as positive); tn denotes the number of true negatives (an instance that is negative and classified as negative); fp denotes the number of false positives (an instance that is negative and classified as positive) and fn denotes the number of false negatives (an instance that is positive and classified as negative).

Precision is the probability that a (randomly selected) retrieved document is relevant. While recall is the probability that a (randomly selected) relevant document is retrieved in a search.

According to [95] an ROC curve visualises the following:

1. It shows the tradeoff between sensitivity and specificity (any increase in sensitivity will be accompanied by a decrease in specificity).

2. The closer the curve follows the left-hand border and then the top border of the ROC space, the more accurate is the test.

3. The slope of the tangent line at a cutpoint gives the likelihood ratio (LR) for that value of the test.

Accuracy of an experiment is measured by the Area Under the ROC Curve (AUC). An area of 1 represents a perfect test; an area of 0.5 represents a worthless test.

Accuracy of performance is defined as:

$$Accuracy = \frac{tp + tn}{tp + tn + fp + fn} \tag{3.24}$$

A rough guide for classifying the accuracy of a diagnostic test is the traditional academic point system [95]:

- 0.90-1 = excellent (A)
- 0.80-0.90 = good (B)
- 0.70-0.80 = fair (C)
- 0.60-0.70 = poor (D)
- 0.50-0.60 = fail (F)

In summary the ROC curve shows the ability of the classifier to rank the positive instances relative to the negative instances.

3.12 Summary

This chapter has provided the reader with the fundamental theoretical and conceptual background of the original and contributor work that is presented in chapters 4-7 of this thesis. More specific theoretical and conceptual knowledge that is required by the individual chapters will be presented within the relevant chapters.

Chapter-4 provides the readers with the pre-processing algorithms and techniques that are used within the context of the research presented in this thesis.

Chapter 4

Recognition of Military Personnel

4.1 Introduction

Current security challenges such as impersonation, disguise, information and identity theft etc, have made it imperative for organisations and individuals to setup surveillance systems to help improve security. It is not uncommon that military environments are under serious threats on a daily basis from terror groups or organisations. These groups seek ways to destroy a country's military force (security base): who help defend against potential domestic and foreign attacks. Suicide bombing, information gathering and leakages, insider attack etc, are various ways these individuals could attack military organisations.

Given the above observations there is the need for an automated computer vision based surveillance system to recognise military personnel wearing a given uniform type within the military camps or environments. Such a system can determine the flow of persons or personnel in and out of the environments and within, so that various discrepancies and threats as mentioned above, can be identified, validated, minimised or fully eliminated. To this end, in this chapter we design, implement and analyse an automated, computer vision based, military personnel recognition system that is based on texture, colour and SURF features. A GLCM texture implementation [96] is used to extract texture features, while a 256 bin colour histogram is used to compute the colour features. These basic features, i.e., texture, colour and SURF features are used as descriptors within the proposed appearance-based personnel recognition system. The classification of personnel is achieved using a SVM classifier, firstly on the multi-category classification task of Army, Air Force, Navy caps into camouflage and plain types, respectively and secondly, camouflage uniform classification into Army, Air Force and Navy types, respectively. In order to maximise the recognition accuracy, CFS is used to select discriminative features and improve recognition results. After the categorisation

into camouflage and plain caps, additional SURF features are used to further categorise a plain cap's badge into Army, Air Force and Navy types, accordingly.

For clarity of presentation this chapter is divided into three further sections. Section 4.2 presents the proposed system. Section 4.4 presents the experimental results and a comprehensive analysis of the performance of the proposed system. Section 4.5 finally concludes the chapter.

4.2 The Proposed System

The block diagram of the proposed system is illustrated in figure 4.1 for a quick overview and for the purpose of clarity.

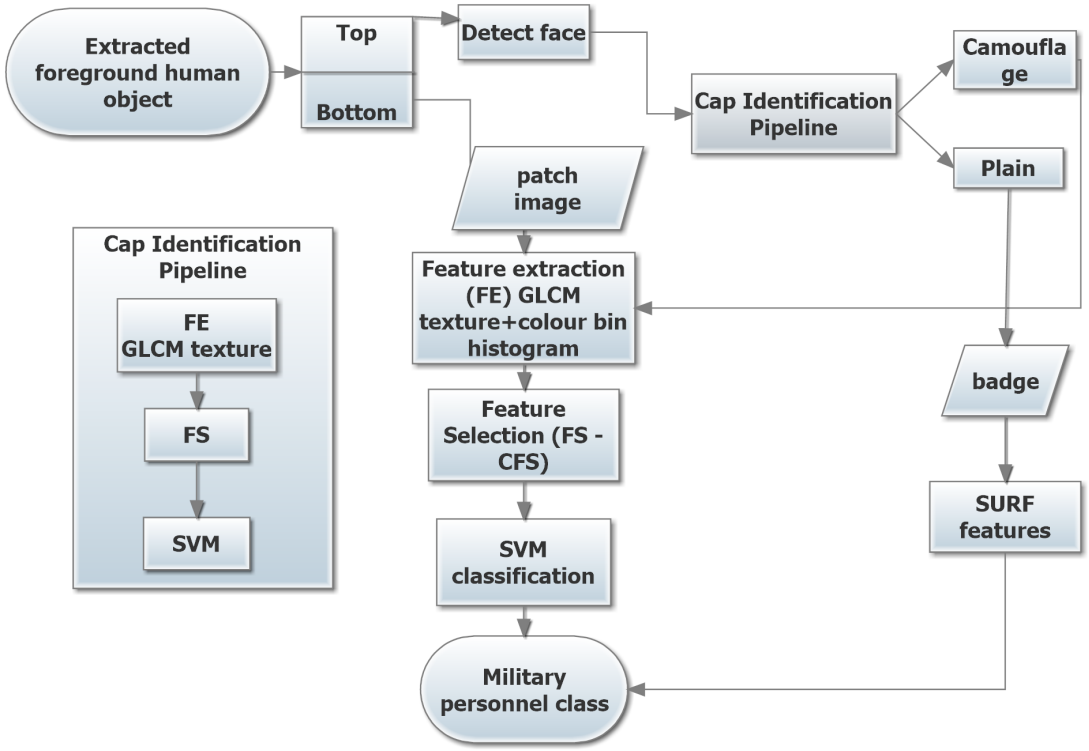


Figure 4.1: An overview of the proposed method for military personnel recognition

The above block diagram assumes that the detected moving object is a human being and the background area around the human has been removed already, i.e. an amount of pre-processing of the original video data has been performed (see section 4.2.1).

The recognition of military personnel's arm of service is based primarily on analysing the personnel's cap and the uniform. Therefore the first step is to identify the head area and the uniform area of a detected, moving, human being. A face detector is used to localise the head region, within the detected human's top part of the body, and thus determining the cap area, which should lie above the

face and thus to localise attention on the cap region. Based on texture analysis, the cap is first classified into camouflaged or plain types, i.e. caps that consist of no camouflage. If the cap is of plain type SURF features are used to recognise the type based on the recognition of the badge. If the cap is camouflaged then the analysis enters the camouflaged type recognition stage. Similarly the body region is analysed for the personnel's uniform type. This is done by first acquiring a region-of-interest of the potential uniformed area of a personnel's upper body area. Selecting the ROI results in increasing the chances of obtaining pixel samples that relate to the uniform area rather than to the background of the body. Subsequently from this sample area, texture and colour features are used to classify the uniform types based on the classification of the type of camouflage. Feature selection is used to reduce the initial, large set of parameters and improve results prior to classifying using SVM.

4.2.1 Pre-processing

Initially, the sample image is segmented to obtain the foreground object by the method in [25] which is a learning-based system for detecting and localizing objects in images using HOG features (see figure 4.2) and making further refinements using the grabcut algorithm [80] which is an interactive image segmentation technique (see figure 4.3). It is noted that as a result of using the grabcut algorithm, the human body region area has been approximately segmented.

Figure 4.2: Detected people results

Figure 4.3: Segmented foreground images using grabcut

Subsequently, the foreground of the detected and extracted human image is segmented into two parts namely the, top-half and bottom-half (see figure 4.4). From the top-half of the human body, a simple and efficient face detector which uses haar-like features and AdaBoost feature selection [97] helped identified the cap region. Therefore each object is now divided into three different regions of military clothing, namely; the bottom half, the top half (excluding the upper part including the face and cap regions) and the cap region itself.

Figure 4.4: The two segmented body parts of a detected human

Finally, the three regions are partitioned into equal patches of size 50×50 pixels (see figures 4.5 & 4.6) for the purpose of providing a reliable centre sample areas that do not contain object boundary and background pixels, for the uniform and cap area feature extraction and selection presented in this chapter.

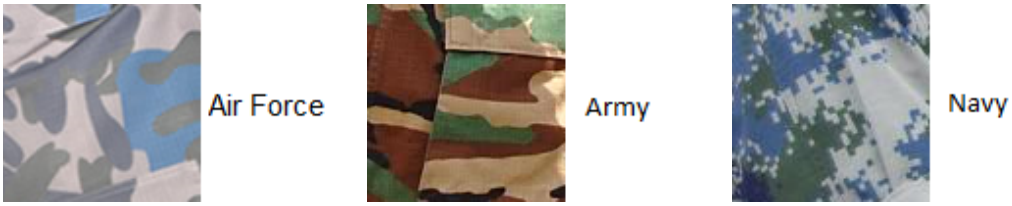


Figure 4.5: Examples of camouflage image patches

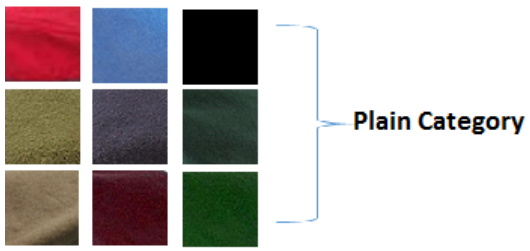


Figure 4.6: Examples of plain image patches

4.2.2 Feature extraction for camouflage type identification

A significant number of colour and texture features are extracted from all image patches (figures 4.5 & 4.6) separated in section 4.2.1 and is used for providing samples for training and testing purposes.

The basic colour and texture features extracted are as follows:

1. **Hue colour histogram**: A one dimensional colour histogram with 256 bins is obtained from the hue colour channel of the patch represented in the HSV colour domain.

2. **GLCM texture features**: Initially, a GLCM was derived for each patch using the MATLAB implementation of GLCM, "graycomatrix". Twenty two statistical texture features were extracted from the GLCM representation of the image patch. The texture features extracted from the GLCM matrix are listed as follows:

- Contrast, Correlation, Energy, Sum of squares: variance, Sum of average, Sum of variance, Sum of entropy, Difference of variance, Difference of entropy, Information measure of correlation, Information measure of correlation 2 as defined in [38]

- Autocorrelation, Cluster prominence, Cluster shade, Dissimilarity, Entropy, Homogeneity, Max probability as defined in [90]

- Inverse difference normalised, Inverse difference moment normalised as defined in [14]

- Correlation, Homogeneity as defined in [96]

The 256 colour bins were combined with the 22 texture features for the appearance based categorisation of the camouflage into army, navy, air force types in the uniformed areas excluding the cap region.

Although the same texture and colour features may be used in the categorisation of the cap into army, navy, air force types, due to the presence of both plain and camouflaged caps, initially only the texture features should be used for categorisation into either plain or camouflaged types. In the event the cap is categorised into a camouflaged type, further categorisation is done following the process described for other areas of the uniform. In the event the categorisation is that the cap is of the plain nature an attempt is then made to recognise the badge of the arms of service to categorise it into a type (see section 4.2.4). It is noted that the original RGB colour representation is first converted to the HSI colour space (see section 3.1.2.1) before bin values are extracted from the H channel as features used for recognition.

A close visual inspection of image patches illustrated in figures 4.5 & 4.6, indicates similarity of patterns but differences in colour between Army and Air Force camouflage and similarity of colour but differences in patterns between the Army and Navy camouflage. Therefore the above colour and texture features (relates to patterns) when combined should provide a reasonably accurate means of contrasting between the army, navy and air force camouflages.

4.2.3 Feature selection

To improve recognition accuracy and reduce the feature dimension and processing time, the discriminative features for classification were selected. In other words the full set of combined colour and texture features that were presented in section 4.2.2 was first reduced with the aim of optimising classification accuracy.

It is noted that feature selection helps to improve machine learning.

There are two approaches to feature selection; wrapper based and filter based approaches [35]. The method adopted here is the filter based approach CFS (see section 3.9). The CFS based approach was selected as the feature selection approach as it performed better than the wrapper based approach and is not algorithm specific [35]. The CFS approach helps to rank feature subsets according to the correlation based on the heuristic "merit" as reported by [58].

4.2.4 Recognition of arm of service

For all recognition tasks the popular SVM multi-class classifier was utilised.

We assume that the uniform of all military personnel is of camouflaged type but the cap can be of either camouflaged or plain type. The camouflage type recognition will allow one to categorise the arm of service of the military personnel. For this purpose both texture and colour features are combined and used following the feature extraction and selection stages described in sections 4.2.2 and 4.2.3

respectively. However, as the cap has high possibility of being plain, we first use only the texture features presented in section 4.2.2 alone to categorise the cap into plain of camouflaged type. If the cap is camouflaged, the camouflaged uniform type recognition process described above is followed. In contrast if the cap is plain then the categorisation is done by the detection and recognition of the type of the badge on the cap using the method in section 4.2.5.

4.2.5 Recognition of arm of service of a plain cap

Although a camouflaged cap can be recognised following steps described in section 4.2.2 (feature extraction), 4.2.3 (feature selection) and 4.2.4 (recognition), the recognition of the type of a plain cap calls a simple but completely different approach. The idea here is to detect and recognise the logo of the arm of service that will be on the front of the cap. We use SURF features (see section 3.4) and recognition using a feature matching approach.

4.3 Experimental Setup

The software tool utilised in our experiments are MATLAB and Weka using a PC configuration of intel core i5 dual core processor 2.70Ghz with 6GB ram.

We designed a model presented in figure 4.7 to run the military personnel categorisation experiment.

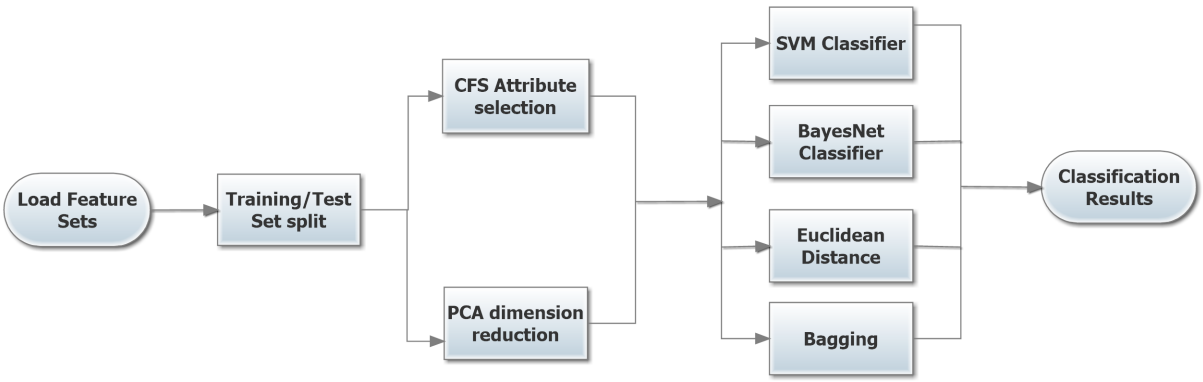


Figure 4.7: Classifiers and feature selection algorithms comparison model

Figure 4.7 above includes two feature reduction techniques; CFS & PCA and four classification algorithms; SVM, Bayes Network, Nearest Neighbour Euclidean Distance (NNED), and Bagging respectively.

From the experiments conducted using the proposed colour and texture features, the following results in table 4.1 below were obtained.

Table 4.1: Classification Accuracy using four classifiers

Classifier + Selector	Accuracy	Classifier + Selector	Accuracy
SVM + CFS	94.1%	SVM + PCA	91.4%
BayesNet + CFS	91.8%	BayesNet + PCA	85.1%
NNED + CFS	91%	NNED + PCA	83.9%
Bagging + CFS	86.7%	Bagging + PCA	84.3%

From results obtained from table 4.1, we observe that SVM + CFS gave the optimal accuracy; therefore it was adopted for all the experiments conducted in this thesis. The specific parameters used for SVM and CFS algorithms are: CFS search method - BestFirst -N 5, where N is the termination point. SVM is tabulated below:

Table 4.2: SVM parameters

Parameter name	Parameter value
complexity	1.0
epsilon	1.0E-12
filterType	Normalise training data
kernel	PolyKernel with exponent 1.0

4.4 Experimental Result and Analysis

A number of experiments were conducted to analyse the performance of the proposed military people identification approaches. For all experiments, accuracy is used as a measurement of success.

Classification accuracy is defined as:

$$Accuracy = \frac{tp + tn}{tp + tn + fp + fn} \tag{4.1}$$

4.4.1 Camouflaged type recognition army, navy, air force

An initial experiment was conducted for the classification of camouflaged uniforms into the three categories, using only using the 22 original set of texture features. For the purpose of training and testing the classifier, a total of 510 image patches (170 each from each type) were used. Fifty percent of the total sets were used for training and fifty for testing. A low classification accuracy of 71% was recorded. A feature selection using CFS selected 9 features; only maintained an accuracy figure of 68%. These experiments concluded that the texture features only cannot

be effectively used for the uniform type classification. Although increasing the training set and test set could lead to higher efficiencies a significant improvement of accuracy cannot be expected using texture features only. Including the colour features can result in an improved classification accuracy.

In the second set of experiments a total of 256 colour bin values were extracted and combined with the texture features giving a total of 278 features. However in the classification of camouflaged uniforms into Army, Air Force and Navy categories, CFS [35] was used to select discriminate features from the original 278 feature set of 22 texture features and 256 colour features to a total of 46 features that comprised of 8 texture features and 38 colour features, recording a slight improvement of accuracy to 94% from 92.5% when the full feature set was used.

Table 4.3 tabulates further performance related metrics that can be used to evaluate the performance of the proposed camouflaged recognition approach when both the full feature sets and the selected feature sets are used. The performance values of True, false positive rates, precision, recall, F-Measure and ROC area are reported. F-measure is a measure of an experiment's accuracy. It is the harmonic mean of precision and recall given as:

$$F = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \tag{4.2}$$

Table 4.4 tabulates the related confusion matrices.

Table 4.3: True, false positive rates, precision, recall, F-Measure and ROC area performance values

TP	FP	Precision	Recall	F-Measure	ROC Area	Class
			Whole features			
0.946	0.019	0.967	0.946	0.957	0.977	Air Force
0.939	0.046	0.906	0.939	0.922	0.967	Army
0.888	0.046	0.899	0.888	0.893	0.924	Navy
0.925	0.036	0.926	0.925	0.926	0.957	Weighted Average
			Selected feature sets			
0.946	0.012	0.978	0.946	0.962	0.984	Air Force
0.927	0.035	0.927	0.927	0.927	0.971	Army
0.913	0.057	0.88	0.913	0.896	0.932	Navy
0.929	0034	0.931	0.929	0.93	0.964	Weighted Average

Table 4.4: Confusion matrix for personnel recognition

Whole feature sets				Selected feature sets		
Air Force	Army	Navy		Air Force	Army	Navy
88 (95%)	0	5 (5%)	Air Force	88 (95%)	0	5 (5%)
2 (3%)	77 (93%)	3 (4%)	Army	1 (1%)	76 (92.8%)	5 (6.2%)
1 (1%)	8 (8%)	71 (91%)	Navy	1 (1%)	6 (5%)	73 (94%)

It is shown that the Air Force camouflage type has the highest level of recognition accuracy Navy indicating the least accuracy or highest level of confusion. A closer visual comparison of the three camouflage types (see figure 4.5) reveals that this is expected. Air Force camouflage has a unique colour combination and a texture pattern as compared to the Navy camouflage. The Navy camouflage in comparison on the other hand shares the green colour with the Army camouflage and a unique high-level texture pattern compared to the other two.

4.4.2 Cap type recognition camouflaged vs plain

In the classification of the cap type into plain and camouflaged categories, as mentioned in section 4.2, only the original 22 texture features (i.e. all texture features) were considered. For each of the camouflaged (i.e. Army, Air Force and Navy camouflaged, respectively) vs the corresponding plain cap classification tasks, respectively 6, 4 and 12 textures features were selected via the use of CFS in the classification process (see Table 4.5 for a summary of features selected for different classification tasks). In all of the above experiments the classifier used is the SVM classifier.

Table 4.5: Selected Features using CFS on the proposed feature sets

Camouflage categorisation	**Army Camo vs Plain**	**Air Force Camo vs Plain**	**Navy Camo vs Plain**
CP, CS, EN, ET, HG1, HG2 SE, IMC, Bins 1, 20, 31, 33, 42, 43, 45, 46, 53, 64, 67, 68, 85, 89, 90, 128, 132, 149, 151, 155, 156, 157, 159, 162, 163, 165, 168, 169, 172, 173, 174, 175, 176, 177, 202, 210,233, 252	CT, D, EN, ET, SE, IMC	CT, CR, ET, SE	CT, CR, CP, EN, ET, MP, SA, SE, DV, DE, IMC2, IDMN

Notations used: Camo - camouflage, CP - Cluster Prominence, EN - Energy, HG1 & HG2 - Homogeneity, SE - Sum of Entropy, IMC - Info measure of Correlation, CS - Cluster Shade, D - Dissimilarity, CR - Correlation, CT - Contrast, ET - Entropy, MP - Maximum Probability, SA - Sum of Average, DV - Difference of Variance, DE - Difference of Entropy, IDMN - Inverse Difference Moment Normalised.

4.4.3 Plain cap type recognition using badges

If the cap type recognition experiment presented in section 4.4.2 revealed that the cap is of the plain type (i.e. no camouflage) then the best option is to carry out its categorisation into the three types using the badge that will be present at the front of the cap. The assumption that a badge will be present that is unique to the arm of service is reasonable given the regulations governing the armed forces. The idea of using colour as to perform the discrimination will not be a sound judgment given illumination, colour constancy, colour balancing problems that are inherent in most captured videos.

Therefore we use SURF features on the cap badge, for the classification of the cap into Army, Air Force and Navy types (see figures 4.8,4.9, and 4.10). Experimental results indicate a high accuracy of recognition despite variations in scale, orientation etc., (see figures 4.11, 4.12 and 4.13).

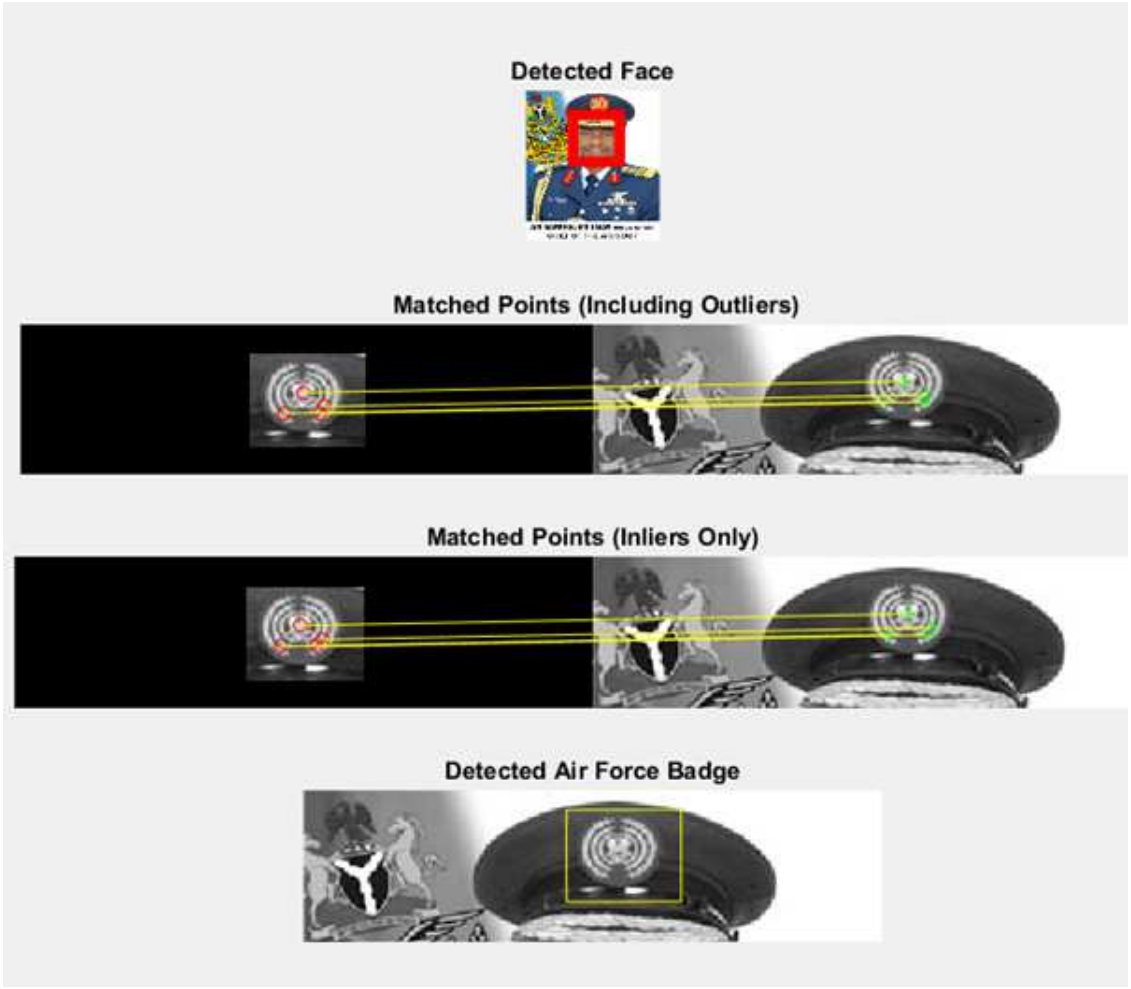


Figure 4.8: Air Force cap badge matching

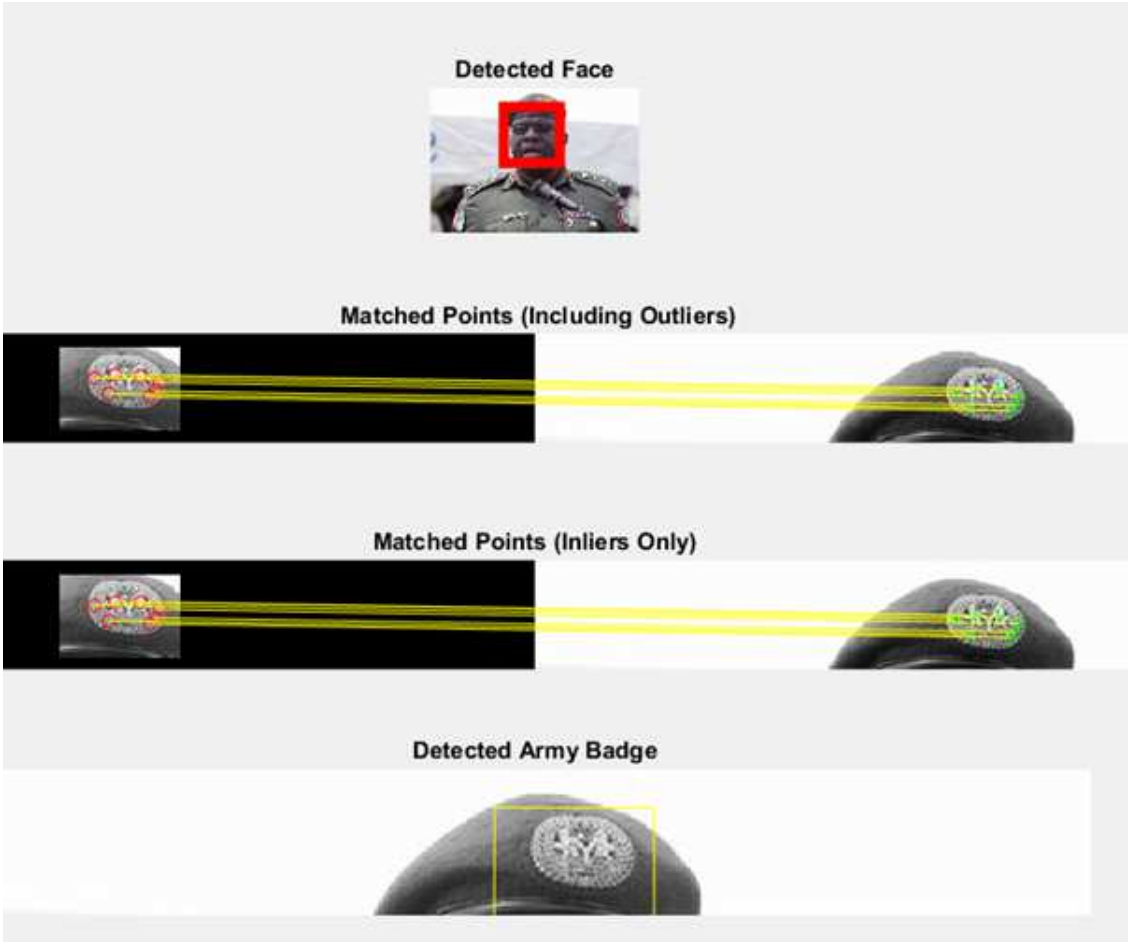


Figure 4.9: Army cap badge matching

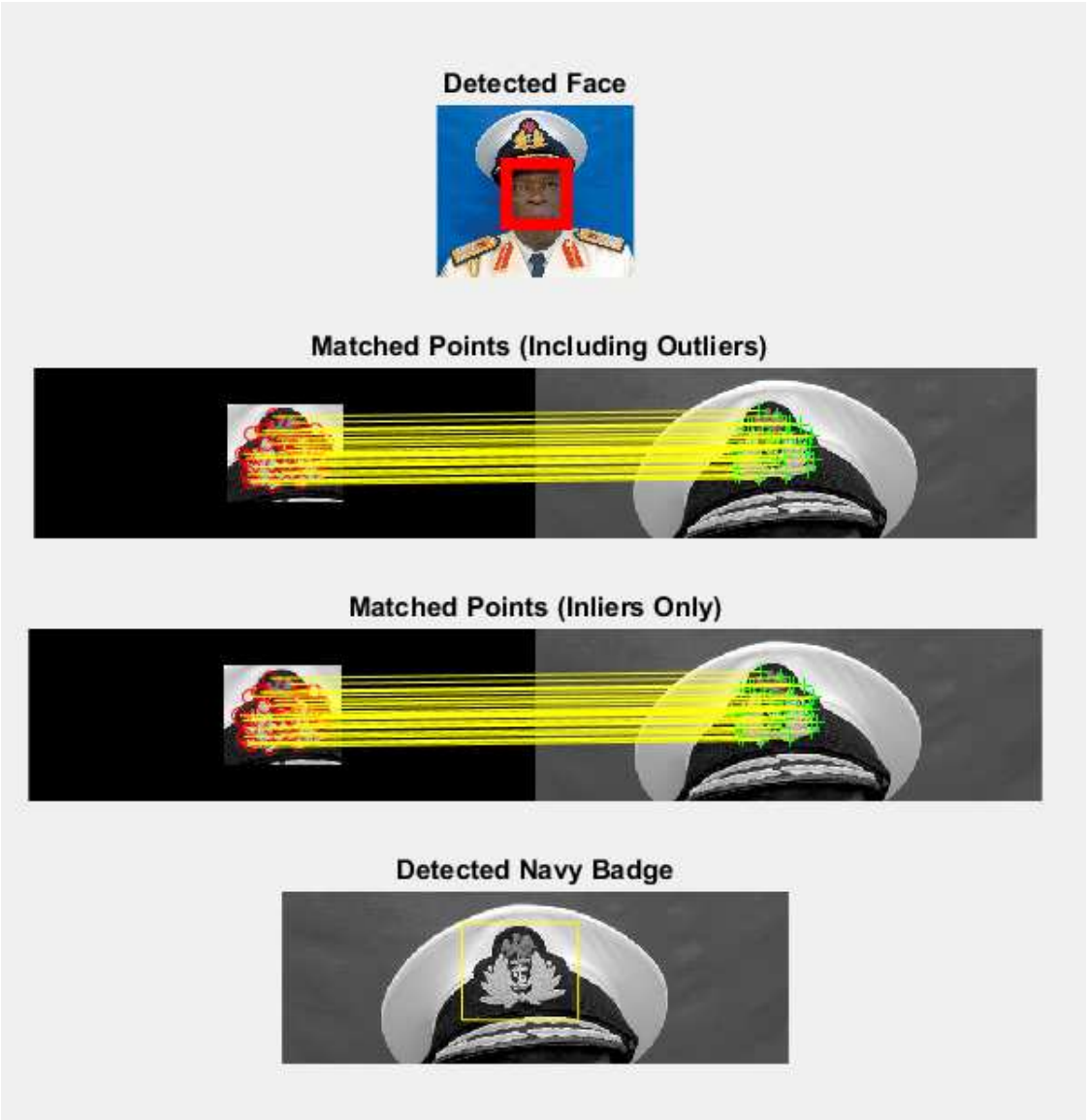


Figure 4.10: Navy cap badge matching

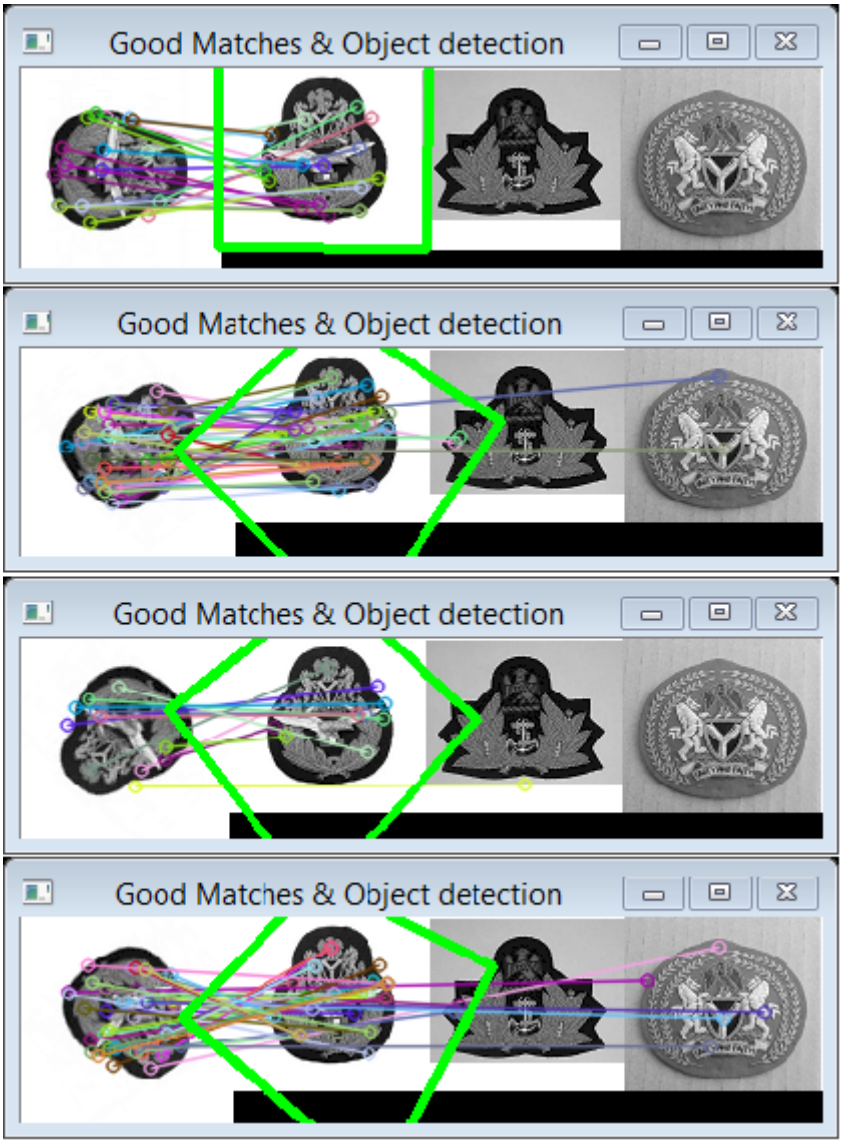


Figure 4.11: Using SURF to recognise Air Force cap badge type

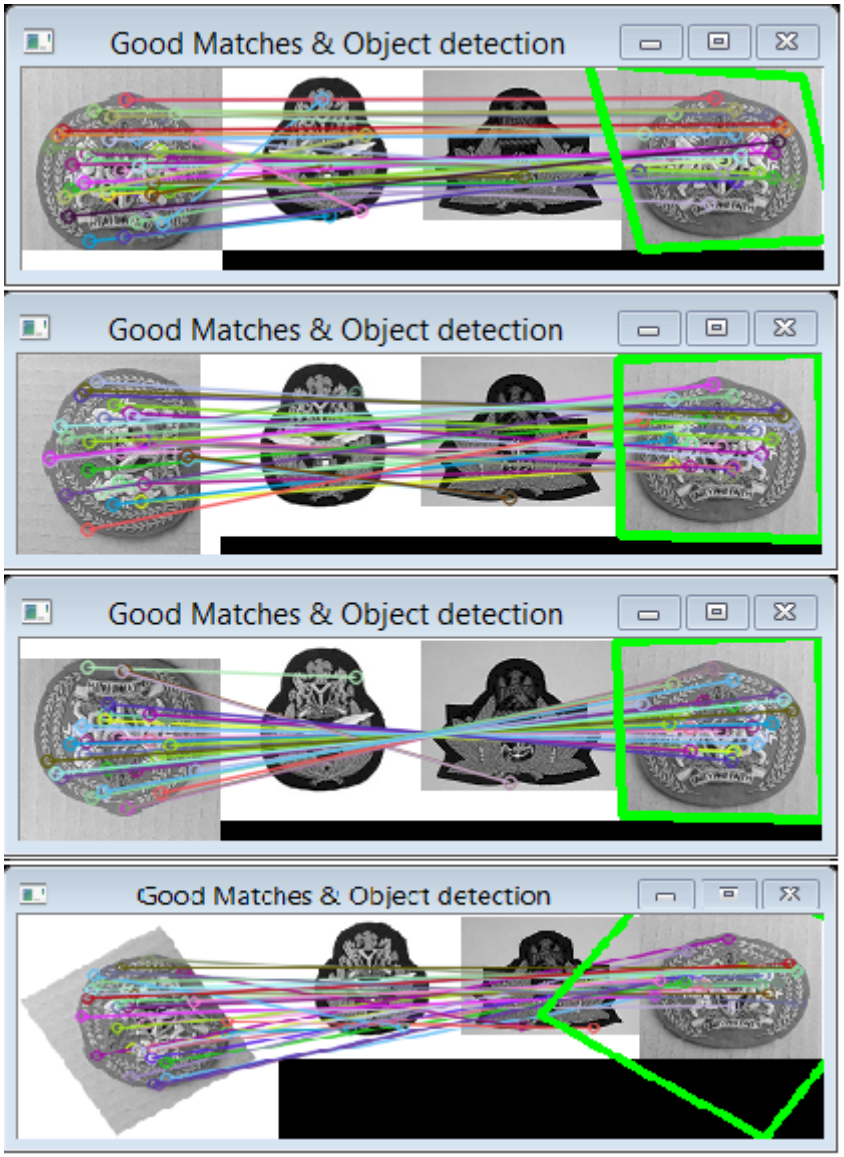


Figure 4.12: Using SURF to recognise Army cap badge type

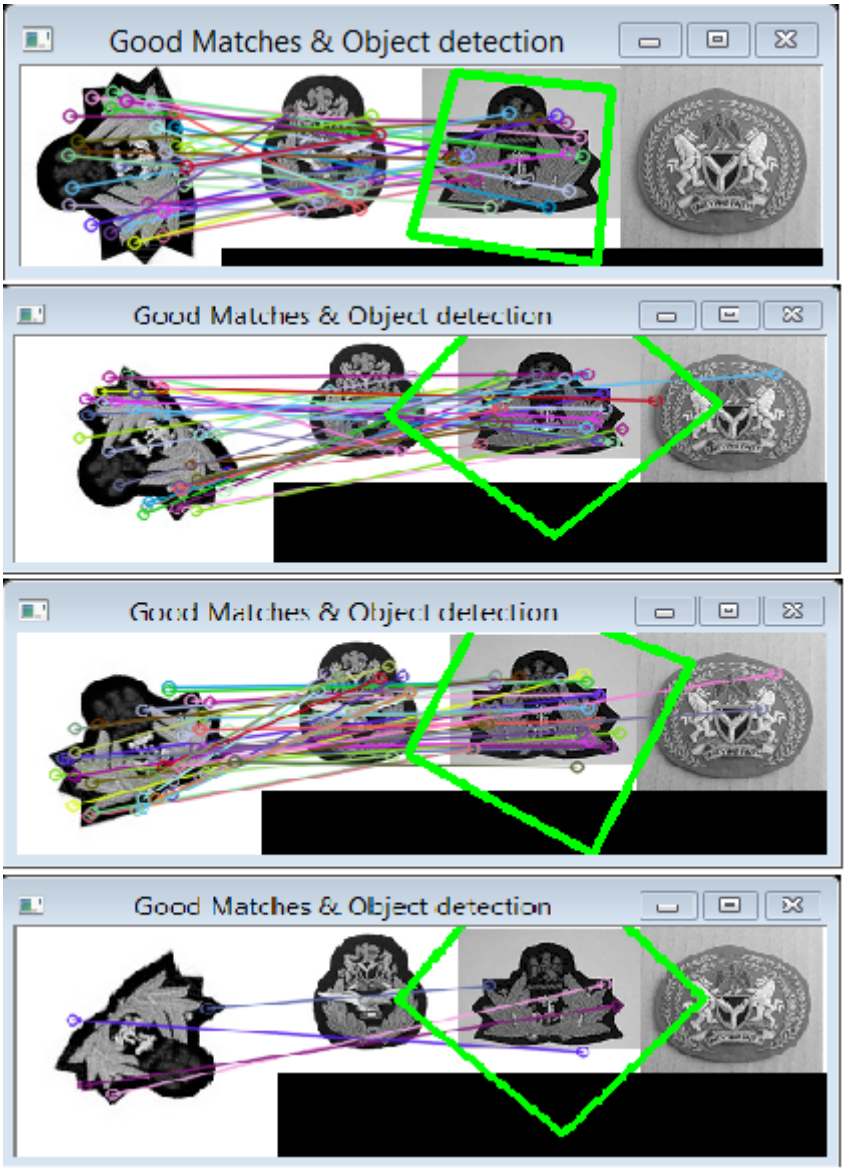


Figure 4.13: Using SURF to recognise Navy cap badge type

Figures 4.11, 4.12 & 4.13 shows cap badges of Air Force, Army and Navy with each serving as query into the three badges in the database. Result at different orientation shows accurate recognition of the three cap badges.

4.4.4 Effect on hue, saturation and texture on camouflage type recognition

Detailed visual inspections that were done on mis-classifications revealed that different levels of saturation and intensity could cause the hue to be less dominant and thus have less significance in the recognition process. Nevertheless Hue should play a more significant role as the uniforms are of particular colours combinations. In contrast, a change of hue can happen due to fading of colours on the uniform

due to wear and tear or weathering or due to changes in scene illumination (i.e. due to contrast and excess lighting/exposure) and a change of intensity can occur due to scene illumination and under exposed imaging.

Therefore, a further experiment was conducted to investigate the effect of separately using the various colour channels of the HSI colour representation in camouflaged uniform classification, i.e. using H, S and I values separately, with their combinations, alongside texture features. The specific feature combinations included in the experiments were as follows:

- GLCM texture and histogram of Saturation
- GLCM texture and histogram of Intensity
- GLCM texture and histogram of Hue
- GLCM texture and histogram of Hue and Saturation
- GLCM texture and histogram of Hue and Intensity
- GLCM texture and histogram of Saturation and Intensity
- GLCM texture and histogram of Hue, Saturation and Intensity

Experimental results were recorded as shown in table 4.6 and figure 4.14

Table 4.6: Experimental Results for camouflage classification using different combinations of colour channels and GLCM texture Features.

Features Extracted	Recognition Accuracy
Saturation and Texture	81.9%
Intensity and Texture	77%
Hue and Texture	94%
Hue, Saturation and Texture	89.8%
Hue, Intensity and Texture	92%
Saturation, Intensity and Texture	83.5%
Hue, Saturation, Intensity and Texture	89.8%

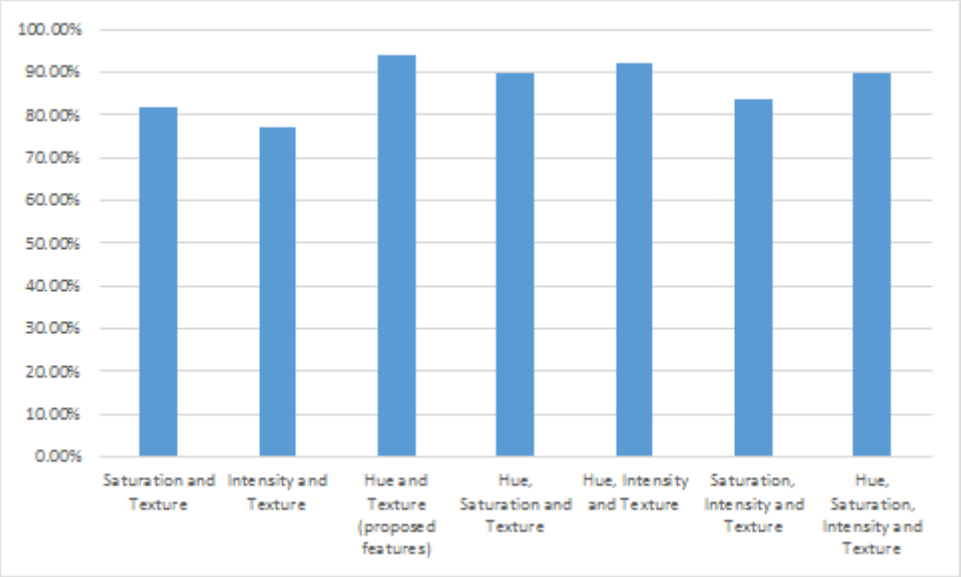


Figure 4.14: Experimental Results for camouflage classification using different combinations of colour channels and GLCM texture Features.

From table 4.6 and figure 4.14 it can be seen that the highest level of classification accuracy (94%) is achieved when using only Hue with the texture features. It is clearly better than the use of all three colour components, i.e. when Saturation and Intensity is included with Hue and Texture features. The least performance is indicated when only intensity features are included with texture. This conditions means that colour has been totally ignored from being considered. The intensity distribution that is spread on the camouflage within different colour patches contributes marginally to the case when only texture features are included (71%, see section 4.4.1) and hence shows a slightly better accuracy rate of 77%.

4.4.5 Comparison of proposed vs benchmark algorithms

We compared our technique with the techniques proposed in [68] and [58]. The result of the accuracies are tabulated in table 4.7 and illustrated in figure 4.15. For the purpose of fair comparison we implemented the technique proposed in [58] twice; **(a).** firstly, we split image patch into three segments and extract features from each segment - (Technique proposed in[58](a)); **(b).** secondly, we extract features directly from the image patch - (Technique proposed in[58](b)). For all experiments we used SVM as the classifier.

Table 4.7: Recognition accuracies for various techniques using SVM classifier

Feature techniques	Accuracy		
	CFS features	Whole features	AUC
RGB 32Bin Histogram in [68]	70%(23)	86.7%(96)	78%
Normalised 2D Histogram in [68]	45%(60)	56.5%(1024)	62%
RGB 32Bin + Shape Hist in [68]	70%(26)	87%(136)	77.8%
Local Shape Features in [68]	69%(30)	72.5%(75)	78%
Technique proposed in [58](a)	71%(12)	71%(31)	82.7%
Technique proposed in [58](b)	72.5%(20)	74%(93)	83%
Proposed technique (Hue and Texture)	94%(46)	92.5%(278)	96.4%

We can see from the above table that proposed technique recorded highest accuracy with normalised 2D histogram demonstrating the lowest recognition performance (see figure 4.15).

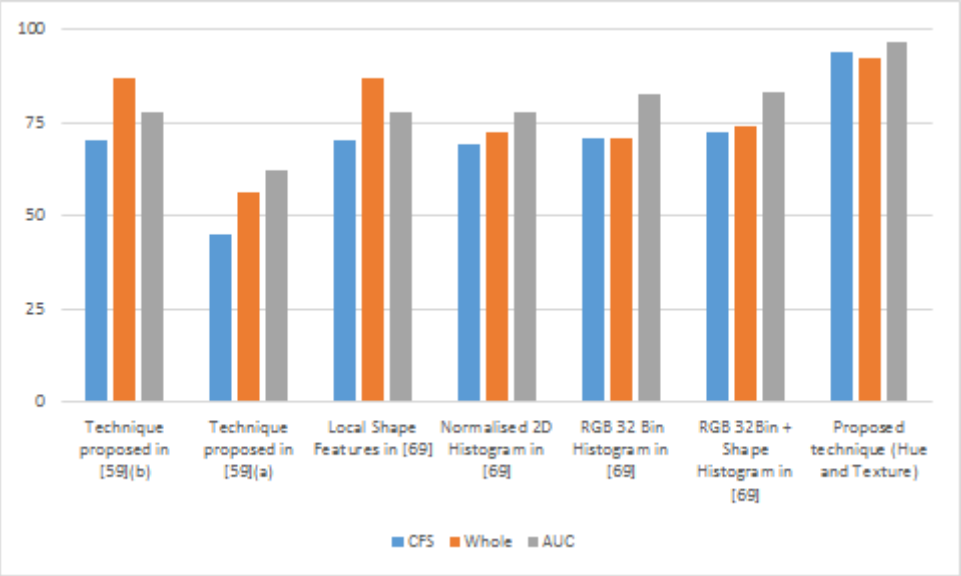


Figure 4.15: Recognition accuracies including area under curve (AUC) for various techniques using SVM classifier.

Further experimental results from the compared approaches are shown in tables 4.8 and 4.9 below:

Table 4.8: True, false positive rates, precision, recall, F-Measure and ROC area performance values for whole feature sets when using different approaches.

Proposed technique						
TP	FP	Precision	Recall	F-Measure	ROC Area	Class
0.946	0.019	0.967	0.946	0.957	0.977	Air Force
0.939	0.046	0.906	0.939	0.922	0.967	Army
0.888	0.046	0.899	0.888	0.893	0.924	Navy
0.925	0.036	0.926	0.925	0.926	0.957	Weighted Average
Technique proposed in [58] (a)						
TP	FP	Precision	Recall	F-Measure	ROC Area	Class
0.86	0.216	0.696	0.86	0.769	0.836	Air Force
0.61	0.087	0.769	0.61	0.68	0.824	Army
0.738	0.091	0.787	0.738	0.761	0.896	Navy
0.741	0.135	0.748	0.741	0.738	0.851	Weighted Average
Technique proposed in [58] (b)						
TP	FP	Precision	Recall	F-Measure	ROC Area	Class
0.806	0.21	0.688	0.806	0.743	0.819	Air Force
0.598	0.11	0.721	0.598	0.653	0.804	Army
0.713	0.12	0.731	0.713	0.722	0.872	Navy
0.71	0.15	0.712	0.71	0.707	0.831	Weighted Average
Local Shape Features in [68]						
TP	FP	Precision	Recall	F-Measure	ROC Area	Class
0.559	0.123	0.722	0.559	0.63	0.719	Air Force
0.915	0.081	0.843	0.915	0.877	0.942	Army
0.725	0.206	0.617	0.725	0.667	0.768	Navy
0.725	0.136	0.728	0.725	0.721	0.806	Weighted Average
Normalised 2D Histogram in [68]						
TP	FP	Precision	Recall	F-Measure	ROC Area	Class
0.505	0.228	0.56	0.505	0.531	0.667	Air Force
0.585	0.197	0.585	0.585	0.585	0.755	Army
0.613	0.229	0.551	0.613	0.58	0.73	Navy

0.565	0.218	0.565	0.565	0.564	0.715	Weighted Average
RGB 32 Bin Histogram in [68]						
TP	FP	Precision	Recall	F-Measure	ROC Area	Class
0.849	0.093	0.84	0.849	0.845	0.884	Air Force
0.939	0.04	0.917	0.939	0.928	0.972	Army
0.813	0.069	0.844	0.813	0.828	0.878	Navy
0.867	0.068	0.866	0.867	0.866	0.911	Weighted Average
RGB 32Bin + Shape Histogram in [68]						
TP	FP	Precision	Recall	F-Measure	ROC Area	Class
0.849	0.074	0.868	0.849	0.859	0.897	Air Force
0.939	0.052	0.895	0.939	0.917	0.967	Army
0.825	0.069	0.846	0.825	0.835	0.884	Navy
0.871	0.065	0.87	0.871	0.87	0.915	Weighted Average

Table 4.9: True, false positive rates, precision, recall, F-Measure and ROC area performance values for selected feature sets when running different approaches.

Proposed technique						
TP	FP	Precision	Recall	F-Measure	ROC Area	Class
0.946	0.012	0.978	0.946	0.962	0.984	Air Force
0.927	0.035	0.927	0.927	0.927	0.971	Army
0.913	0.057	0.88	0.913	0.896	0.932	Navy
0.929	0.034	0.931	0.929	0.93	0.964	Weighted Average
Technique proposed in [58] (a)						
TP	FP	Precision	Recall	F-Measure	ROC Area	Class
0.849	0.191	0.718	0.849	0.778	0.848	Air Force
0.585	0.11	0.716	0.585	0.644	0.79	Army
0.725	0.114	0.744	0.725	0.734	0.872	Navy
0.725	0.141	0.726	0.725	0.721	0.837	Weighted Average
Technique proposed in [58] (b)						
TP	FP	Precision	Recall	F-Measure	ROC Area	Class
0.828	0.21	0.694	0.828	0.755	0.823	Air Force
0.585	0.11	0.716	0.585	0.644	0.791	Army
0.7	0.12	0.727	0.7	0.713	0.856	Navy
0.71	0.15	0.712	0.71	0.706	0.823	Weighted Average
Local Shape Features in [68]						
TP	FP	Precision	Recall	F-Measure	ROC Area	Class
0.538	0.154	0.667	0.538	0.595	0.682	Air Force
0.817	0.087	0.817	0.817	0.817	0.927	Army
0.738	0.223	0.602	0.738	0.663	0.763	Navy
0.69	0.154	0.695	0.69	0.688	0.786	Weighted Average
Normalised 2D Histogram in [68]						
TP	FP	Precision	Recall	F-Measure	ROC Area	Class
0.57	0.222	0.596	0.57	0.582	0.703	Air Force
0.573	0.156	0.635	0.573	0.603	0.791	Army
0.7	0.206	0.609	0.7	0.651	0.768	Navy

0.612	0.196	0.612	0.612	0.61	0.751	Weighted Average
RGB 32 Bin Histogram in [68]						
TP	FP	Precision	Recall	F-Measure	ROC Area	Class
0.72	0.16	0.72	0.72	0.72	0.781	Air Force
0.695	0.133	0.713	0.695	0.704	0.814	Army
0.763	0.12	0.744	0.763	0.753	0.812	Navy
0.725	0.139	0.725	0.725	0.725	0.801	Weighted Average
RGB 32Bin + Shape Histogram in [68]						
TP	FP	Precision	Recall	F-Measure	ROC Area	Class
0.72	0.148	0.736	0.72	0.728	0.785	Air Force
0.732	0.139	0.714	0.732	0.723	0.823	Army
0.75	0.114	0.75	0.75	0.75	0.803	Navy
0.733	0.134	0.734	0.733	0.733	0.803	Weighted Average

The following plot is the weighted average of the proposed technique vs the methods in [58] and [68].

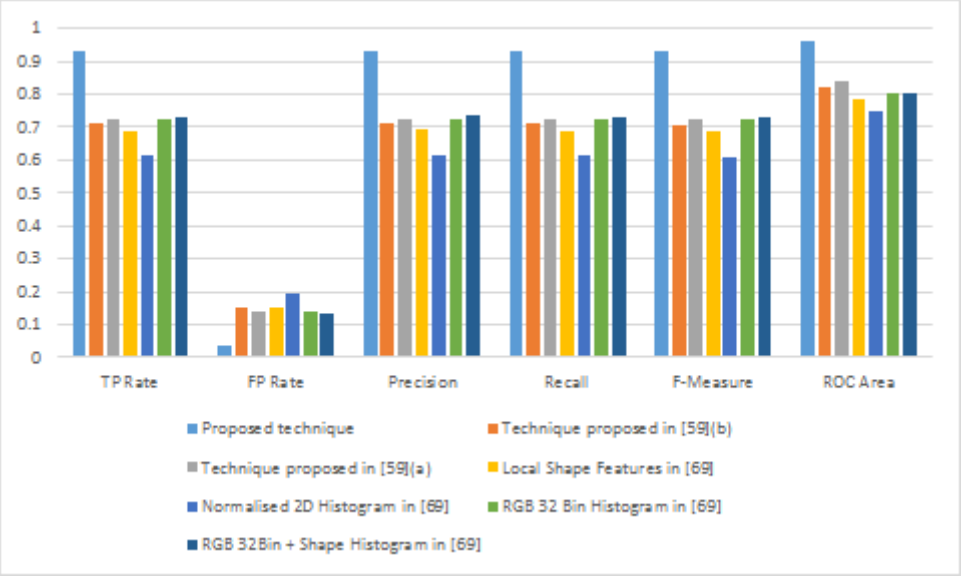


Figure 4.16: TP, FP rates, precision, recall, F-measure and ROC area plot of proposed technique vs methods in [58] and [68]

Table 4.10: Confusion matrix for personnel recognition using whole and selected feature sets on different experiments

Proposed technique						
Whole feature sets				Selected feature sets		
Air Force	Army	Navy		Air Force	Army	Navy
88	0	5	Air Force	88	0	5
2	77	3	Army	1	76	5
1	8	71	Navy	1	6	73
Technique proposed in [58] (a)						
Whole feature sets				Selected feature sets		
75	10	8	Air Force	77	10	6
20	49	13	Army	19	48	15
14	9	57	Navy	15	9	56
Technique proposed in [58] (b)						
Whole feature sets				Selected feature sets		
80	9	4	Air Force	79	10	4
20	50	12	Army	18	48	16
15	6	59	Navy	13	9	58
Local Shape Features in [68]						
Whole feature sets				Selected feature sets		
52	5	36	Air Force	50	6	37
7	75	0	Army	13	67	2
13	9	58	Navy	12	9	59
Normalised 2D Histogram in [68]						
Whole feature sets				Selected feature sets		
47	17	29	Air Force	53	18	22
23	48	11	Army	21	47	14
14	17	49	Navy	15	9	56
RGB 32 Bin Histogram in [68]						
Whole feature sets				Selected feature sets		
79	4	10	Air Force	67	13	13
3	77	2	Army	17	57	8
12	3	65	Navy	9	10	61
RGB 32Bin + Shape Histogram in [68]						
Whole feature sets				Selected feature sets		
79	4	10	Air Force	67	13	13
3	77	2	Army	15	60	7
9	5	66	Navy	9	11	60

For the purpose of detailed analysis of the performance of the proposed approach, the classification performance is finally evaluated using the ROC curve that helps visualise performance, in detail. The ROC curves for the proposed approach and those of [58] and [68] are illustrated in figure 4.17.

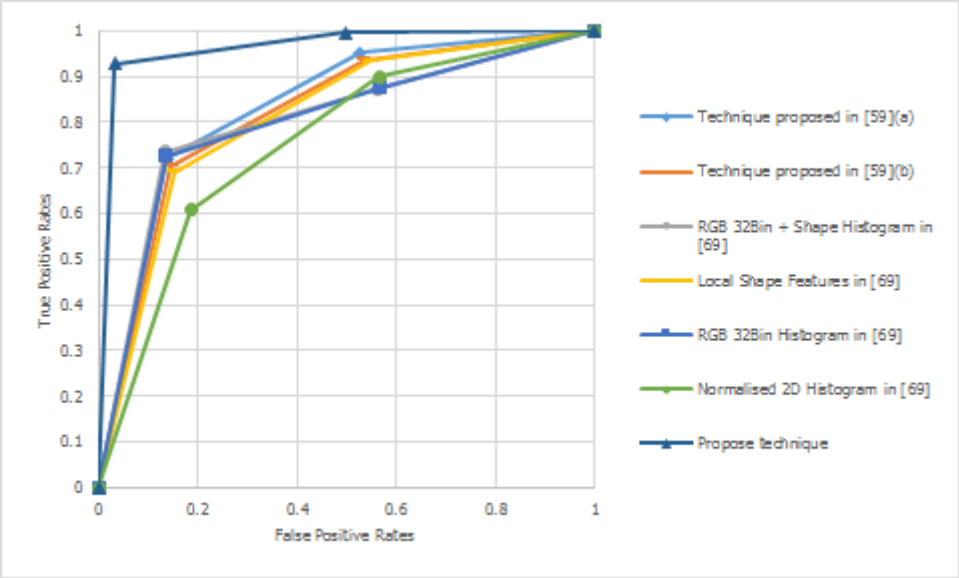


Figure 4.17: Proposed technique vs [58] and [68] techniques

From the ROC curve in figure 4.17, we see that the proposed approach has an excellent performance compared to other techniques with AUC of 96%. Method in [58] was good with [68] performing fairly; except with one technique which was poor at 62% AUC.

4.5 Conclusions

This chapter proposed an appearance-based technique that help recognise the arm of service of a military personnel based of the identification of the camouflage of the uniform and the recognition of the badge on the cap in cases where the cap is plain. The camouflage type categorisation into army, navy and air force indicated a 94% accuracy when the original feature set comprising of 278 features was reduced to 46 features. In the case where the cap was plain (i.e. no camouflage) a texture feature based classification approach using a support vector machine gave accuracy levels of 100%, 90% and 100% in the identification of Army, Navy and Air Force categories respectively using SURF matching technique.

The proposed systems can be implemented within a fully automated, real-time, military arm of service recognition system that integrates the various classification algorithms presented in the chapter, within a single logically organised unit. Such

a system can help a military base to check and monitor the following: appropriate or inappropriate dressing, absence from duty post, impersonation, disguise and completeness or incompleteness of personnel presence in military camp or environment.

Chapter 5

Vehicle Type Recognition

5.1 Introduction

In the recent past, concerns directly associated with vehicle related crime have risen internationally. As described by the Interpol report in [42] vehicle crime is:

> *...a highly organized criminal activity affecting all regions of the whole world and with clear links to organized crime and terrorism. Vehicles are not only stolen for their own sake, but are also trafficked to finance other crimes. They can also be used as bomb carriers or in the perpetration of other crimes.*

To this end, there are many situations where access to vehicles needs to be monitored automatically to manage and control their movement to and from secure sites, motorways and across international borders. Although number plate recognition provides a level of security based on the license information gathered, in an era where vehicle cloning prevails, any additional vehicle identification data can help to improve the robustness against such unlawful activities. The identification of vehicle type and keeping a count of each type passing certain known locations will help this process.

Further, the increase of the cost of building and maintaining motorways have forced many governments to consider privatising motorways resulting in a need for toll collection from their users. The number of toll roads present is growing fast internationally and so is the crime rate to avert the payment of the correct toll. Toll is normally charged based on vehicle type and the varied tariffs used means that when a human observer is not present the systems can be fooled by a vehicle that is charged a higher rate being driven through a gate meant to be for a type that is charged less. The automatic identification of the vehicle type can help take preventive measures to stop this crime.

The exponential increase of road traffic over the years has caused serious concerns about the level of pollution caused by vehicular traffic. Especially the production of ozone is considered a by-product of nitrogen dioxide caused directly due to internal combustion engines, when exposed to direct sunlight. The larger the power of a vehicle engine, the larger would be the impact it will have on the creation of the secondary pollutants such as, ozone. The counting of various types of vehicles that uses a motorway on an hourly or daily basis will help in estimating the emitted and formed air pollutants from vehicles [74]. Simply detecting vehicles and tracking them will allow the monitoring of total road usage and the estimation of their speed that also has an impact on approximating the pollution levels [74].

The above needs solicits the importance of the design, development, implementation and the installation of a computer vision based automated vehicle counting and type recognition system, which is the key focus presented in this chapter.

In observing the techniques proposed in literature summarised in section 2.3, it can be concluded that vehicles are recognised and classified at different angles under different conditions using different feature sets, classification techniques and hence algorithms. In other words, a change of camera angle requires a change of features that needs to be extracted for classification. The classification technique that performs best will also change. Further, most techniques have been tested either on rear or front views only. In practice once a camera is installed in an outdoor environment with the hope of capturing video footage for vehicle type recognition, it is likely that due to wind or neglect in installation, the camera could turn in due course. If the vehicle type recognition system was dependent significantly on the angle of view, the system would thus fail to operate accurately. Further at the point of installation practical problems may be such that the camera position and orientation will have to be changed as compared to the fixed angular view that it has originally being designed for. This will either require the system to be re-redesigned using different feature sets, classifiers and algorithms or the system having to go through a camera re-calibration processes, which is typically non-trivial and time consuming. It would be ideal if at the new orientation the captured content could still be used for classification.

Given the above observations we propose a novel algorithm for vehicle type recognition and subsequent counting, which is independent of the camera view angle. We adopt a strategy that uses multiple features that are scale and rotation invariant, leading to the accurate classification of vehicles independent of the camera angle.

For clarity of presentation this chapter is divided into four sections. Apart from

this section, which introduces the reader to the problem domain and identifies the research gap based on existing work, presented in the literature in section 2.3, the remaining sections are structured as follows: the proposed vehicle type recognition algorithm is presented in section 5.2. Section 5.3 is focused on experimental results and a performance analysis with concluding remarks provided in section 5.4.

5.2 Research Methodology

This section introduces the reader to the proposed methodology, presenting in detail the functionality of each module/stage of the proposed vehicle type recognition system under three main topics: vehicular object segmentation; feature extraction; and vehicular object classification. Figure 5.1 illustrates a block diagram of the proposed system.

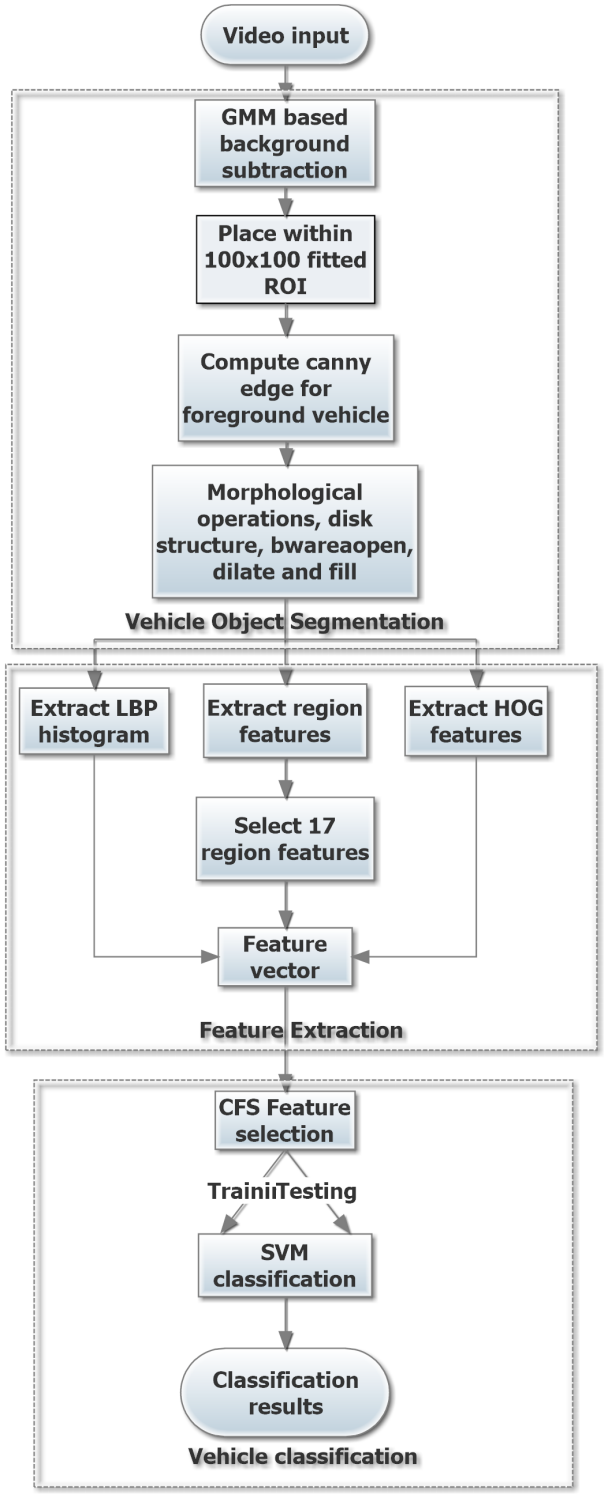


Figure 5.1: Proposed methodology for vehicle classification

The analysis of the performance of the proposed system for vehicle type recognition was conducted on datasets gathered from two low medium resolution cameras that were installed on the roadside of the Sohar Highway, Oman. They were of pixel resolution 320×240, and the frame rate was 25 FPS. The data used in the experimental analysis consisted of 10 hours video footage, captured during

daytime.

5.2.1 Vehicular object segmentation

The video frames were first segmented using a GMM (section 3.7) based foreground/background subtraction algorithm [71, 76] that detects moving objects. Due to the close-up view settings used in capturing the video footage from a motorway environment, it can be assumed that all foreground objects picked up by the above algorithm are moving vehicles only. The segmented vehicular object regions need further processing to ensure that the segmented regions more appropriately represent the true shape of a vehicle. For this purpose a Canny Edge Detector was first used to estimate the edges of the segmented object and the segmented region was subsequently refined using several morphological operators [32] that included, disk structure, bwareaopen, dilation and filling. The contribution of each of the operators in improving the segmented vehicular object shapes is demonstrated by the experimental results presented in figure 5.2.

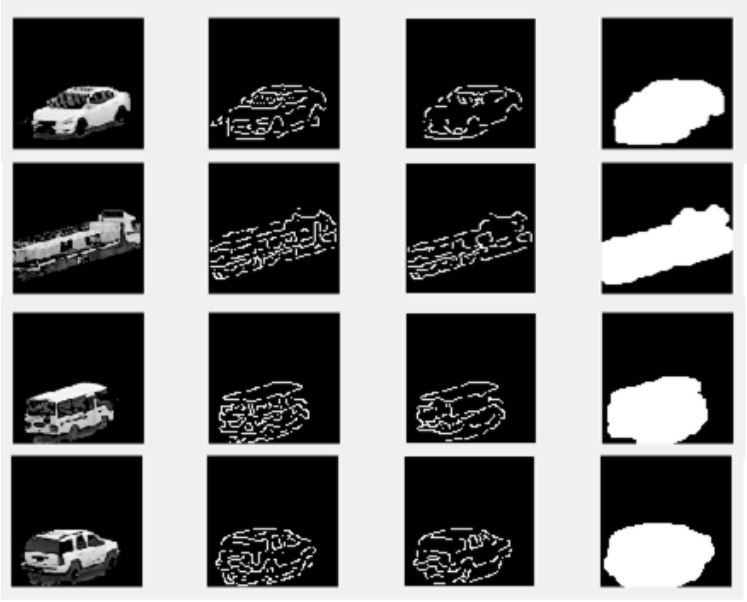


Figure 5.2: Diagram shows original vehicle, after edge detection, after removing extra edges and after dilate and fill operations respectively

After the extraction of the foreground vehicular objects, they are placed within the tightest fitting square shaped Regions of Interest (RoIs). These are subsequently resized to a normalised size of 100×100 pixels that are the regions used by subsequent stages for further processing.

5.2.2 Feature extraction

Feature extraction is performed on square shaped windows (normalised to 100×100 pixel areas) surrounding the segmented foreground objects, with the background pixels within the square area set to zero.

Firstly for the purpose of training, we manually extracted vehicle image samples normalised to a size of 100 × 100 pixels from the recorded video footage frames. Figures 5.3 and 5.4 illustrate some examples of segmented foreground objects.

Figure 5.3: Training samples some segmented vehicles from front/rear view dataset

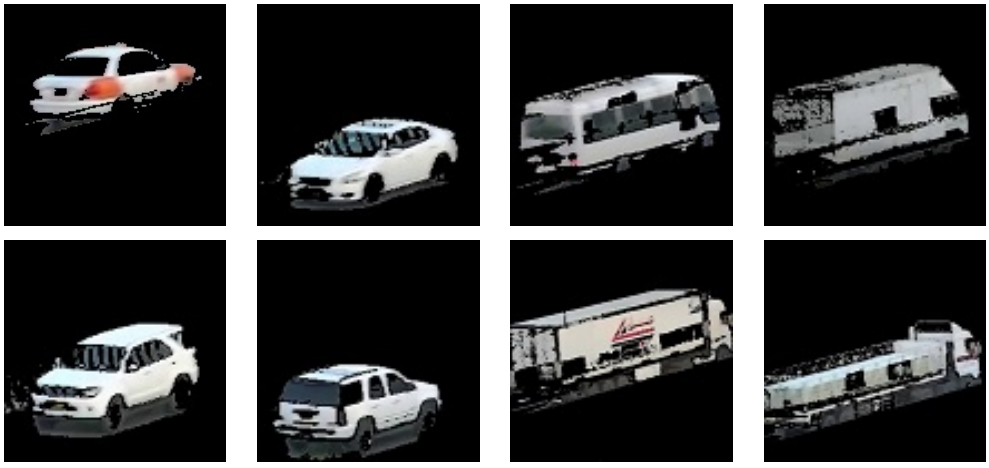


Figure 5.4: Training samples some segmented vehicles from angular view dataset

For testing purposes the ROIs are automatically segmented following the process described in section 5.2.1. Note that once the segmented foreground region is extracted it is first enclosed within a tightest fitting square area that is subsequently normalised to a size 100 × 100 pixels. The pixels within the square area but outside the object's ROI is set to zero. The features are calculated on the above mentioned 100 × 100 square regions. The following sections describe the feature extraction process.

We propose the use of 17 simple scalar Region descriptors as features, alongside HOG and LBP histogram features. They can be detailed as follows.

5.2.2.1 Region descriptors/features

We propose the initial use of 17 Region Features, which can be defined as follows:

1. **Area** in [26]: The total number of pixels that are included in the ROI within the square area.

2. **Centroid** in [26]: Horizontal and vertical coordinates of center of mass are computed as the two features that represent the centroid.

3. **Bounding Box** in [41]: The smallest rectangle containing the ROI. Bounding box feature is of the form $[x,\ y,\ width]$; where $x,\ y$ specifies the upper-left corner of the bounding box, and width is in the form $[x_{width}\ y_{width}...]$ and specifies the length of the bounding box along each dimension.

4. **Eccentricity** in [41]: The Eccentricity characteristic is the ratio of the length of the maximum chord A to the maximum chord B, which is perpendicular to the ROI enclosed within the rectangle.

5. **Major Axis Length** in [41]: The length (in pixels) of the major axis of the ellipse that has the same second moments as the ROI.

6. **Minor Axis Length** in [41]: the length (in pixels) of the minor axis of the ellipse that has the same second moments as the ROI.

7. **Orientation** in [41]: The angle (in degrees) between the x-axis and the major axis of the ellipse that has the same second-moments as the ROI.

8. **Filled Area** in [41]: The number of pixels in Filled Image; where filled image is a binary image (logical) of the same size as the bounding box of the ROI.

9. **Convex Area** in [41]: The number of pixels within the convex hull of the ROI, with all pixels within the hull filled in.

10. **EquivDiameter** in [41]: the diameter of a circle having the same area as the ROI.

11. **Solidity** in [41]: The proportion of the pixels in the convex hull that are also within the ROI. Computed as $Area/ConvexArea$.

12. **Extent** in [41]: The proportion of the pixels in the bounding box that are also in the ROI. Computed as the Area divided by area of the bounding box.

13. **Perimeter** in [41]: The perimeter is the length of the boundary of the object ROI, in pixels.

Note that since horizontal and vertical coordinates of the centroid are computed as two separate centroid features and bounding box features include four component features, namely, the x and y co-ordinates of the top left hand corner and the width and height of the bounding box there are altogether a total of 17 region features that will be considered.

5.2.2.2 HOG features

The HOG features were extracted as defined in section 3.5. An HOG feature set of length 144 was computed thus:

$$\begin{aligned} \vec{x} &= (\vec{I}./\vec{z} - \vec{y})./(\vec{y} - \vec{ol}) + 1 \\ length &= \prod[\vec{x}, \vec{y}, k] \end{aligned} \tag{5.1}$$

where $\vec{y}$ is a two element vector [2 2] (block size), k is the number of bins; 9 , $\vec{z}$ is a two element vector [32 32] (cell size), $\vec{I}$ is 100×100 (size of the image), and $\vec{ol}$ is [1 1] ($\vec{y}/2$ - block overlap).

5.2.2.3 LBP features

The LBP histogram features (section 3.6) were extracted from each image enclosed within the 100×100 rectangular area giving a 256 bin histogram.

5.2.2.4 Feature combination

In order to recognise, classify and count vehicle types, we captured appearance and shape information using the proposed feature sets; in doing so, Region Features, HOG Features, and LBP Histogram Features defined above were extracted from the segmented foreground object and were combined to form a feature vector for the classification of vehicles into four categories namely, cars, buses, jeeps and trucks respectively. The extracted Region (17), HOG (144) and LBP histogram (256) features were combined and used for the experiments.

5.2.3 Feature selection

To reduce the feature space and speed-up the processing cycle, we used the CFS [36] approach (see section 3.9) as the feature selector. CFS algorithm helps to rank feature subsets according to the correlation based on the heuristic "merit" as reported by [58]. This reduced the original feature attributes obtained from the segmented foreground vehicle objects to the minimal. In section 5.3 we showed that with feature selection, substantial accuracy improvement for vehicle classification using both types of views was achieved.

5.3 Experimental Analysis

A number of experiments were conducted to evaluate the performance of the proposed algorithm in vehicle type recognition. The experiments were conducted on video footage captured by a general purpose, non-calibrated, CCTV camera installed on the side of Sohar Highway, Oman, in the city of Sohar. As the robustness of the algorithm to the vehicle's angle of approach to the camera axis and real-time performance capability are two important design criteria, further experiments were conducted to evaluate in detail the accuracy and speed of the proposed algorithm.

Two video datasets were collected for training and testing, by installing the camera appropriately to capture front/rear (F/R) views of the vehicles and side / angular views. The first dataset was collected during a short duration (15 minutes) and captured the views of the vehicles in line with the motorway lanes. This was achieved by filming from an overhead bridge with the camera installed rigidly on a tripod. The second dataset was captured over a 10 hour period of daytime and recorded footage at approximately a 45^0 angle from the direction of the movement of vehicles. It is this angle that we consider a more practical direction of view for a camera installed in the roadside. The experimental results for the two datasets are presented, combined and separated to enable subsequent, direct comparison. The idea is to prove that the proposed algorithm can produce accurate results regardless of the angle of operation as long as training has been done on sample images that have been recorded at a similar angle.

In general, the set of input-output sample pairs that are used for the training of the classifier can be represented as,

$$(x_1, y_1), (x_2, y_2), ..., (x_N, y_N) \tag{5.2}$$

where the input x_i denotes the feature vector extracted from image I and the output y_i is a class label. Since we are categorising into vehicle types, the class label y_i encodes the four vehicle types, namely, cars, buses, jeeps and trucks; while the extracted feature x_i encodes one of the combinations of the feature sets described above, i.e. Region, LBP and HOG features; 1). Region; 2). LBP; 3). HOG; 4). Region and LBP (RL) ; 5). Region and HOG (RH); 6). Region, LBP and HOG (RLH); 7). LBP and HOG (LH) respectively.

Note that all of the above seven feature set combinations were tested to determine which combinations results in the best accuracy. As two datasets were used, namely; F/R view dataset and angular view dataset. The experimental results are presented separately in sections 5.3.1 and 5.3.2 respectively, as follows:

5.3.1 Experiments on the front and rear view dataset

The dataset consisted of approximately 100 different vehicles and was split 50:50 for the purpose of training and testing. The vehicles captured and thus used in experimentation only consisted of two vehicle types, namely, cars and buses (unfortunately due to short duration of test data recording no jeeps and trucks were captured) and hence the classification was of a binary nature, i.e. into these two classes.

We conducted experiments using different feature attributes; 1). Region; 2). LBP; 3). HOG; 4). RL; 5). RH; 6). RLH; 7). LH. Various success rates were recorded. Using region features, we recorded 93% prediction accuracy when using the entire set of feature attributes and the same percentage accuracy when CFS selected 3 discriminating features from the original 17. Using LBP features only, we recorded 79% recognition accuracy using the entire set of feature attributes with significant improvement of recognition accuracy to 90% when CFS selected 8 discriminating features from the original 256. Using HOG features only, we recorded a 97% recognition accuracy using the entire set of feature attributes with accuracy dropping to 94%, when CFS selected 23 discriminating features, from the original set of 144. Using RL features, we recorded 99% recognition accuracy using the entire set of feature attributes with improvement to 100% recognition accuracy when CFS selected 8 discriminating features from the possible total of 273. Using RH features, we recorded 96% recognition accuracy using the entire set of feature attributes, with an improvement to 97% recognition accuracy when CFS selected 10 discriminating features from the total of 161. Using LH features, we recorded 96% recognition accuracy using the entire set of feature attributes, with an improvement to 97% recognition accuracy when CFS selected 24 discriminating features from a total of 400. Finally, using RLH features, we recorded 97% recognition accuracy when using the entire set of feature attributes with same level of accuracy of 97% indicated when CFS selected 16 discriminating features from the original 417.

A summarisation of these results and observations are recorded in the first third of the table 5.1.

Figure 5.5 below shows an example of a classified vehicle from the F/R dataset.

Figure 5.5: Some examples of recognised vehicle from front/rear dataset

5.3.2 Experiments on angular view dataset

The second dataset obtained at an angle of approximately 45^0 to the direction of vehicular movement consisted of sufficient number of examples of all four types of vehicles that can be used for training purposes. Therefore the classification was carried out into four categories cars, jeeps, trucks and buses respectively. A 50:50 split was used for training and testing.

We conducted experiments using all of the seven different selections of feature attributes; 1). Region; 2). LBP; 3). HOG; 4). RL; 5). RH; 6). RLH; 7). LH. Various success rates were recorded. Using region features only, we recorded 85.7% recognition accuracy using the entire set of feature attributes with an improvement to 86% recognition accuracy when CFS selected 9 discriminating features from the original 17. Using LBP features, we recorded 74% recognition accuracy using the entire set of feature attributes, with a significant improvement to 77% recognition accuracy when CFS selected 20 discriminating features from the original 256. Using HOG features, we recorded 92.7% recognition accuracy using the entire set of feature attributes with accuracy dropping to 89% recognition accuracy when CFS selected 34 discriminating features from the original 144. Using RL features, we recorded 89% recognition accuracy using the entire set of feature attributes with an improvement to 96% recognition accuracy when CFS selected 26 discriminating features from the original 273. Using RH features, we recorded a 95% recognition accuracy using the entire set of feature attributes with the accuracy dropping to 93% when CFS selected 22 discriminating features from the original 161. Using LH features, we recorded a 93% recognition accuracy using the entire set of feature attributes with an improvement to 94.7% recognition accuracy when CFS selected 47 discriminating features from the original 400. Finally, using RLH features, we recorded a 94% recognition accuracy using the entire set of feature attributes with significant improvement of accuracy to 97% when CFS selected 37 discriminating

features from the original set of 417.

A summarisation of these results and observations are recorded in the second third of the table 5.1.

Figure 5.6 below shows some examples of classified vehicles from the angular view dataset.

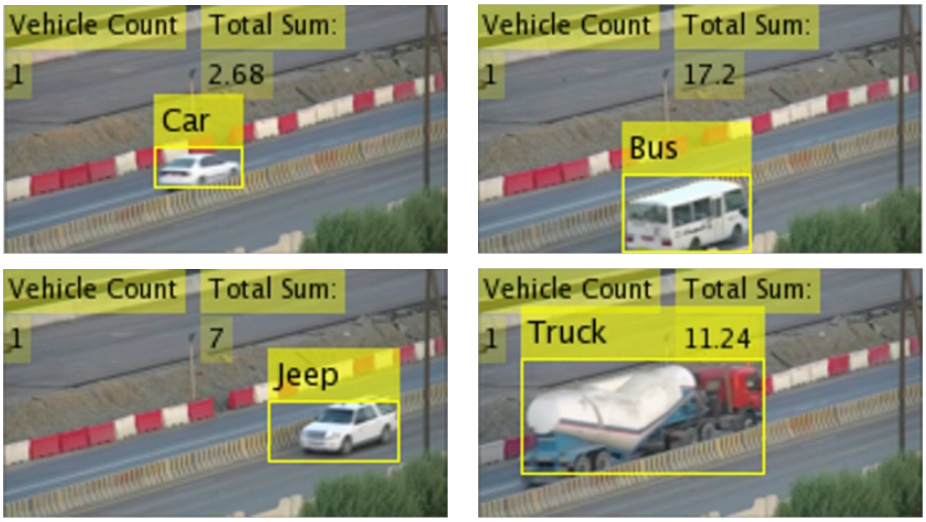


Figure 5.6: Some examples of recognised vehicle from angular dataset

5.3.3 Analysis of results

Table 5.1 summarises the recognition accuracies achieved when using the two datasets (i.e. F/R and angle), with and without feature selection. It also gives an indication of the number of features in each category, i.e., Region, HOG, LBP and their various feature combinations that remain after feature selection is applied. The table also includes experimental results when the two datasets were combined for both training and testing purposes. [Note: these are included in the bottom third of the table. CV = combined view.].

Table 5.1: Classification accuracy results with selected features

Features	Whole accuracy	Selected accuracy	Selected features	View angle
HOG	97%	94%	23	F/R
Region	93%	93%	3	F/R
LBP	79%	90%	8	F/R
RL	99%	100%	4R,4L	F/R
RH	96%	97%	3R,7H	F/R
LH	96%	97%	6L,18H	F/R
RLH	97%	97%	3R,8H,5L	F/R
HOG	92.7%	89%	34	Angle
Region	85.7%	86%	9	Angle
LBP	74%	77%	20	Angle
RL	89%	96%	9R,17L	Angle
RH	95%	93%	8R,14H	Angle
LH	93%	94.7%	22L,25H	Angle
RLH	94%	97%	7R,14H,16L	Angle
HOG	90%	87.8%	35	CV
Region	75%	74%	7	CV
LBP	74%	80%	23	CV
RL	84%	82.5%	7R,15L	CV
RH	91.5%	83.5%	9R,11H	CV
LH	89.5%	91.8%	20L,25H	CV
RLH	91.5%	90.8%	8R,12H,15L	CV

The overall conclusion when observing the results tabulated in table 5.1 is that the feature combinations RL, RH, LH and RLH performs best as against using a single set of features all being either Region, LBP or HOG features.

Results tabulated in table 5.1 shows that the experiments on the first dataset (that consists of vehicles captured from their F/R) indicates higher accuracy figures as compared to experiments with the second dataset (angular view). There are many reasons for this. It is noted that with the F/R dataset the classifications were done only between two classes, namely cars and buses. This was due to the practical reason that during the short duration (15 minutes) in which the video footage of F/R dataset was captured, only a very few samples of trucks and jeeps appeared in the footage. This made it impossible to find sufficient samples to train the classifier. Classifying between two vehicle classes which are relatively distinct (i.e. cars vs buses) as in the experiments, will be more accurate as compared to discriminating between four vehicular classes that have some class pairs, which are

harder to discriminate between (e.g. cars vs jeeps and jeeps vs mini buses). This argument is justified when analysing the confusion matrices of tables 5.2 and 5.3 for the two datasets using the feature set of RLH. Further the angular dataset was significantly larger, though producing a lower accuracy provides a more accurate and trusted estimate of the performance accuracy of the proposed approach.

Table 5.2: Confusion matrix for Angular view dataset using RLH feature

	Car	Jeep	Bus	Truck
Car	1480 (98%)	0	0	0
Jeep	60 (1.4%)	1380 (88%)	20 (0.7%)	0
Bus	20 (0.7%)	60 (1.4%)	1480 (98%)	0
Truck	0	0	0	1500 (100%)

Table 5.3: Confusion matrix for F/R view dataset using RLH feature

	Car	Bus
Car	500 (100%)	10 (1%)
Bus	20 (2%)	470 (97%)

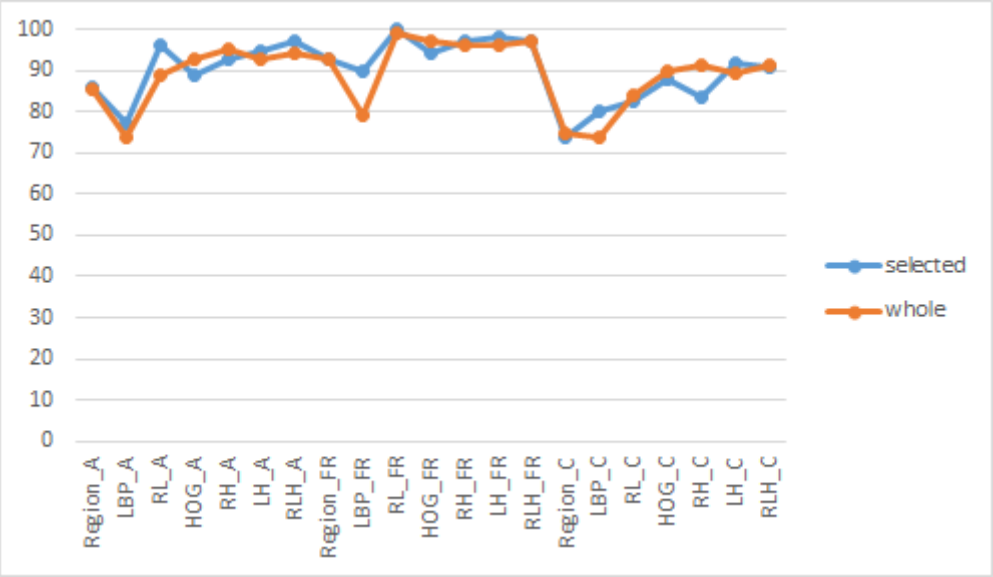


Figure 5.7: Accuracy on both datasets

Figure 5.7 plots the accuracy of various techniques, with and without feature selection, for comparison purposes. We see that the feature combination techniques, in particular the RLH technique performed generally better in all experiments.

Table 5.4: Speed of processing using varying feature attributes

Features	F/R Whole	F/R Selected	Angle Whole	Angle Selected	CV Whole	CV Selected
HOG	0.02sec	0.01sec	0.08sec	0.04sec	0.17sec	0.05sec
Region	0.01sec	0sec	0.03sec	0.02sec	0.03sec	0.02sec
LBP	0.05sec	0.01sec	0.2sec	0.03sec	0.37sec	0.04sec
RL	0.03sec	0.01sec	0.14sec	0.03sec	0.32sec	0.05sec
RH	0.01sec	0sec	0.09sec	0.03sec	0.14sec	0.04sec
LH	0.03sec	0.01sec	0.15sec	0.04sec	0.27sec	0.05sec
RLH	0.02sec	0.01sec	0.13sec	0.04sec	0.6sec	0.07sec

Using different combinations of features will result in spending different amount of time for modelling. Table 5.4, and Figure 5.8 illustrate that when the whole feature set is used, time required for modelling increased; this is due to the fact that when the number of feature attributes are large, more time is required for the modelling to complete successfully. However the careful analysis of the results also indicate that feature selection can improve the classification result and reduce the feature set to a reduced number of discriminative features that result in making the time requirement for classification minimal.

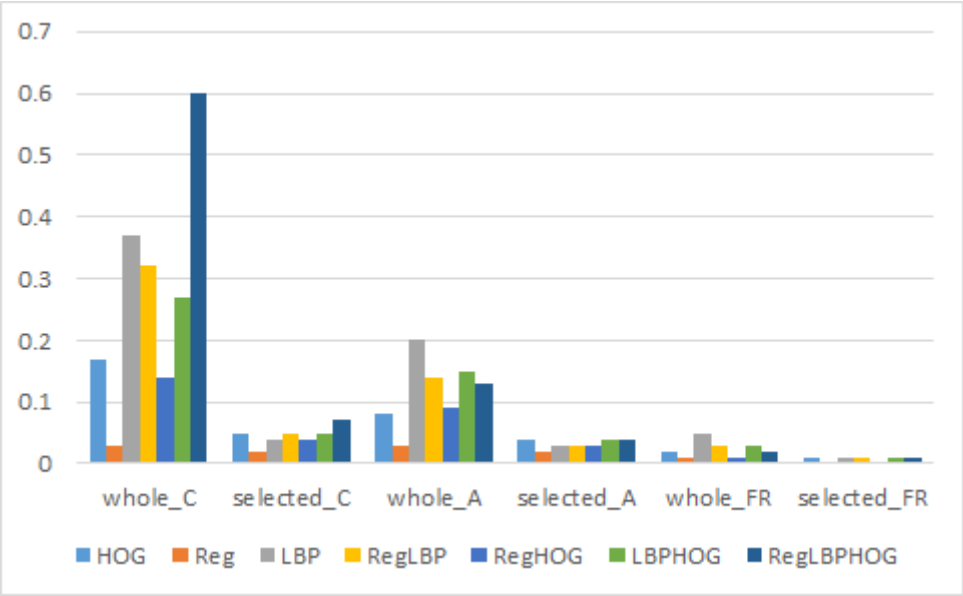


Figure 5.8: Speed of processing: whole vs selected. Notations used: C - combined view, FR - frontrear view, A - angular view

For the purpose of detailed analysis of the performance of the proposed approach, the classification performance is evaluated using the True, false positive rates, precision, recall and ROC area values of the two views and combined (see tables 5.5, 5.6 and 5.7 below for details).

Table 5.5: True, false positive rates, precision, recall, and ROC area performance values for angular view dataset

TP	FP	Precision	Recall	ROC Area	Class
Whole features					
0.959	0.027	0.922	0.959	0.972	Car
0.918	0.035	0.893	0.918	0.96	Bus
0.897	0.014	0.959	0.897	0.968	Jeep
1	0	1	1	1	Truck
0.943	0.019	0.944	0.943	0.975	Weighted Average
Selected feature sets					
1	0.018	0.949	1	0.991	Car
0.945	0.013	0.958	0.945	0.964	Bus
0.949	0.005	0.987	0.949	0.983	Jeep
1	0	1	1	1	Truck
0.973	0.009	0.974	0.973	0.985	Weighted Average

Table 5.6: True, false positive rates, precision, recall, and ROC area performance values for F/R view dataset

TP	FP	Precision	Recall	ROC Area	Class
Whole features					
0.98	0.041	0.962	0.98	0.97	Car
0.959	0.02	0.979	0.959	0.97	Bus
0.97	0.03	0.97	0.97	0.97	Weighted Average
Selected feature sets					
0.98	0.041	0.962	0.98	0.97	Car
0.959	0.02	0.979	0.959	0.97	Bus
0.97	0.03	0.97	0.97	0.97	Weighted Average

Table 5.7: True, false positive rates, precision, recall, and ROC area performance values for combined view dataset

TP	FP	Precision	Recall	ROC Area	Class
Whole features					
0.872	0.034	0.928	0.872	0.952	Car
0.904	0.04	0.911	0.904	0.95	Bus
0.925	0.042	0.816	0.925	0.953	Jeep
1	0	1	1	1	Truck
0.915	0.031	0.917	0.915	0.96	Weighted Average
Selected feature sets					
0.902	0.049	0.902	0.902	0.953	Car
0.872	0.033	0.924	0.872	0.951	Bus
0.881	0.045	0.797	0.881	0.941	Jeep
1	0	1	1	1	Truck
0.908	0.034	0.91	0.908	0.956	Weighted Average

In table 5.8, we compare the average of true, false positive rates, precision, recall and ROC area performance values on the dataset. Figure 5.9 tabulates these results as bar graphs for ease off interpretation.

Table 5.8: Weighted average of true, false positive rates, precision, recall, and ROC area performance values on all datasets

TP	FP	Precision	Recall	ROC Area	datasets
RLH feature sets					
0.97	0.009	0.97	0.97	0.99	Angle
0.97	0.03	0.97	0.97	0.97	F/R
0.91	0.034	0.91	0.91	0.96	CV
HOG feature sets					
0.89	0.035	0.90	0.89	0.96	Angle
0.94	0.061	0.94	0.94	0.94	F/R
0.88	0.045	0.88	0.88	0.94	CV
Region feature sets					
0.86	0.046	0.86	0.86	0.94	Angle
0.93	0.058	0.94	0.93	0.94	F/R
0.74	0.094	0.74	0.74	0.87	CV
LBP feature sets					
0.77	0.076	0.77	0.77	0.90	Angle
0.90	0.101	0.90	0.90	0.90	F/R
0.80	0.074	0.80	0.80	0.92	CV

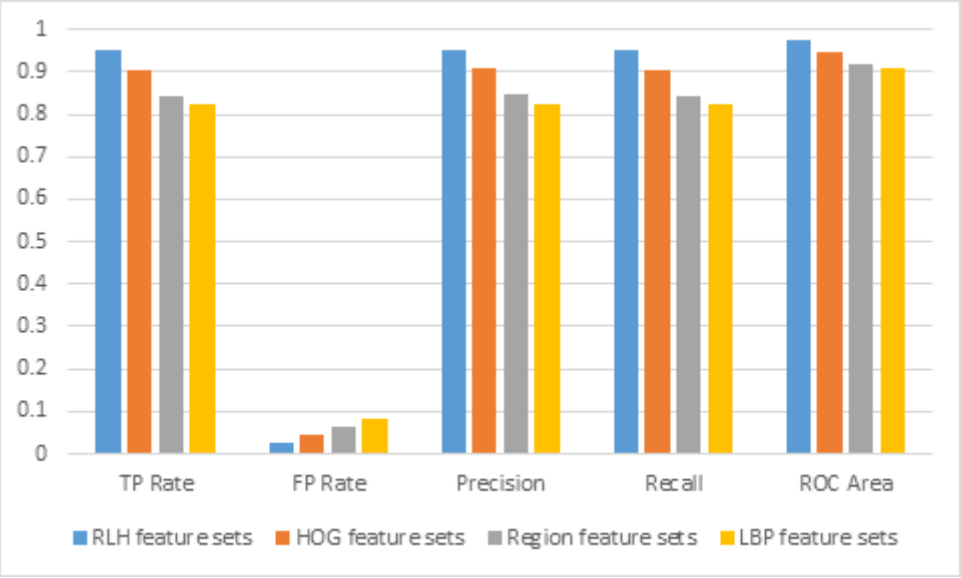


Figure 5.9: Weighted average plot of TP, FP rates, precision, recall and ROC area for RLH, HOG, Region and LBP feature sets

In summary the ROC curve shows the ability of the classifier to rank the positive instances relative to the negative instances. The table below shows the true, false positives including the AUC values on all datasets using the RLH feature

combination.

Given the above observations and facts, we plot the ROC graphs of the proposed approach when tested with the F/R datasets and angular datasets, in figures 5.10, 5.11 and 5.12 below.

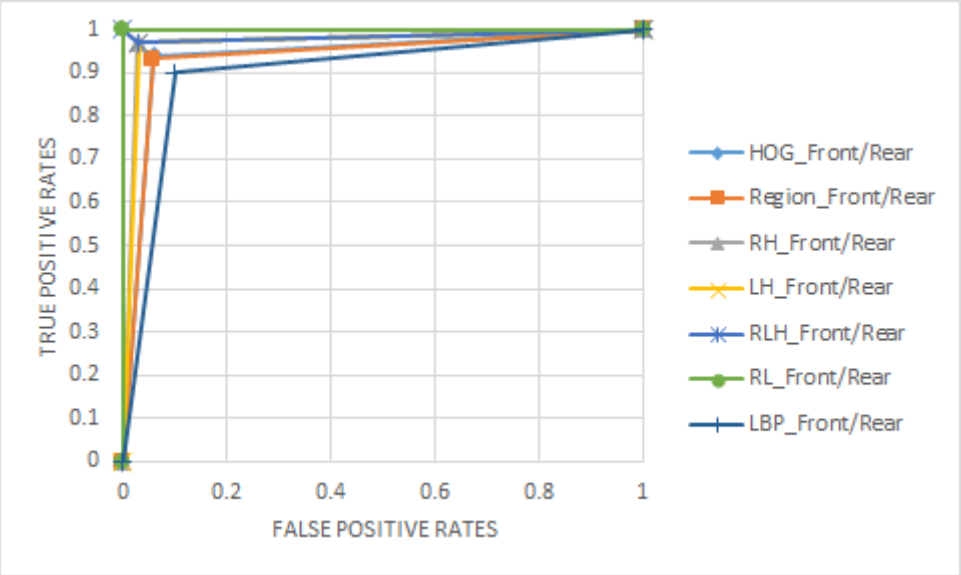


Figure 5.10: ROC curve of front/rear view datasets

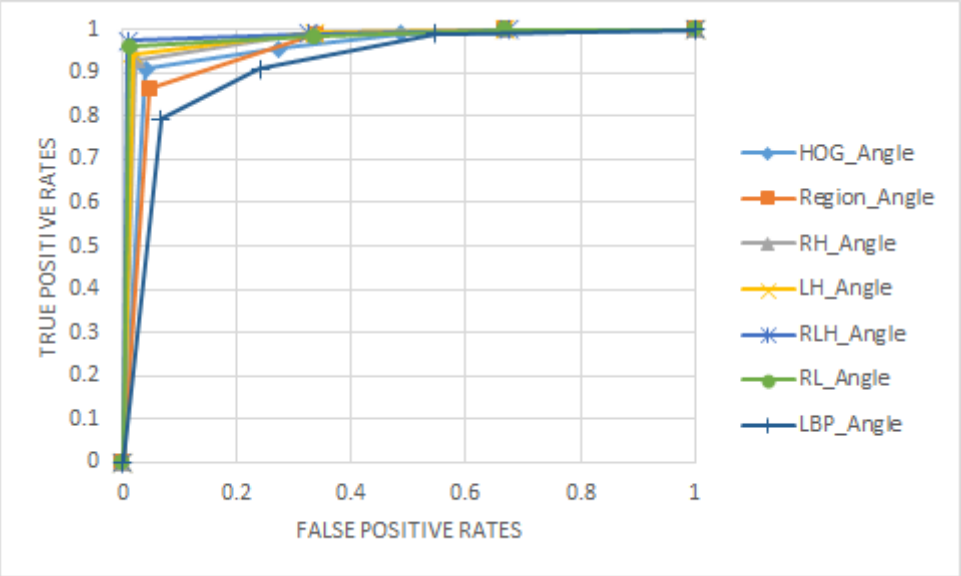


Figure 5.11: ROC curve of angular view datasets

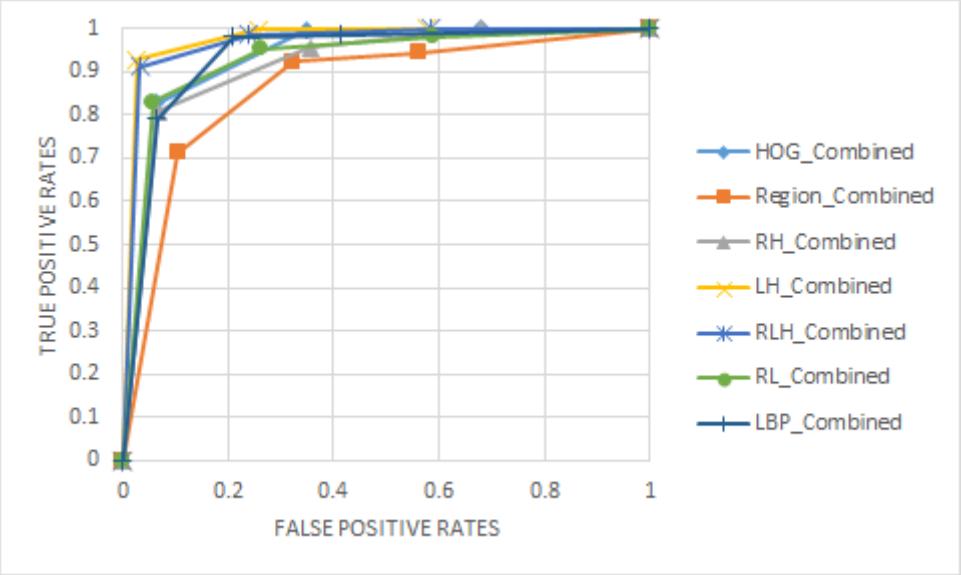


Figure 5.12: ROC curve of combined view datasets

The average AUC value for the classification of using the proposed feature combination on F/R and angular datasets is 97%, which is greater than 90% (section 3.11). Therefore the average performance (across the classification of various vehicle types) of the algorithm can be concluded to be excellent.

It is noted that each point on the ROC curve represents a TPR/FPR pair, corresponding to a particular decision threshold. The AUC is a measure of how well a parameter can distinguish between groups. ROC curves can also be used to compare performance of two or more experiments (see figures 5.10,5.11 and 5.12).

For the purpose of detailed analysis, we present the F-measure analysis of the classification performances in table 5.9.

Only the results when feature selection was used have been tabulated. From the table 5.9, we see the impact feature combinations on the percentage accuracy values indicated. It is clear that combining two or all three types of features enables a more accurate overall performance.

A final observation is that the accuracy levels are better when the training and testing are both done on footage captured within a limited angle. When all data is combined the accuracy drops. However this drop of accuracy is not significant to rule out that the proposed approaches will work regardless of the angle of approach of the vehicle.

Table 5.9: F-measure recognition percentages

Features	Angular view	Front/Rear view	Combined view	Average value
HOG	89%	94%	87.8%	90%
LBP	77%	90%	80%	82%
Region	86%	93%	74%	84%
RH	93%	97%	84%	91%
RLH	97%	97%	91%	95%
RL	96%	100%	83%	93%
LH	95%	97%	92%	94.6%

5.4 Conclusion

In this chapter we have proposed a real-time vehicle type recognition and counting system that can be re-used, independent of the direction of view. The system is based of detecting a vehicle and using a combination of features of Region, Local Binary Pattern and Histogram Oriented Gradient, to identify the vehicle type. Further we show that using a suitable feature selection approach both the speed and the accuracy of the algorithms can be significantly increased. Average accuracy figures reaching 95% has been achieved on CCTV video footage captured via a general purpose, non-calibrated camera on the side of a motorway during a ten hour recording period.

Chapter 6

Night-time Detection and Recognition

6.1 Introduction

Human object detection and vehicle type recognition have always been popular application domains that have been served by computer vision techniques since they are fundamental to a number of video analytic and surveillance scenarios and can be quite effective even as a standalone system for basic level security provision. The applications could range from video analytic/forensics, where the objective would be to analyse a crime scene for the benefit of corporate/government bodies and/or military establishments or applications that will employ them for environment monitoring related surveillance activities. Recently, human and vehicle detection have found many uses in application domains that incorporate such a need as part of some core functionality namely, intelligent transportation systems, smart vehicles and in robotics.

This ever-increasing range of applications, especially in mission-critical situations or wherein human safety may be compromised, necessitates the development of a reliable and robust human detection and vehicle type recognition system. Consequently, a number of detection and recognition techniques have been developed and are already in use. However, techniques that were initially designed for daytime images fail when applied in their original form on night-time images. The primary reason is that conventional night-time images suffer either from low light conditions or from bright and intense light sources that tend to flood the entire image, such as the dazzle of headlights from oncoming vehicles. Consequently, thermal images tend to offer a better alternative for analyzing night-time scenes than conventional night-time images. Thermal images, on the other hand, have their own drawbacks such as the lack of colour and texture information, which

may be the very features required by the aforementioned techniques.

Given the above observations this chapter proposes the use of contour-related feature extraction from thermal images, which are largely unaffected by widely varying lighting conditions. We show that the proposed technique based on the **CEN**sus **T**ransformed histog**R**am **O**riented **G**radient (CENTROG) descriptors (see section 3.8) is able to classify vehicles and detect pedestrians at night-time based on captured thermal images.

For clarity of presentation this chapter is divided into several sections. Apart from this section which is a general introduction to the problem domain section 6.2 provides the theoretical background behind CENTRIST and CENTROG descriptors. Section 6.3 subsequently details the proposed approach for human object detection and vehicle type classification. Section 6.4 provides the experimental results and a detailed analysis and finally section 6.5 concludes the chapter.

6.2 CENTRIST and CENTROG Descriptors

Census Transformed Histogram for encoding sign information (CENTRIST) is a visual description technique that was originally proposed by Wu et. al. [102] that is used to detect topological sections or scene categories. It extracts the structural properties from within an image, while filtering out the textural details. It employs the Census Transform (CT) [107] technique in which an 8-bit value is computed in order to encode the signs of comparison between neighbouring pixels. Census Transform compares the intensity value of a pixel with its eight surrounding neighbours (see example below).

Example CT:

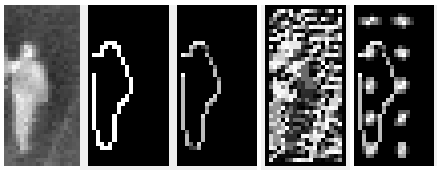
$$\begin{array}{|c|c|c|} \hline 26 & 75 & 65 \\ \hline 26 & \mathbf{46} & 22 \\ \hline 26 & 40 & 65 \\ \hline \end{array} \Rightarrow \begin{array}{ccc} 1 & 0 & 0 \\ 1 & & 1 \\ 1 & 1 & 0 \end{array} \Rightarrow (10011110)_2 \Rightarrow CT = 158$$

From CT example above, it can be seen that if the pixel under consideration is larger than (or equal) to one of its eight neighbours, a bit 1 is set in the corresponding location; else a bit 0 is set. The eight bits generated from intensity comparisons can be put together in order and converted to a base-10 value (e.g., binary to decimal conversion). This is the computed CT value for the pixel under consideration. The so-called CENTRIST descriptor therefore is the histogram of the CT image generated from an image.

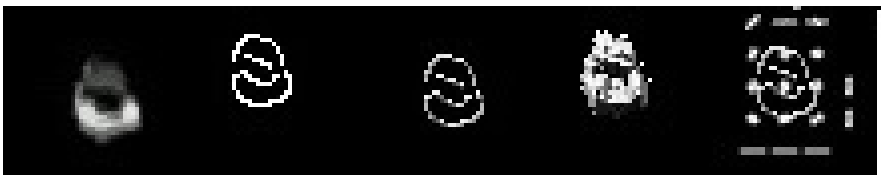
In order to compute the CENTROG features (the proposed technique), after the image structure has been captured, we compute CT on captured edge image, thereafter HOG is computed from the transformed edge image. The HOG works

by counting the occurrences of gradient orientation in localized portions of an image. The HOG captures local object appearances and shape, which can often be characterized rather well by the distribution of local intensity gradients, or edge directions as reported in [77]. Gradient is computed by applying $[1, 0, 1]$ and $[1, 0, 1]^T$ in horizontal and vertical directions within an image. Gradient information is collected from local cells into histograms using tri-linear interpolation. On the overlapping blocks composed of neighbouring cells, normalisation is performed. CENTROG descriptor therefore is the HOG on the CT generated image. The resultant images are shown below, see figure 6.1. CENTROG is a very useful technique which helps to capture local and global structure of a particular image effectively when colour and texture information are missing in a given image.

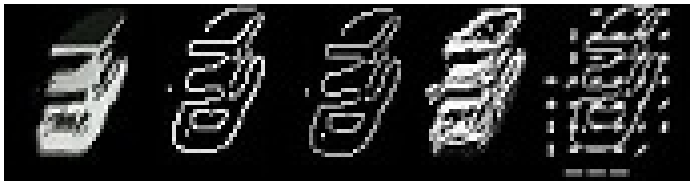
CENTROG compared against CENTRIST in pedestrian detection and vehicle type recognition. Results obtained shows that CENTROG is a better alternative for pedestrian detection and vehicle type recognition for night-time thermal images.



(a) Pedestrian sample



(b) Car sample



(c) Truck sample

Figure 6.1: Samples showing original image with processed images after CT, edge and HOG operations (a) Pedestrian sample, (b) Car sample, (c) Truck sample

The resultant images as shown in figure 6.1, Parts (a), (b) and (c) shows the; original, edge, CT-edge, CT and HOG on CT-edge images, respectively.

6.3 Proposed System Description

The proposed system consists of pedestrian detection and vehicle type recognition subsystems. These are described as follows:

6.3.1 Pedestrian detection

Night-time thermal pedestrian images of resolution 360×240 pixels were obtained from publicly available database of thermal images in [19]. Human figures were manually extracted as rectangular regions of 20×40 pixels as sample regions for training for the presence of human figures. Further, additional 20×40 pixel regions were extracted from the background regions as samples for training for the absence of a human figure. Canny edge detection (section 3.2) was applied on all of the extracted 20×40 image regions (i.e. both positive and negative sample regions for human object recognition) followed by the computation of the CT. HOG features were then extracted. The resulting feature sets were used to train an SVM classifier for pedestrian detection. The flow diagram for pedestrian detection is shown in figure 6.2 below.

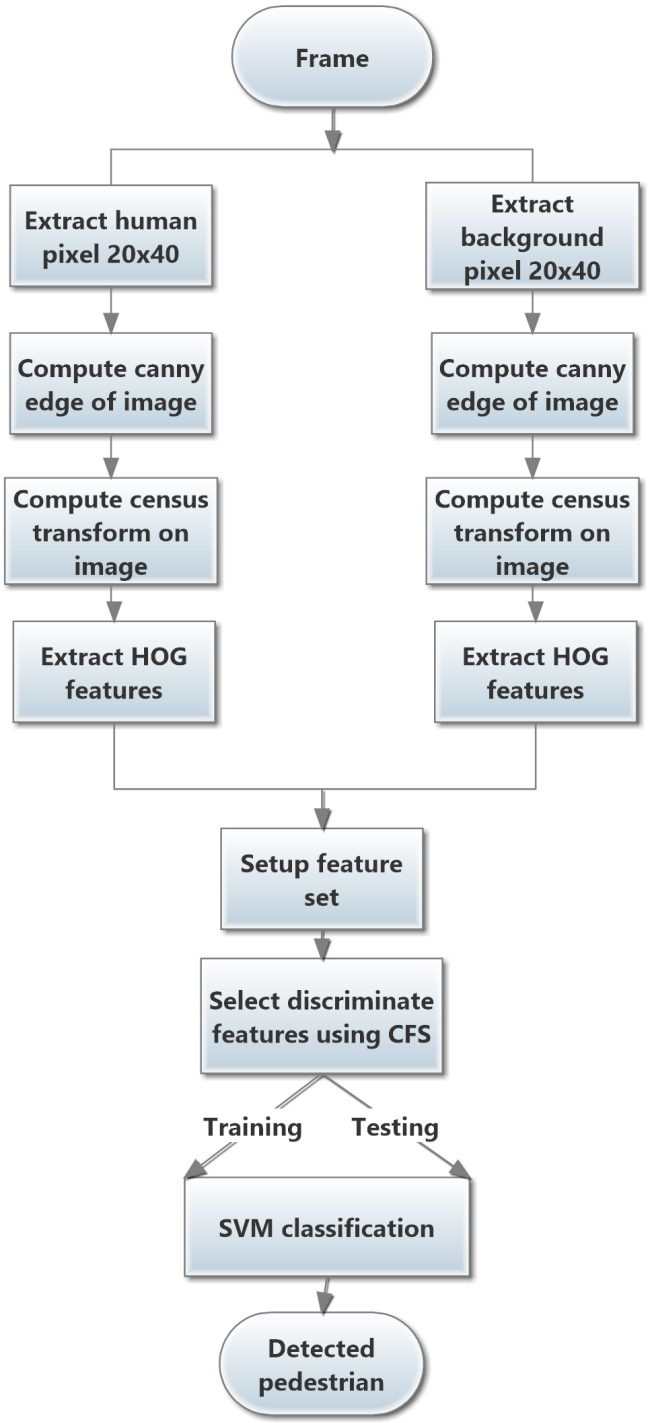


Figure 6.2: Proposed night-time pedestrian detection technique

6.3.2 Vehicle type recognition

Night-time thermal vehicle images were retrieved from a publicly available video dataset in [89], which were then segmented using the GMM (section 3.7) based background subtraction technique. The GMM technique uses a method to model each background pixel by a mixture of k-Gaussian distributions. The weight of the mixture represents the quantum of time for which the pixel values stay unchanged

in a scene.

The resolution of the video dataset used was 720×480 pixels. Within these, a ROI from co-ordinate locations [127.5, 149.5, 401, 262] was extracted as it can be assumed that all foreground objects picked up by the above algorithm in this region were moving vehicles only. These co-ordinate locations were selected since they represented a region of the image wherein the vehicles were located when it was closest to the camera and hence offered the best view. This resulted in a ROI with a resolution of 401×262 pixels, which was then resized to 100×66 in order to maintain the aspect ratio. From the dataset used, trucks and cars categories were classified using SVM binary classifier and used as a training set and a test set. Canny edge detection was applied on the extracted images followed by computation of the CT. HOG features were then extracted and employed to train the SVM classifier for vehicle type recognition. The flow diagram is as shown in figure 6.3.

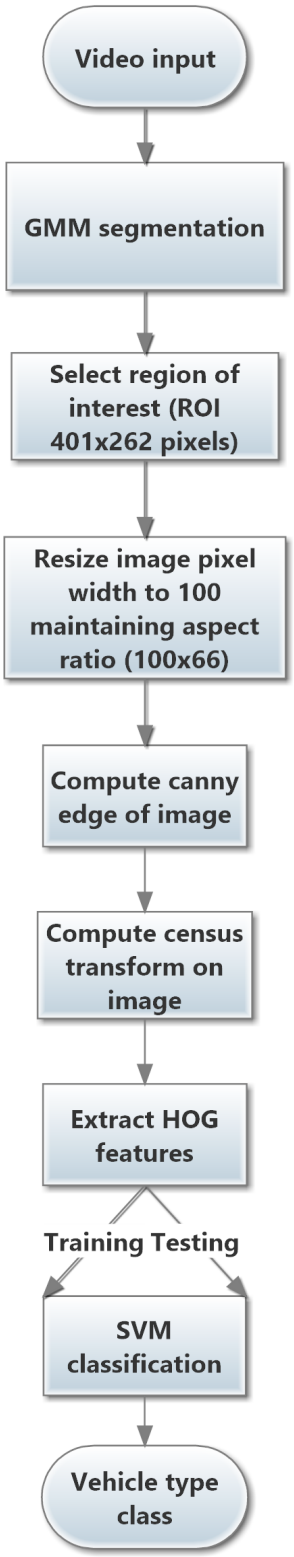


Figure 6.3: Proposed night-time vehicle classification technique

6.4 Experiments and Performance Analysis Results

A number of experiments were conducted to evaluate the performance of the proposed algorithm on pedestrian detection and vehicle type recognition in night-time thermal images. The experiments were conducted on thermal image dataset given in [19] for pedestrian detection. Similarly, thermal video dataset given in [89] was used for vehicle type recognition. The results obtain from these experiments are discussed in the following subsections.

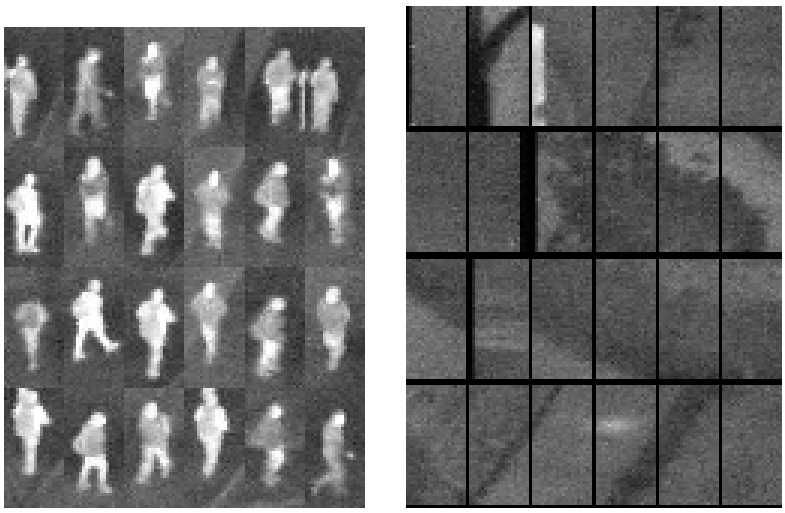
6.4.1 Pedestrian detection experiments

The following are the parameters associated with the image dataset used to capture images. The dataset comprises images captured under different environmental conditions.

Name	Description
Sensor	Raytheon 300D thermal sensor core, 75 mm lens, Camera mounted on rooftop of 8-story building, Gain/focus on manual control.
Data	Pedestrian intersection on the Ohio State University campus, Number of sequences = 10, Total number of images = 284, Format of images = 8-bit grayscale bitmap, Image size = 360 × 240 pixels, Sampling rate = non-uniform, less than 30Hz.

Table 6.1: Pedestrian Camera parameter

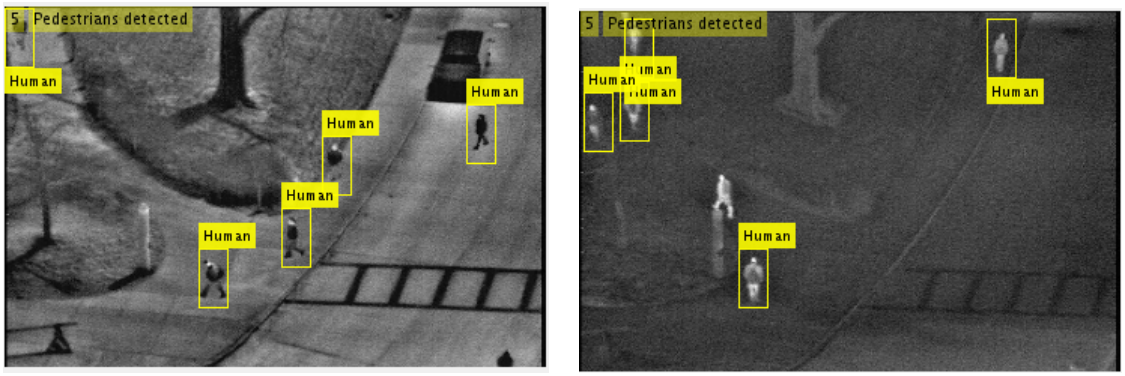
Sections within these images consisting of humans were manually extracted. These were rectangular regions of 20×40 pixels. A total of 942 pedestrian image sections were extracted, half of which were used for training and the remaining half for testing. Similarly, a total of 2494 background image sections with dimensions of 20×40 pixels were also extracted and half of them were used for training and the remaining half for testing. Samples of extracted pedestrian and non-pedestrian image sets can be seen in figure 6.4.



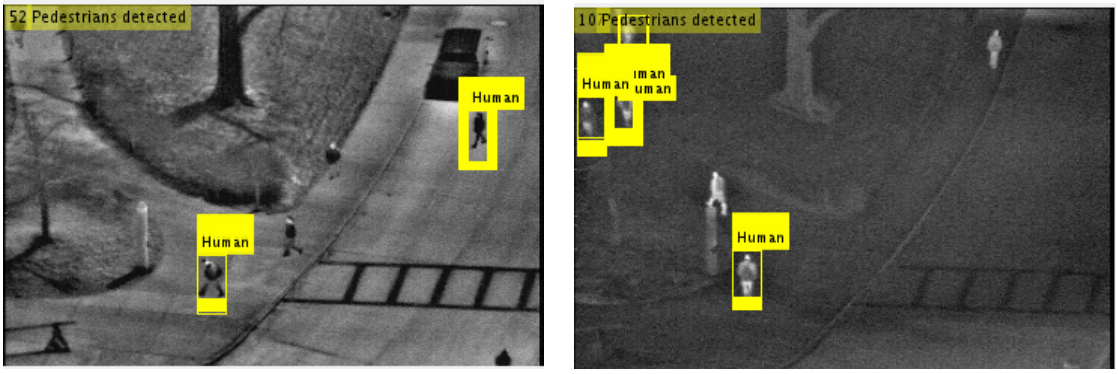
(a) Pedestrian (b) Non-pedestrians

Figure 6.4: Some examples of extracted 20×40 pixel pedestrian and non-pedestrian images

The proposed technique was evaluated using a sliding window approach for annotating detected humans accordingly. To detect pedestrians in a given image sample, the whole image is scanned with a sliding window of width 20 pixels and a height of 40 pixels. Binary classification using SVM was conducted on feature sets of length 144. Experiments were conducted using CENTROG and compared with CENTRIST feature descriptor. Experimental results obtained showed that CENTROG (the proposed technique) outperformed CENTRIST by recording a detection accuracy of 97% versus 94%. figure 6.5 shows the results of detected pedestrians using the two approaches.



(a) CENTROG



(b) CENTRIST

Figure 6.5: Detected pedestrians using (a) CENTROG and (b) CENTRIST

From figure 6.5, it can be observed that CENTRIST failed to detect some pedestrians and flagged a few false alarms, while CENTROG did not. However, CENTROG failed to detect one pedestrian due to an object that elongated the pedestrian in the image (figure 6.5a, image on the right).

Table 6.2 tabulates further performance related metrics that can be used to evaluate the performance of CENTROG based recognition approach vs the CENTRIST based recognition approach. Table 6.3 tabulates the related confusion matrices.

Table 6.2: True, false positive rates, precision, recall, F-measure and ROC area performance values for pedestrian detection

TP	FP	Precision	Recall	F-Measure	ROC Area	Class
CENTROG						
0.94	0.017	0.95	0.94	0.95	0.96	Pedestrian
0.98	0.06	0.98	0.98	0.98	0.96	Non-pedestrian
CENTRIST						
0.91	0.05	0.88	0.91	0.90	0.93	Pedestrian
0.96	0.09	0.97	0.96	0.96	0.93	Non-pedestrian

In justifying the experiment, we present the confusion matrix in table 6.3 below.

Table 6.3: Confusion matrix for pedestrian detection

	Pedestrian	Non-pedestrian
CENTROG		
Pedestrian	429 (93.5%)	28 (6.5%)
Non-pedestrian	22 (1.8%)	1239 (98.2%)
CENTRIST		
Pedestrian	421 (90%)	43 (10%)
Non-pedestrian	56 (4.7%)	1198 (95.3%)

6.4.2 Experiments on vehicle type recognition

The following are specifications of the camera used to capture the video footage within the dataset provided:

Description
FLIR SR-19 Thermal Camera, White Box, Black Box, Total Video Footage Captured: 63 min of ROBB DRIVE and 1-80 OVER-PASS

Table 6.4: Vehicle Camera parameter

After segmentation using the GMM foreground/background subtraction technique, 650 truck and 650 car images were selected, half of which were utilized for training and the remaining half for testing (see figure 6.6 for an example of

extracted vehicle image sets). Binary classification (i.e. car vs truck) using SVM was conducted on feature sets of length 2772. A number of experiments were conducted using CENTRIST and CENTROG feature descriptors. Results from these experiments showed an accuracy of 100% for the CENTROG based technique in contrast to an accuracy of 92.7% demonstrated by the CENTRIST based technique

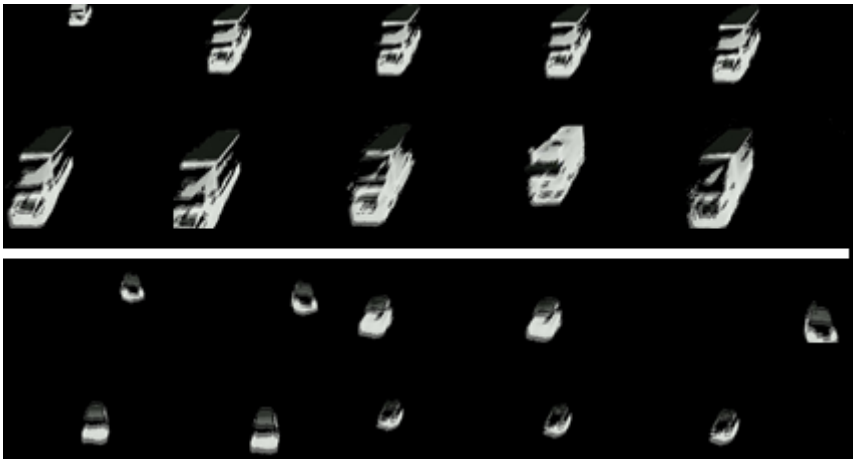


Figure 6.6: Some examples of extracted segmented vehicles

The CENTROG technique was tested on a number of randomly selected images, the results of which are depicted in figure 6.7. As can be observed, CENTROG was able to successfully recognize all vehicle types.

Figure 6.7: Classified vehicles using CENTROG feature descriptor

This is in contrast to the application of the CENTRIST technique on the same data set which resulted in some cars being wrongly classified as a truck (see figure 6.8).

Figure 6.8: Classified vehicles using CENTRIST feature descriptor

Table 6.5 tabulates further performance related metrics that can be used to evaluate the performance of CENTROG recognition approach vs the CENTRIST recognition approach. Table 6.6 tabulates the related confusion matrices.

Table 6.5: True, false positive rates, precision, recall, F-measure and ROC area performance values for vehicle type recognition

TP	FP	Precision	Recall	F-Measure	ROC Area	Class
CENTROG						
1	0	1	1	1	1	Truck
1	0	1	1	1	1	Car
CENTRIST						
0.96	0.10	0.91	0.96	0.93	0.93	Car
0.90	0.04	0.95	0.90	0.93	0.93	Truck

In justifying the experiment, we present the confusion matrix in table 6.6 below.

Table 6.6: Confusion matrix for vehicle type recognition

	Car	Truck
CENTROG		
Car	325 (100%)	0
Truck	0	325 (100%)
CENTRIST		
Car	322 (99.1%)	3 (0.9%)
Truck	1 (0.3%)	324 (99.7%)

6.4.3 Performance evaluation using ROC curves

In this section a further comprehensive performance evaluation of the proposed approach is carried out using ROC curves.

The Area Under Curve (AUC) is a measure of how well a parameter can distinguish between two contrasting groups of values.

Given the above observations and facts, we plot the ROC graphs of the proposed approach CENTROG based approach when tested on the pedestrian and vehicle datasets against the CENTRIST based approach, in figures 6.9 and 6.10 respectively.

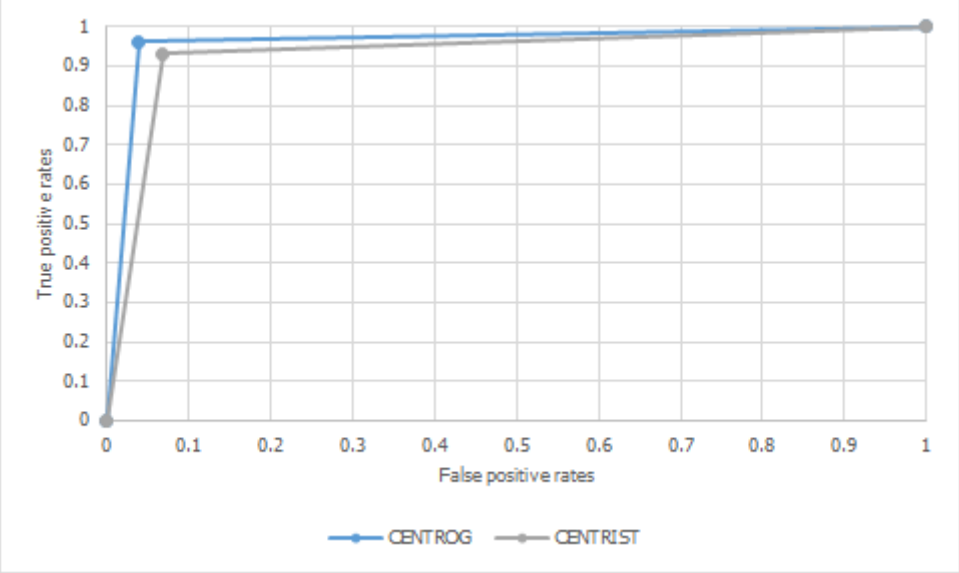


Figure 6.9: ROC curves showing the performance of CENTROG vs CENTRIST feature descriptors on the pedestrian detection experiment

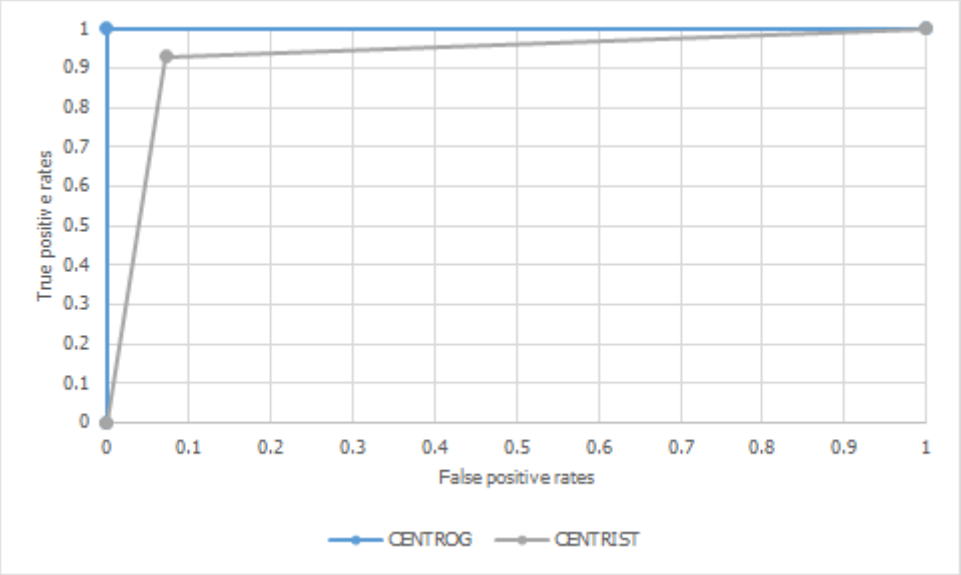


Figure 6.10: ROC curves showing the performance of CENTROG vs CENTRIST feature descriptors on the vehicle type classification experiment

From figures 6.9 and 6.10 it is seen that the average AUC value for the classification of pedestrian dataset is 96% when using the CENTROG descriptor and 93% when using the CENTRIST descriptor. Although they both have excellent performances, the proposed feature descriptor has a higher performance value of 96%. Similarly, the average AUC value for the classification of vehicle dataset is 100% when using the CENTROG descriptor and 92.7% when using the CENTRIST descriptor. Therefore the proposed CENTROG descriptor based approach out performances the CENTRIST descriptor based approach by a percentage of 7.3%, which is a significant performance improvement.

Tables 6.7 and 6.8 tabulate full performance comparison data when using CENTROG and CENTRIST feature descriptors, whilst detecting pedestrians and recognizing vehicle types on the thermal image dataset, respectively. As can be observed, the CENTROG based approach outperforms the CENTRIST based approach in detecting both pedestrians and in recognizing vehicle types.

Table 6.7: Performance analysis on pedestrian detection

Technique	True Positive	False Positive	Precision	Recall	F-measure	ROC-Area
CENTROG	97%	5%	97%	97%	97%	96%
CENTRIST	94%	8%	94%	94%	94%	93%

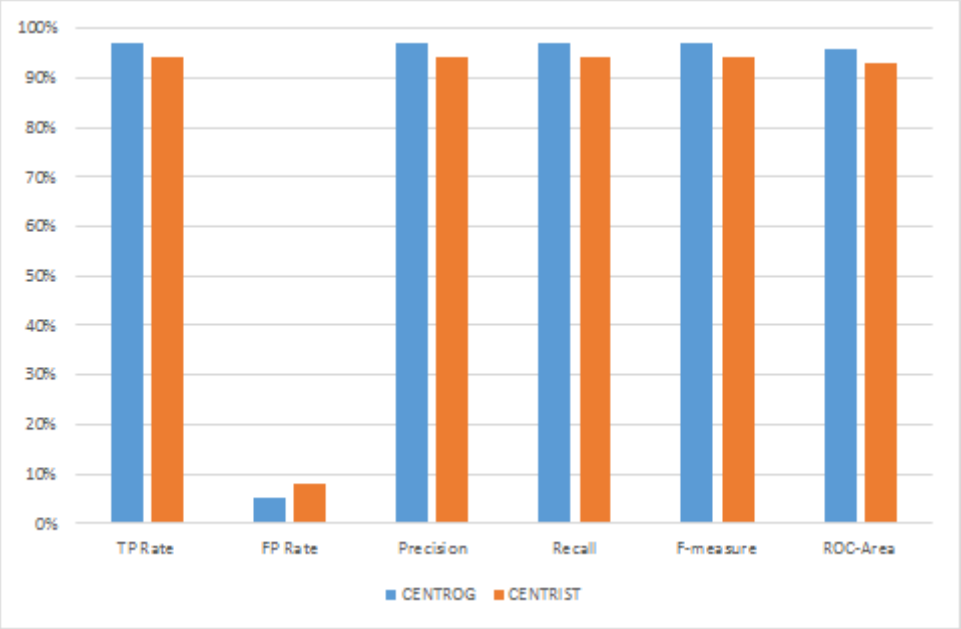


Figure 6.11: Performance analysis plot on pedestrian detection

Table 6.8: Performance analysis on vehicle classification

Technique	True Positive	False Positive	Precision	Recall	F-measure	ROC-Area
CENTROG	100%	0%	100%	100%	100%	100%
CENTRIST	92.7%	7.3%	92.9%	92.7%	92.7%	92.7%

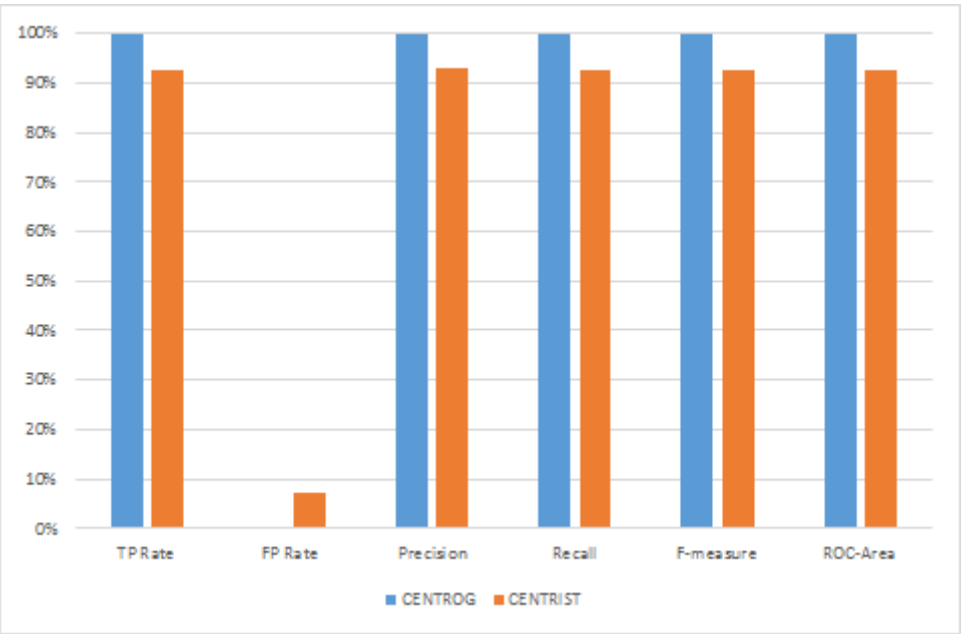


Figure 6.12: Performance analysis plot on vehicle type recognition

Further experiments were performed on the pedestrian dataset by combining

CENTROG and CENTRIST feature descriptors. Experimental result gave 98.8% accuracy, which is a slight improvement over CENTROG technique which recorded 97%.

In order to reduce the computation complexity and hence the computation time, a subset of discriminating features were chosen from the entire feature set and used within the experiments (see section 3.9).

Table 6.9 tabulates and compares the recognition accuracies, the number of features used and the processing time required for training (i.e. building the model) before and after the use of the feature selection algorithm. The improvement of speed obtainable is clear whilst maintaining the accuracy at the same level.

Table 6.9: Processing time and accuracy rates after feature selection for both pedestrian detection and vehicle type recognition

Features	Total Features	Processing Time	Accuracy	Selected Features	Processing time	Accuracy
CV	2772	1.05 secs	100%	22	0.03 secs	97.3%
CP	144	1.02 secs	97%	27	0.15 secs	95.9%
CTV	256	0.26 secs	92.7%	37	0.05 secs	90.2%
CTP	256	1.57 secs	94%	35	0.14 secs	93.2%
CTEI	512	2.54 secs	95.6%	92	0.48 secs	96%
CCT	656	1.62 secs	98.9%	92	0.33 secs	98%

Notations used: CV - CENTROG vehicle, CP -CENTROG pedestrian, CTV - CENTRIST vehicle, CTP - CENTRIST pedestrian, CTEI - CENTRIST edge and image, CCT - CENTROG and CENTRIST.

In order to improve pedestrian detection, we combined extracted features of histogram of CT image and edge image (CENTRIST Edge and Image). It recorded a slight improvement for pedestrian detection from 94% to 95.6% using the combined CENTRIST features. Similarly, we combined CENTRIST and CENTROG, which improved slightly from 97% to 98.9% for pedestrian detection.

6.5 Conclusion

This chapter proposed a feature-based technique for pedestrian detection and vehicle classification in night-time thermal images. The features were extracted by applying Histogram Oriented Gradient feature extraction on Census Transformed images and hence is termed as CENTROG. A linear SVM classifier was trained on the features obtained from the two datasets (pedestrian and vehicle). The proposed technique was implemented and compared with the CENTRIST technique.

Experimental results showed that the proposed CENTROG based approach outperformed the CENTRIST based approach in detecting pedestrians (3% realative improvement) as well as recognizing vehicle types (7.3% relative improvement), thereby exhibiting a higher detection and classification accuracy. Further experiments revealed that combining CENTROG and CENTRIST feature descriptors offered the best performance (1.9% relative improvement over CENTROG). Finally the impact of the CFS on the processing time taken for training (i.e. building the model) before and after the use of the feature selection algorithm was also analysed. Results indicated a significant reduction (1.02 seconds - vehicles recognition, 0.87 seconds - pedestrian detection) in time taken for detection and classification in contrast to employing the entire feature set. Reduction in processing time implies that the proposed technique can be employed in real-time detection and classification scenarios.

Chapter 7

People Re-identification by Low-Level Features and Mid-level Attributes

This chapter presents a novel approach to people re-identification, a task that is considered as of fundamental importance in modern video analytic/forensic systems.

7.1 Introduction

A fundamental task for a distributed multi-camera surveillance system is to recognise individuals in diverse scenes obtained using two or more cameras at different times and locations. Person re-identification is a long term people surveillance and monitoring task, where individuals or a group of people are differentiated from several possible targets in diverse scenes, obtained from different cameras distributed over a network of locations of substantial distances, in the presence of occlusions, difference in view angles, lighting conditions and time.

In a surveillance scenario, an individual disappearing from a particular camera view needs be matched with similar human objects present in one or more other views obtained at different physical locations, over a period of time, and be differentiated from numerous other human objects in the same views. In a typical surveillance / video monitoring task, it can help to find out if a particular individual who enters and exits a building is the same person identified within another different building; within a public space, work environment, university campus, school, train station, airports etc. It is noted that in answering the above question the views of surveillance footage may be taken from different, angles and distances, backgrounds, lighting conditions and various degrees of occlusions.

As reported by [47], concentrating errors, biasness, matching errors and human surveillance costs, has given rise to the need for the automation of re-identification tasks. Despite the past and present efforts to solve the automation of the re-identification problem using various techniques [30], it still remains a research area, where much research effort are needed, due to the fact that conventional biometrics such as face recognition has failed as a result of insufficient region of interest (ROI) detail for extracting robust features.

Further in exploiting other visual features such as appearance of a person, most features used in literature have not been sufficiently discriminative enough for low quality inter-camera differentiation, due to changes in a person's appearance, differences in view angles, changes in lighting conditions, presence of background clutter and occlusion etc [30].

Although in general, significant feature variations could be present in a significant variety of clothes worn by people, vast majority of public may choose to wear ordinary clothes with similar appearance in daily living. Such characteristics which bear a mid-level semantic meaning can be exploited for a person re-identification task. In this chapter, we will consider mid-level semantic attributes as valued variables for the person re-identification problem. For example, we will consider the trouser to either be coloured or bright.

In this chapter we propose a selective parts-based approach for low-level feature representation of a pedestrian and for mid-level feature attribute detection for human description. This approach helps to reduce misalignment, avoidance of the background and helps in clothing attributes detection, which help improve re-identification accuracy.

A specifically captured dataset alongside existing publicly available dataset; Viewpoint Invariant Pedestrian Recognition (VIPeR) were used in the experiments conducted.

For clarity of presentation the chapter is divided into a number of sections as follows: immediately following this section is section 7.2, which describes the proposed method for person re-identification. Section 7.3 describes how the parts of a holistic human figure were detected to enable detailed clothing attribute detection. Section 7.4 shows us the list of clothing attributes used for the proposed person re-identification task. Section 7.5 gives us the results of the various experiments performed. Section 7.6 presents experimental analysis with their respective performance results, while section 7.7 concludes this chapter.

7.2 Proposed Re-identification Framework

This section presents the operational details of the proposed human object re-identification system. The process of re-identifying a person in a video surveillance system generally includes three broad steps: human object detection; feature capture and representation and object classification (see figure 7.1).

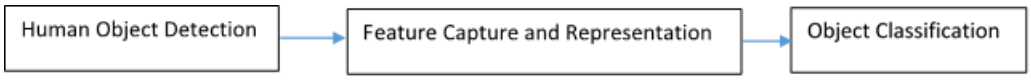


Figure 7.1: Human re-identification process

Figure 7.2 illustrates the detailed block diagram of the proposed person re-identification system. Sections 7.3 7.5 presents the underlying algorithmic details of each of the functional blocks of the figure 7.2 below.

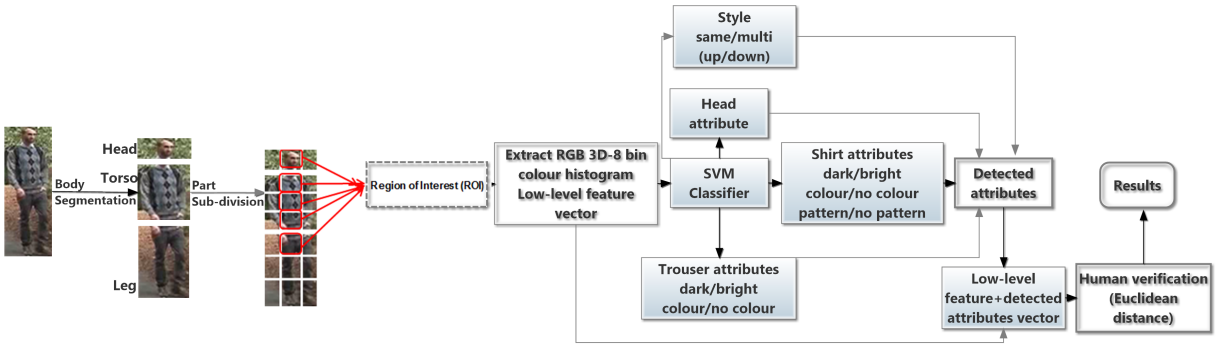


Figure 7.2: Proposed system for person re-identification

Fundamentally, in the proposed system, the re-identification of a person is carried out by jointly making use of so-called low-level features of a person's appearance (i.e. a detailed colour histogram of central body part regions, see section 7.3.1) and so-called mid-level features captured from a person's head, torso and leg regions (e.g. dark head, coloured shirt, dark trouser etc). More specifically the low-level feature representation of person's appearance is defined by detailed colour histograms which are normalised and obtained in regions of an initial body part segmentation and a subsequent sub-division (see section 7.3.1); while the mid-level feature representation of a person's appearance is defined by a higher-level description of the same regions that determines for example whether the shirt/trouser is dark/coloured or not and head is dark or not etc. The details of these functional blocks can be described in the following section.

7.3 Human Body Part-based Feature Representation

Prior to the detection and analysis of a human body parts or segments for subsequent feature extraction, the full human body needs to be detected in a scene. For this purpose we utilised the object detection technique of [17] which uses HOG features for human localisation. Once the full body is identified as defined within a single rectangle, a body part segmentation and a subsequent sub-division is carried out. Finally the a detailed feature analysis is carried out (see section 7.4 and 7.5) within the above regions that is finally used for person re-identification (see section 7.4.3).

7.3.1 Body region segmentation and sub-division

Assuming a standing and upright human, body region segmentation and sub-division helps subsequent capture of specific features of a segmented human object. This segmentation is performed by splitting the rectangular region containing the complete human figure into three parts, namely; head, torso, and leg (see figure 7.2). Further sub division of these three regions into smaller regions of interest (ROIs) is done by further splitting the; head region into three horizontally separated, equally sized sub-regions, the torso and leg regions are divided into equally sized, 3×3 rectangular sub-regions, as depicted by figure 7.2. In order to minimise the effects of consideration of the background regions in further analysis, only the middle rectangular patch is selected from the head region and the four middle patches, placed vertically, are selected from the torso and the leg regions, for subsequent capture of low-level colour histogram features and further attribute selection.

7.3.2 Low-level feature extraction and representation

The next step after body regions segmentation and sub-division is the colour histogram based feature detection and representation of the five centrally spaced regions. For each of the five said regions a so-called RGB 3D-8 bin colour histogram is extracted by (see figure 7.2) dividing each colour channel (i.e. R,G and B) into 8 bins and concatenating into a single feature vector of length $8 \times 8 \times 8 = 512$. Consequently, the appearance of a person is described by a feature vector, obtained by concatenating features of the five centrally located patches; giving a total feature length of 2560 (see figure 7.3).

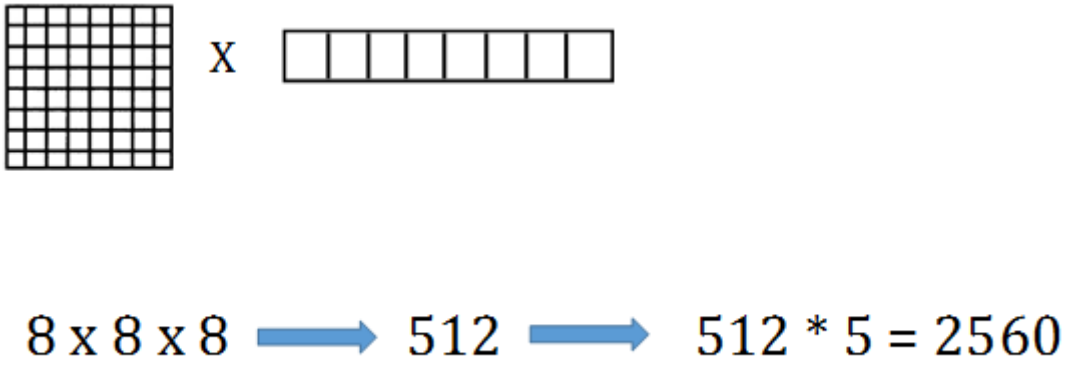


Figure 7.3: Low-level feature concatenation

7.4 Clothing Attribute Representation

Aimed at creating a more detailed representation of a human figure by adding further higher level features to the low-level feature descriptor obtained above (see section 7.3.2) the said five regions are further analysed to determine seven attributes that determines a higher-level appearance of the human body.

Figure 7.4 illustrates the seven attributes defined. One attribute is defined from the head-region, namely the 'head-colour'. Three attributes are defined from the shirt region, namely the 'shirt-colour', 'shirt-brightness' and 'shirt-pattern', Two attributes are defined from the trouser region, namely, 'trouser-colour' and 'trouser-brightness'. Finally, one attribute is defined for describing the overall appearance, namely, 'clothing-style'. Each of the above attributes can take two possible values as tabulated in table 7.1. Hence the value of each of the attributes can be represented by a binary number 1, or 0, for e.g. dark-shirt with 1 and non-dark shirt with 0.

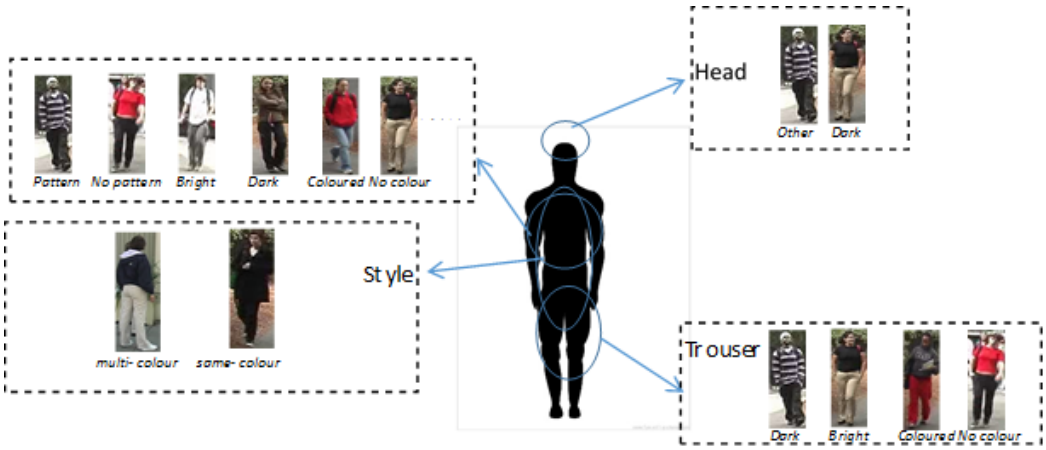


Figure 7.4: Definition of medium-level attributes

Table 7.1: Attributes description and values

Number	Attributes	Value1	Value2
1	Shirt-Colour	Coloured	No Colour
2	Shirt-Brightness	Bright	Dark
3	Shirt-Pattern	Patterned	No Pattern
4	Clothing-Style	Single colour up/down	Multi-colour up/down
5	Head-Colour	Dark	Other
6	Trouser-Brightness	Dark	Bright
7	Trouser-Colour	Coloured	No No colour

7.4.1 Clothing attribute value determination

The medium-level attribute values of test human objects were determined by using a Support Vector Machine (SVM) classifier to train on hand annotated attributes with known values from known sample regions of a training image dataset (see section 7.6.1).

As a result of the above each detected human figure's medium-level features will be represented by a seven element vector with each element being either a zero or a one.

7.4.2 The combined feature vector

Figure 7.5 illustrates the combined feature vector that comprises of the low-level 3D-8 bin colour histogram features and medium level features that are represented by the above mentioned attributes. This combined feature vector defines the detected human and will subsequently be used in human re-identification.

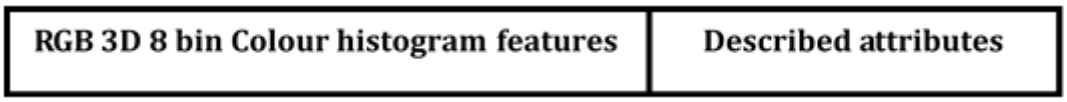


Figure 7.5: Total feature length

7.5 Experiments

Two datasets were used for experiments, a self-captured set of new content and the most popular database used by other researchers, i.e., VIPeR.

7.5.1 Self-captured dataset

The captured database has 118 frames which comprises of footage relevant to 6 different people taken from two different cameras. All images are scaled to a size of 128×48 pixels. In our experiments the cameras are named as A and B and the set of images captured by Cam B are used as the gallery images and the set of images captured by the Cam A are used as the probe image set. The performance of the proposed algorithm for person re-identification is evaluated by matching each test image in Cam A against the images in Cam B, the gallery image set. Figure 7.6 shows some examples of the detected persons in the self-constructed dataset. This dataset contains predominantly indoor images with challenges in illumination changes due to changes in artificial lighting within the building.

Figure 7.6: Samples from the self-captured data set

7.5.2 The VIPeR dataset

The VIPeR dataset contains 632 pedestrian image pairs captured by two cameras having different viewpoint, pose and lighting. Images are scaled to size 128×48 pixels. In our experiments we name the two camera as Cam A and Cam B. In the experiments conducted the set of images captured by the Cam B are considered the gallery set and those captured by the Cam A are considered as the probe image set. The algorithmic performance is evaluated by matching each test image in Cam A against the Cam B gallery.

Some selected example images from the VIPeR dataset are illustrated in figure 7.7

Figure 7.7: VIPeR data samples

7.5.3 Evaluation and metrics used

The database used for evaluation be it the VIPeR dataset or the self-captured dataset is first divided into two sets, i.e., the training image set and test image set. Approximately half of the images are used for training and the remaining half is used for testing. We train an SVM classifier on both the training and validation portions, while re-identification performance is reported on the held out test portion.

A person from the query image set is re-identified using a distance metric between itself and each of the candidate images in the gallery image set.

The low-level, distance measure, d^L, between a query image, I_q and a candidate image from the gallery image set I_g is defined as follows:

$$d^L(I_q, I_g) = \sum_l d_l^L(L_l(I_q), L_l(I_g)) \tag{7.1}$$

where $L_l(I_q)$ and $L_l(I_g)$, refers to the extracted type l low-level features from the query and gallery images i.e I_q and I_g respectively and d_l^L is the corresponding distance measure for the feature type l.

For the clothing attributes, the distance measure is defined as follows:

$$d^A(I_q, I_g) = \sum_a d_a^A(A_a(I_q), A_a(I_g)) \tag{7.2}$$

where $A_a(I_q)$ and $A_a(I_g)$ are the attribute encoding 'a' of the query image I_q and the candidate gallery image I_g

Given the above definitions, the Euclidean distance metric between a query image and a gallery image based on the low-level features is defined as follows:

$$d^L = \sqrt{\sum_i (q(l_i|x_{q,i}) - g(l_i|x_{g,i}))^2} \tag{7.3}$$

where $l_i|x_{q,i}$ refers to the i^th low-level feature of the query image given all other features of the query image and $l_i|x_{g,i}$ refers to the i^th low-level feature of the gallery image given all other features of the gallery image.

Similarly, the Euclidean distance metric between the query image and a gallery image based on the attribute-space is defined as follows:

$$d^A = \sqrt{\sum_i (q(a_i|x_{q,i}) - g(a_i|x_{g,i}))^2} \tag{7.4}$$

where all terms can be defined in a manner similar to that defined in equation 7.3.

In literature, the standard performance evaluation metrics used in person re-identification are matching performance at rank n, cumulative matching characteristic (CMC) curves, and normalised Area Under the CMC Curve (nAUC) [47]. The matching performance at rank n reports the probability that the correct match occurs within the first n ranked results from the gallery image set. This is obtained by calculating the Euclidean distances between a query image and all images in the gallery image set and ordering the matches in ascending order of matching error. The match with the smallest error is considered the rank-1 image and so on. The CMC curve plots the recognition for all rank values, n, and the nAUC summarises the area under the CMC curve (Note: the ideal nAUC is 1.0 and nAUC of 0.5 defines match obtains simply by 'chance').

However, the measures used for the performance evaluation of the proposed person re-identification algorithm are limited to the rank score illustrated by the associated cumulative matching characteristic (CMC) curves.

7.6 Experimental Results and Analysis

This section presents the experimental results and a detailed analysis. The performance of the proposed approach was considered using three different matching metric measures namely, a) matching based on low-level features only b) matching based on medium-level attribute signatures only and c) matching based on both low level features and attributes, combined.

7.6.1 Attributes detection

After the extraction of low-level colour features they can be used in the colour based recognition of values of the seven attributes of a human figure defined in Table 7.1. The VIPeR database was used for the attribute training and testing. From the images captured for Camera A, each attribute value was manually annotated. The manually annotated information from Camera A, for a given attribute (say for e.g. shirt-colour) was used in training an SVM. The testing was done on images captured by Camera B. Each attributes value was determined using the relevant trained SVM. This training and testing processes were carried out for each attribute, separately, using a different SVM. Table 7.2 records the detection accuracies obtained for each of the attributes. The highest accuracy has been obtained for 'Style' and the lowest accuracy has been recorded for the Head region in deterring whether it is dark or not. The latter is due to the high possibility of presence of individuals with darker skin tone and these individuals getting mixed up with people who are turning the back of their head to the camera.

Table 7.2: Attributes classification accuracies based on VIPeR dataset

Number	**Attributes**	**Value1**	**Value2**	**Detection accuracy**
1	Shirt-Colour	Coloured	No Colour	79.4%
2	Shirt-Brightness	Bright	Dark	73.4%
3	Shirt-Pattern	Patterned	No Pattern	87.8%
4	Clothing-Style	Single colour up/down	Multi-colour up/down	90.7%
5	Head-Colour	Dark	Other	66.5%
6	Trouser-Brightness	Dark	Bright	70.9%
7	Trouser-Colour	Coloured	No No colour	76.4%
	Mean			**77.9%**

The average accuracy for the detected attributes is 77.9%.

7.6.2 Matching performance analysis

Figure 7.8 illustrates the CMC curves when low-level features and attributes are used for the representation of detected people, both as individual metrics and together, i.e. as a combined metric. When the combined feature set is used the figure 7.9 illustrates the same graphs plotted within the narrow range of Rank-1 to Rank-20.

The results indicate that up to Rank-5 the combined feature set performs better than the individual feature sets. However above Rank-5 a better accuracy of recognition is demonstrated when using the Attributes only. This indicates that the detailed low level colour histogram features add details to the person's Attributes making the matching more accurate at up to Rank-5. However the use of low-level colour features only is not recommended due to relatively poor performance. A detailed study revealed that the low-level colour features although providing details for higher ranked matches, when used independently varies significantly between images of even the same person. Having the Attributes considered in addition allows the combined features to more accurately define an object.

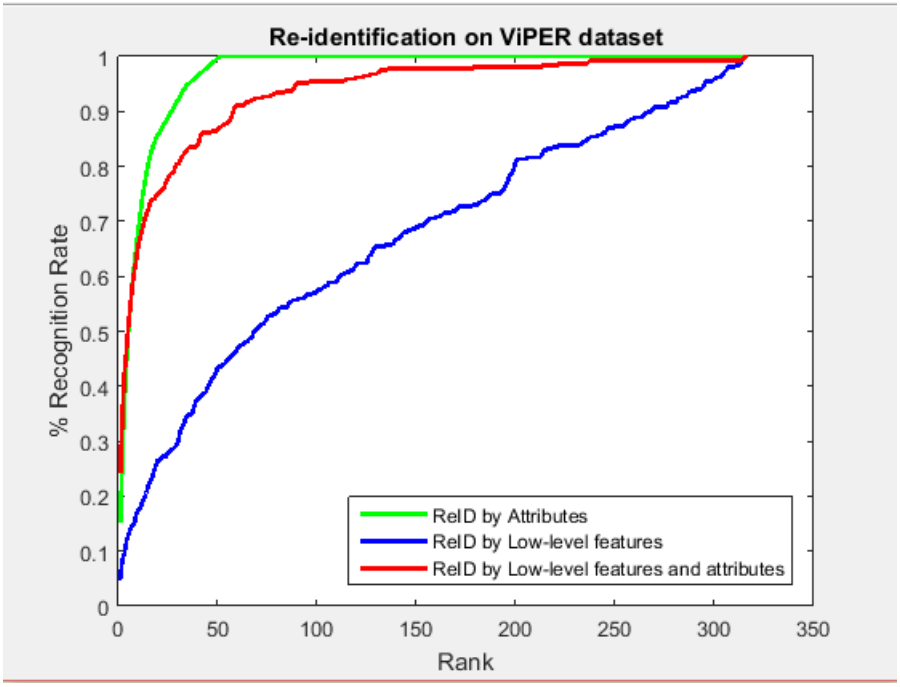


Figure 7.8: Cumulative matching characteristic curves of proposed technique

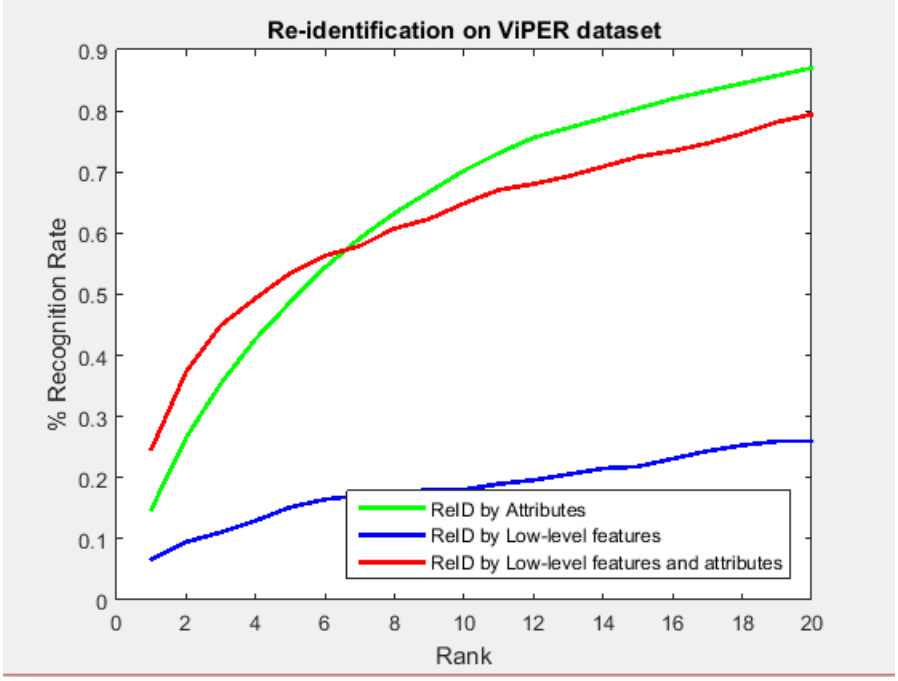


Figure 7.9: Cumulative matching characteristic curves of proposed technique plotted within the narrow range of Rank-1 to Rank-20

The average accuracy obtained by averaging over all Rank's was 62%, 97.6% and 92.1% respectively for low-level features, attributes and their combination.

Table 7.3 compared the performance of the proposed approach to that of the method proposed in [47] that proposed a low-level feature based approach dependent on colour and textures for initial attribute detection and an subsequent

attribute only based approach for person re-identification. The results have been tabulated for the same set of training and test images obtained from the VIPeR image database.

The results tabulated in Table 7.3 show that the at Rank-5 and above the proposed approach when only the Attributes are used and the combined set of Attributes and low-level features are used performed significantly better than the method proposed in [47] a method popularly used as a benchmarking algorithm in literature. However at Rank-1 the proposed method when only the Attributes are used performs less accurately as compared to the benchmark algorithms. It is noted that the benchmark algorithm of [47] is based on a larger (hence more detailed set of medium-level features) set of attributes (21 attributes) as compared to the number of attributes used by the proposed technique (7 attributes). This is the likely reason for it to perform better than the proposed algorithm at Rank-1 when only the Attributes are use. However when the combined low-level colour features and medium-level Attributes are used the proposed algorithm works better. This is due to the additional detail of the objects definition included by the low-level colour attributes that are used in the proposed approach.

The proposed low-level feature set only includes colour features from the RGB representation of the image. However the low-level feature set that the algorithm in [47] uses for attribute detection uses both colour features and texture features. The colour features, show less in number is spread across three different representations of object colour (RGB, HS and YCbCr). Our detailed investigation revealed that when colour features of the same object when represented in different colour features are used, a significant amount of redundant information is used in the training process. This affects the accuracy. Further global texture features are very much subjected to changes due to background clutter, over/under exposed images etc, that could also affect in a negative manner if texture features are also used alongside colour features.

Table 7.3: Person re-identification accuracy

VIPeR	Rank1	Rank5	Rank10	Rank20
Attributes	15.5	50	68.4	85.8
Low-level features and attributes	24.7	54.4	65.5	75
Low-level features	5.1	13	17.4	26.6
Method in [47]	21.4	41.5	55.2	71.5
Self-constructed				
Proposed technique	5	35.6	56	74.6

Figure 7.10 illustrates bar graphs comparing the performance at different Rank

scores. Results in Table 7.3 also tabulates the performance of the proposed approach when combined features are used and the self-captured dataset with more challenging images are used for experimentation.

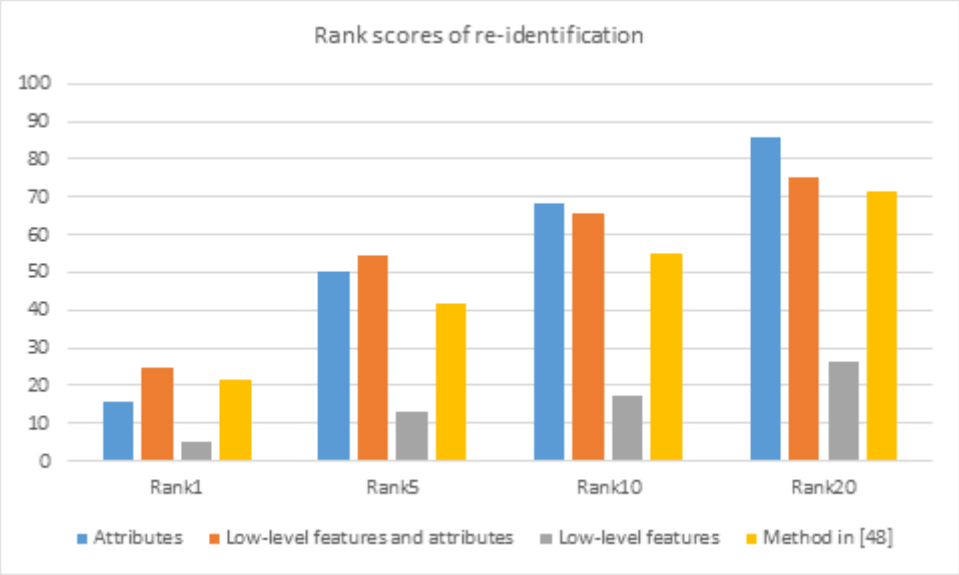


Figure 7.10: Rank scores re-identification performance

Figure 7.11 illustrates an example of matching gallery images for a probe image from the query image dataset when using the VIPeR dataset (top row) and the self-captured (bottom row) dataset. It is seen that the query image matches with a number of candidates from the gallery image database where the person has turned with respect to the camera angle of view.

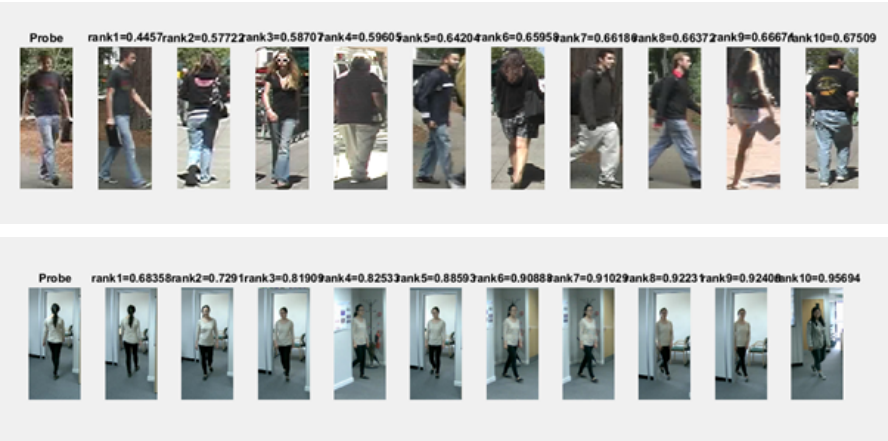


Figure 7.11: Human re-identification on both datasets

The above results indicate the superior performance obtainable from the proposed approach.

7.7 Conclusion

In this chapter, we have shown that detailed colour featured captured in known localities of a human figure in the form of a 3D colour histogram with a finite number of bins can be used to accurately determine attributes of a human body that can then be used together with the low-level colour features for person re-identification. Accuracy figures of approximately 75% and 85% have been obtained when using combined Attribute and low-level features and Attributes only, respectively at a rank of Rank-20.

Chapter 8

Conclusions and Future Work

This thesis presented a number of novel approaches to the general areas of object detection, recognition and people re-identification.

The first original contribution made within the context of the research presented in this thesis was a novel approach to military personnel recognition i.e. a system that helps count the number of military personnel present within an environment so as to enable the counting of personnel. Such a system can be used for alerting for missing personnel or reporting an increase of the number personnel within a secured premises enabling the generation of early warnings. Apart from using the camouflage type of the personnel's uniform we proposed the use of the badge on the cap to recognise the arm of service of some personnel.

The second original contribution made within the context of the research presented in this thesis was a vehicle type recognition algorithm. A unique feature of this novel approach was it ability to recognise the vehicle type irrespective of the angle of view of the camera. This contribution alleviates the a major challenge associated with requiring re-calibration of camera and re-training on captured data, should the camera change its direction of view.

The third contribution of this thesis focused on night-time people detection and vehicle type recognition using thermal image sets. The proposed technique based on a novel featured named as CENTROG is a variant of the well-known CENTRIST approach in which HOG features are extracted from a CT transformed image.

The fourth and final contribution of the research presented within the context of this thesis is an algorithm for person re-identification. In the approach proposed the person re-identification was done based on detected medium-level Attributes of a person combined with a low-level, detailed, colour histogram feature vector. The Attributes themselves were first determined based on using known information from detailed colour histograms from corresponding Attributes of a training database of images. Both the use of colour histogram details only and Attribute

details only and then using the combined feature set were investigated.

The following section presents the conclusion reached.

8.1 Conclusion

Two techniques were presented and implemented [58, 68] for military personnel recognition in chapter 4 of this thesis. However, both techniques were challenged in their ability to categorise military personnel based on the appearance of their uniforms due to the presence of similarities of colour and texture of a camouflage of a personnel's uniform. Particularly, it was observed that an army camouflage appearance pattern is similar to that of the air force camouflage appearance pattern while the army and navy uniforms have similar colour contents. Colour and texture features alone were unable to categorise between classes based on experimental results obtained. We explored colour and GLCM texture features to effectively differentiate between military persons' class of arm of service. It was shown that colour plays a vital role in camouflage person recognition especially when the camouflage appearance between classes have colour variations. We also showed the importance of selecting discriminative features using feature selection in particular the use of the CFS algorithm, so as to speed up processing and improve recognition accuracy. In the same work, we showed that the cap badge of a military person can categorise between classes using matching of SURF features. With the proposed system, any military personnel on AWOL can be detected, an alert can be signalled for a check of suspicious persons within an environment. Therefore the proposed military personnel arm of service recognition system can complement any existing face recognition based security technology by integrating the two system. The proposed system was simple algorithmically and fast and can be implemented for a real-time military monitoring system.

A novel technique for vehicle type recognition irrespective of angle of view was proposed in chapter 5. The integration of region, HOG and LBP features showed that single individual feature for vehicle type recognition cannot adequately categorise vehicle types in different view related scenarios. To demonstrate the performance of the proposed algorithm, data were combined from datasets obtained from two datasets of different views (front/rear and angular views) using the proposed feature combination approach. An overall average recognition accuracy of 95% was recorded in combined view datasets, which means a vehicle can be recognised irrespective of direction of movement or view with the need for only a single initial training requirement. We also showed the importance of selecting discriminative features using CFS, so as to speed up processing and improve recognition accuracy. With this system, the bottlenecks associated with the need for

re-calibration and re-training were eliminated, since, only a once for all training is required for vehicle type classification in any direction or angle of view. This system will provide assistance in any toll collection facility and in situations where there is the need to keep count of a particular vehicle movement in a location at a particular time. For example, in determining vehicle air pollution, this system can help give information about the proportions of each vehicle type that pollutes a particular environment.

In chapter 6, the use of CENTROG features was proposed for the detection of pedestrians and recognition of vehicles at night-time when using thermal images. We proposed a feature set that can both detect people and recognise vehicles at night-time. This approach is useful in a driver assistance system, in which a single feature set can be processed and used for pedestrian detection as well as for vehicle type recognition. We compared the use of the proposed CENTROG features against the use of known CENTRIST features for the recognition of people and vehicle in thermal images datasets. Results obtained showed that CENTROG can effectively detect pedestrians and recognise vehicles at night-time and performs significantly better than the known CENTRIST feature based approaches. We also showed the importance of selecting discriminative features using CFS, so as to speed up processing and improve recognition accuracy. This system is implementable in real-time and can serve especially in mission-critical situations or a situation wherein human safety cannot be compromised.

In chapter 7, we proposed the use of low-level RGB colour features and medium-level Attribute features obtained based on the said colour features, both individually and in a combined format for person re-identification. The features and Attributes were obtained from a selected set of regions highly likely to be a parts of a human body, from within a rectangle enclosing the captured whole human body. This specific selection of regions was done so that it is possible to reduce the impact of background clutter that may severely affect recognition performance. The use of seven clothing attributes was proposed for person re-identification along with the use of low-level colour features. The described Attributes were detected using the low-level colour features captured initially and making use of an SVM classifier, giving an overall accuracy of 77.9%. It was shown that combining the low-level colour features with the Attributes described above helps increase recognition accuracy giving a rank-5 accuracy of approximately 54% as against the reported rank-5 accuracy of the 42% by [47]. The proposed system can help improve human tracking performance; track a particular person in a shopping mall; track a particular person in non-overlapping camera in a military or school environment etc.

8.2 Future Work

Although the novel ideas presented in this thesis advances the current state of art and technology in a number of areas related to the application of computer vision and pattern recognition technologies in the application areas of security and surveillance there are further opportunities for improvements and extensions of the proposed algorithms and systems. There is also the possibility of integrating the proposed technology with other vision technologies to enhance overall system performance.

To address the problem of civilian presence in a military environment, the proposed military personnel recognition system can be further extended to enable the recognition of non-military personnel. To this effect a new category/type that is worthy of inclusion is the civilian category. Similarly, the proposed uniform/clothing appearance based personnel recognition algorithm can be integrated with existing face recognition technologies to provide a security system which can be of more significant practical use. Such a system can be used to for example track the whereabouts of a particular known individual military officer within the video/images captured by a distributed camera system to ensure his/her safety and thus contributing towards the general safety of a campsite.

For the day-time vehicle type recognition algorithm proposed in chapter 5, vehicles were identified on a frame-by-frame basis. It is however possible to extend this work so that, vehicles are not recognised on a frame-by-frame basis but rather based on an entire tracked vehicle object. This would allow for the opportunity to further increase the robustness and the accuracy of the proposed system based on assigning a majority voted outcome and/or a position dependent, weighted outcome for the vehicle type. It is possible to test the proposed system under different environmental conditions such as rainy and windy weather; in different geopolitical zones such as vehicles in Africa, Europe, Asia, Australia and America. More extensive testing to evaluate the performance of the algorithm under non-ideal illumination situations could also have been conducted but was not possible due to the lack of test video footage and restricted access to resources.

In Chapter-6, two approaches were proposed for pedestrian detection and vehicle type recognition at night-time. It is possible to integrate both systems to remove if any, bottlenecks, associated with the individual use of algorithms to solve the challenges of a similar practical nature. The proposed system for vehicle type recognition at nigh-time can be further extended to accommodate more vehicular classes, so that we can generalise the applicability of the proposed algorithm on practical vehicle type recognition tasks carried out at night-time. The implementation of the integrated system in real-time so as to demonstrate

how the system can assist in a driver assistance system on-board a vehicle is a further possible practical application or enhancement.

In literature the performance of people re-identification systems have always been demonstrated and evaluated on still images. The possibility of implementing the proposed technique within a real-time video analytic scenario so as to demonstrate the applicability of this system in a real world system, is proposed as future work. It was also revealed that the mid-level attributes detection performance could benefit from some performance improvements. Investigation of the use of additional features, the use of more effective feature reduction techniques and feature combinations are recommended. Further investigating the use of effective feature weighting, based on training data in obtaining the combined feature vector for representing an human object is also recommended.

References

[1] Cong Bai, Wenbin Zou, Kidiyo Kpalma, and Joseph Ronsin. Efficient colour texture image retrieval by combination of colour and texture features in wavelet domain. *Electronics letters*, 48(23):1463–1465, 2012.

[2] Slawomir Bak, Etienne Corvee, Francois Bremond, and Monique Thonnat. Person re-identification using spatial covariance regions of human body parts, 2010.

[3] Slawomir Bak, Etienne Corvee, Francois Bremond, and Monique Thonnat. Multiple-shot human re-identification by mean riemannian covariance grid, 2011.

[4] Sugata Banerji, Atreyee Sinha, and Chengjun Liu. New image descriptors based on color, texture, shape, and wavelets for object and scene image classification. *Neurocomputing*, (0), 2013. http://www.sciencedirect.com/science/article/pii/S0925231213001987.

[5] Federica Battisti, Marco Carli, Giovanna Farinella, and Alessandro Neri. Target re-identification in low-quality camera networks. In *IS&T/SPIE Electronic Imaging*, pages 865502–865502. International Society for Optics and Photonics, 2013.

[6] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. Surf: Speeded up robust features. In *Computer vision–ECCV 2006*, pages 404–417. Springer, 2006.

[7] Yannick Benezeth, Bruno Emile, Hélène Laurent, and Christophe Rosenberger. A real time human detection system based on far infrared vision. In *Image and Signal Processing*, pages 76–84. Springer, 2008.

[8] Peng Bian, Yi Jin, and Nai-ren Zhang. Fuzzy c-means clustering based digital camouflage pattern design and its evaluation. In *Signal Processing (ICSP), 2010 IEEE 10th International Conference on*, pages 1017–1020. IEEE, 2010.

[9] Henri Bouma, Sander Borsboom, Richard J. M. den Hollander, Sander H. Landsmeer, and Marcel Worring. Re-identification of persons in multi-camera surveillance under varying viewpoints and illumination. pages 83590Q–83590Q, 2012. 10.1117/12.918576.

[10] Thirimachos Bourlai, John Von Dollen, Nikolaos Mavridis, and Christopher Kolanko. Evaluating the efficiency of a night-time, middle-range infrared sensor for applications in human detection and recognition. In *SPIE Defense, Security, and Sensing*, pages 83551B–83551B. International Society for Optics and Photonics, 2012.

[11] Raluca Brehar. Pattern recognition system lab 5 - histograms of oriented gradients, retrieved 15th June, 2014 last updated 30th October 2013. http://users.utcluj.ro/~raluca/prs/prs_lab_05e.pdf.

[12] Shaogang Gong Tao Xiang Bryan Prosser, Wei-Shi Zheng. Person re-identification by support vector ranking, 2010.

[13] Shyang-Lih Chang, Fu-Tzu Yang, Wen-Po Wu, Yu-An Cho, and Sei-Wang Chen. Nighttime pedestrian detection using thermal imaging based on hog feature. In *System Science and Engineering (ICSSE), 2011 International Conference on*, pages 694–698. IEEE, 2011.

[14] David A Clausi. An analysis of co-occurrence texture statistics as a function of grey level quantization. *Canadian Journal of remote sensing*, 28(1):45–62, 2002.

[15] Wikipedia Contributors. Computer vision, 2013. http://en.wikipedia.org/wiki/Computer_Vision.

[16] PM Daigavane and PR Bajaj. Real time vehicle detection and counting method for unsupervised traffic video on highways. *International Journal of Computer Science and Network Security*, 10(8), 2010.

[17] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893. IEEE, 2005.

[18] Angela D'Angelo and Jean-Luc Dugelay. People re-identification in camera networks based on probabilistic color histograms. In *IS&T/SPIE Electronic Imaging*, pages 78820K–78820K. International Society for Optics and Photonics, 2011.

[19] James W Davis and Mark A Keck. A two-stage template approach to person detection in thermal imagery. *WACV/MOTION*, 5:364–369, 2005.

[20] Icaro Oliveira de Oliveira and Jose Luiz de Souza Pio. People reidentification in a distributed camera network, 2009.

[21] Yongsheng Dong and Jinwen Ma. Feature extraction through contourlet subband clustering for texture classification. *Neurocomputing*, 116:157–164, 2013.

[22] Francois Bremond Etienne Corvee, Slawomir Bak. People detection and re-identification for multi surveillance cameras, 2012.

[23] Michela Farenzena, Loris Bazzani, Alessandro Perina, Vittorio Murino, and Marco Cristani. Person re-identification by symmetry-driven accumulation of local features. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 2360–2367. IEEE, 2010.

[24] Tom Fawcett. An introduction to roc analysis. *Pattern recognition letters*, 27(8):861–874, 2006.

[25] P. F. Felzenszwalb, R. B. Girshick, and D. McAllester. Discriminatively trained deformable part models, release 4. http://people.cs.uchicago.edu/ pff/latent-release4/.

[26] C Fosu, GW Hein, and B Eissfeller. Determination of centroid of ccd star images. *Int. Arch. Photogram. Remote Sens. Spatial Inform. Sci*, 35:612–617, 2004.

[27] Keren Fu, Chen Gong, Yu Qiao, Jie Yang, and Irene Yu-Hua Gu. One-class support vector machine-assisted robust tracking. *Journal of Electronic Imaging*, 22(2):023002–023002, 2013.

[28] David Gerónimo and Hedvig Kjellstrom. Unsupervised surveillance video retrieval based on human action and appearance. In *Pattern Recognition (ICPR), 2014 22nd International Conference on*, pages 4630–4635. IEEE, 2014.

[29] Stefano Ghidoni, Grzegorz Cielniak, and Emanuele Menegatti. Texture-based crowd detection and localisation. In *Intelligent Autonomous Systems 12*, pages 725–736. Springer, 2013.

[30] Shaogang Gong, Marco Cristani, Chen Change Loy, and Timothy M Hospedales. The re-identification challenge. In *Person Re-Identification*, pages 1–20. Springer, 2014.

[31] Rafael C. Gonzalez and Richard E. Woods. *Digital image processing*. Pearson Prentice Hall, Dorling Kindersley, third edition, 2008.

[32] Rafael C Gonzalez, Richard E Woods, and Steven L Eddins. *Digital image processing using MATLAB*, volume 2 of *Digital image processing using MATLAB, second edition*. Tata McGraw Hill, Haryana, second edition, 2009.

[33] Bill Green. Canny edge detection tutorial, retrieved 15th June, 2015 (2002). http://dasl.mem.drexel.edu/alumni/bGreen/www.pages.drexel.edu/_weg22/can_tut.html.

[34] S. Gupte, O. Masoud, R.F.K. Martin, and N.P. Papanikolopoulos. Detection and classification of vehicles. *Intelligent Transportation Systems, IEEE Transactions on*, 3(1):37–47, 2002.

[35] Mark A Hall. *Correlation-based feature selection for machine learning*. PhD thesis, The University of Waikato, 1999.

[36] Mark A Hall and Lloyd A Smith. Feature selection for machine learning: Comparing a correlation-based filter approach to the wrapper. In *FLAIRS Conference*, pages 235–239, 1998.

[37] Omar Hamdoun, Fabien Moutarde, Bogdan Stanciulescu, and Bruno Steux. Person re-identification in multi-camera system by signature based on interest point descriptors collected on short video sequences, 2008.

[38] R. M. Haralick, K. Shanmugam, and Its'Hak Dinstein. Textural features for image classification. *Systems, Man and Cybernetics, IEEE Transactions on*, SMC-3(6):610–621, 1973.

[39] Sebastian Hommel, Darius Malysiak, and Uwe Handmann. Efficient people re-identification based on models of human clothes. In *Computational Intelligence and Informatics (CINTI), 2014 IEEE 15th International Symposium on*, pages 137–142. IEEE, 2014.

[40] Di Huang, Caifeng Shan, Mohsen Ardabilian, Yunhong Wang, and Liming Chen. Local binary patterns and its application to facial image analysis: a survey. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 41(6):765–781, 2011.

[41] National Computation Infrastructure. Image processing toolbox, Accessed 9 June 2014. https://nf.nci.org.au/facilities/software/Matlab/toolbox/images/regionprops.html.

[42] Interpol. Vehicle crime, Accessed 30 April 2014. http://www.interpol.int/Crime-areas/Vehicle-crime/Vehicle-crime.

[43] Yoichiro Iwasaki, Masato Misumi, and Toshiyuki Nakamiya. Robust vehicle detection under various environmental conditions using an infrared thermal camera and its application to road traffic flow monitoring. *Sensors*, 13(6):7756–7773, 2013.

[44] Thornton Jason, Baran-Gale Jeanette, Butler Daniel, Chan Michael, and Zwahlen Heather. Person attribute search for large-area video surveillance, 2011.

[45] Mehran Kafai and Bir Bhanu. Dynamic bayesian networks for vehicle classification in video. *Industrial Informatics, IEEE Transactions on*, 8(1):100–109, 2012.

[46] Kevin Krucki, Vijayan Asari, Christoph Borel-Donohue, and David Bunker. Human re-identification in multi-camera systems. In *Applied Imagery Pattern Recognition Workshop (AIPR), 2014 IEEE*, pages 1–7. IEEE, 2014.

[47] Ryan Layne, Timothy M Hospedales, and Shaogang Gong. Attributes-based re-identification. In *Person Re-Identification*, pages 93–117. Springer, 2014.

[48] Prof. Judah Levine. Hue, saturation and intensity, retrieved 5th September, 2013 2001. http://www.colorado.edu/physics/phys1230/phys1230_fa01/topic45.html.

[49] Aoxue Li, Luoqi Liu, Kangping Wang, Siyuan Liu, and Shuo Yan. Clothing attributes assisted person re-identification. 2014.

[50] Guoliang Li, Yong Zhao, Daimeng Wei, and Ruzhong Cheng. Nighttime pedestrian detection using local oriented shape context descriptor. In *Proceedings of the 2nd International Conference on Computer Science and Electronics Engineering*. Atlantis Press, 2013.

[51] Ye Li, Bo Li, Bin Tian, and Qingming Yao. Vehicle detection based on the and-or graph for congested traffic conditions. *Intelligent Transportation Systems, IEEE Transactions on*, 14(2):984–993, 2013.

[52] Zhengrong Li, Yuee Liu, Ross Hayward, and Rodney Walker. Color and texture feature fusion using kernel pca with application to object-based vegetation species classification. In *Image Processing (ICIP), 2010 17th IEEE International Conference on*, pages 2701–2704. IEEE, 2010.

[53] Wen-Hung Liao and Ting-Jung Young. Texture classification using uniform extended local ternary patterns. In *Multimedia (ISM), 2010 IEEE International Symposium on*, pages 191–195. IEEE, 2010.

[54] Song Liming and Geng Weidong. A new camouflage texture evaluation method based on wssim and nature image features. In *Multimedia Technology (ICMT), 2010 International Conference on*, pages 1–4. IEEE, 2010.

[55] Yu-Chun Lin, Yi-Ming Chan, Luo-Chieh Chuang, Li-Chen Fu, Shih-Shinh Huang, Pei-Yung Hsiao, and Min-Fang Luo. Near-infrared based nighttime pedestrian detection by combining multiple features. In *Intelligent Transportation Systems (ITSC), 2011 14th International IEEE Conference on*, pages 1549–1554. IEEE, 2011.

[56] Chunxiao Liu, Shaogang Gong, ChenChange Loy, and Xinggang Lin. *Person Re-identification: What Features Are Important?*, volume 7583, book section 39, pages 391–401. Springer Berlin Heidelberg, 2012. http://dx.doi.org/10.1007/978-3-642-33863-2_39.

[57] Qiong Liu, Jiajun Zhuang, and Shufeng Kong. Detection of pedestrians for far-infrared automotive night vision systems using learning-based method and head validation. *Measurement Science and Technology*, 24(7):074022, 2013.

[58] Yanyun Lu, Khaled Boukharouba, Jacques Boonært, Anthony Fleury, and Stéphane Lecœuche. Application of an incremental svm algorithm for on-line human recognition from video surveillance using texture and color features. *Neurocomputing*, 126:132–140, 2014.

[59] Xiao Luo, Qingsheng Luo, Baoling Han, and Cai Gao. Special target recognition and location using differential image detection technology. In *Intelligent Human-Machine Systems and Cybernetics (IHMSC), 2010 2nd International Conference on*, volume 2, pages 116–119. IEEE, 2010.

[60] Ansuman Mahapatra, Tusar Kanti Mishra, Pankaj K Sa, and Banshidhar Majhi. Human recognition system for outdoor videos using hidden markov model. *AEU-International Journal of Electronics and Communications*, 68(3):227–236, 2014.

[61] Tetsu Matsukawa, Toshiya Okabe, and Yuuki Sato. Person re-identification via discriminative accumulation of local features. In *Pattern Recognition (ICPR), 2014 22nd International Conference on*, pages 3975–3980. IEEE, 2014.

[62] MediaCybernetics. Color models, retrieved 5th September, 2013 2005. http://support.mediacy.com/answers/showquestion.asp?faq=35&fldAuto=268.

[63] Sonka Milan, Hlavac Vaclav, and Boyle Roger. *Image Processing Analysis, and Machine Vision*. Cengage Learning, Delhi, third edition, 2008.

[64] Eric Miller. Colour models: Rgb, 20 September 2007. http://graphicdesign.about.com/od/colorbasics/a/rgb.htm.

[65] Ehsan Adeli Mosabbeb, Maryam Sadeghi, and Mahmoud Fathy. A new approach for vehicle detection in congested traffic scenes based on strong shadow segmentation. In *Advances in Visual Computing*, pages 427–436. Springer, 2007.

[66] Yadong Mu, Shuicheng Yan, Yi Liu, Thomas Huang, and Bingfeng Zhou. Discriminative local binary patterns for human detection in personal album. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008.

[67] Ana Cris Murillo, José Jesús Guerrero, and C Sagues. Surf features for efficient robot localization with omnidirectional images. In *Robotics and Automation, 2007 IEEE International Conference on*, pages 3901–3907. IEEE, 2007.

[68] Chikahito Nakajima, Massimiliano Pontil, Bernd Heisele, and Tomaso Poggio. Full-body person recognition system. *Pattern recognition*, 36(9):1997–2006, 2003.

[69] Loris Nanni, Sheryl Brahnam, and Alessandra Lumini. Texture descriptors for generic pattern classification problems. *Expert Systems with Applications*, 38(8):9340–9345, 2011. http://www.sciencedirect.com/science/article/pii/S0957417411001436.

[70] Jun Yee Ng and Yong Haur Tay. Image-based vehicle classification system. *arXiv preprint arXiv:1204.2114*, 2012.

[71] Documentation OpenCV. Background subtraction, Accessed 23rd January 2014. http://docs.opencv.org/trunk/doc/py_tutorials/py_video/py_bg_subtraction/py_bg_subtraction.html.

[72] Celil Ozkurt and Fatih Camci. Automatic traffic density estimation and vehicle classification for traffic surveillance systems using neural networks. *Mathematical and Computational Applications*, 14(3):187, 2009.

[73] J.R. Parker. *Algorithms for Image Processing and Computer Vision*. Wiley Publishing, inc, Indianapolis, second edition, 2011.

[74] Shwetmala Ramachandra T.V. Emissions from indias transport sector: Statewise synthesis. *Atmospheric Environment*, 1(8), 2009.

[75] H.T.P. Ranga, M. Ravi Kiran, S. Raja Shekar, and S.K. Naveen Kumar. Vehicle detection and classification based on morphological technique. In *Signal and Image Processing (ICSIP), 2010 International Conference on*, pages 45–48, 2010.

[76] Douglas Reynolds. Gaussian mixture models. *Encyclopedia of Biometrics*, pages 659–663, 2009.

[77] Irfan Riaz, Jingchun Piao, and Hyunchul Shin. Human detection by using centrist features for thermal images. In *International Conference Computer Graphics, Visualization, Computer Vision and Image Processing*. Citeseer, 2013.

[78] Andrea Cavallaro Riccardo Mazzon, Syed Fahad Tahir. Person re-identification in crowd. *Pattern Recognition Letter*, 2012.

[79] DA Roark, H Abdi, et al. Human recognition of familiar and unfamiliar people in naturalistic video. In *Analysis and Modeling of Faces and Gestures, 2003. AMFG 2003. IEEE International Workshop on*, pages 36–41. IEEE, 2003.

[80] Carsten Rother, Vladimir Kolmogorov, and Andrew Blake. "grabcut": interactive foreground extraction using iterated graph cuts. *ACM Trans. Graph.*, 23(3):309–314, 2004.

[81] Shaogang Gong Ryan Layne, Timothy Hospedales. Person re-identification by attributes, 2012.

[82] Amit Satpathy, Xudong Jiang, and How-Lung Eng. Lbp-based edge-texture features for object recognition. *Image Processing, IEEE Transactions on*, 23(5):1953–1964, 2014.

[83] Riccardo Satta, Giorgio Fumera, Fabio Roli, Marco Cristani, and Vittorio Murino. A multiple component matching framework for person re-identification, 2011.

[84] William Robson Schwartz and Larry S Davis. Learning discriminative appearance-based models using partial least squares. In *Computer Graphics*

and Image Processing (SIBGRAPI), 2009 XXII Brazilian Symposium on, pages 322–329. IEEE, 2009.

[85] P Sengottuvelan, Amitabh Wahi, and A Shanmugam. Performance of decamouflaging through exploratory image analysis. In *Emerging Trends in Engineering and Technology, 2008. ICETET'08. First International Conference on*, pages 6–10. IEEE, 2008.

[86] A. Suresh Shunmuganathan and K.L. Feature fusion technique for colour texture classification system based on gray-level co-occurrence matrix. *Journal of Computer Science*, 8(12):2106–2111, 2012.

[87] Clemens Siebler, Keni Bernardin, and Rainer Stiefelhagen. Adaptive color transformation for person re-identification in camera networks, 2010.

[88] Francois Bremond Monique Thonnat Slawomir Bak, Etienne Corvee. Person re-identification using haar-based and dcd-based signature, 2010.

[89] Marvin Smith, Joshua Gleason, Steve Wood, and Issa Beekun. Vehicle location by thermal images features, Accessed 5th November 2014. http://cs426team11.github.io/vltif/.

[90] L. K. Soh and C. Tsatsoulis. Texture analysis of sar sea ice imagery using gray level co-occurrence matrices. *Geoscience and Remote Sensing, IEEE Transactions on*, 37(2):780–795, 1999.

[91] Guilan Song and Shunqing Tang. Method for spectral pattern recognition of color camouflage. *Optical Engineering*, 36(6):1779–1781, 1997. 10.1117/1.601322.

[92] Chris Stauffer and W Eric L Grimson. Adaptive background mixture models for real-time tracking. In *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, volume 2. IEEE, 1999.

[93] Boris Takac, Andreu Catala, Matthias Rauterberg, and Wei Chen. People identification for domestic non-overlapping rgb-d camera networks. In *Multi-Conference on Systems, Signals & Devices (SSD), 2014 11th International*, pages 1–6. IEEE, 2014.

[94] N.A. Thacker and P.A. Bromiley. Tina 5.0 user's guide, Accessed 5th November 2013 2005 - Accessed 5th November 2013. http://www.tina-vision.net/manuals/user_guide/node119.html.

[95] MD Thomas G. Tape. Interpreting diagnostic tests, retrieved 15th June, 2014. http://gim.unmc.edu/dxtests/roc3.htm.

[96] Avinash Uppuluri. Glcm, retrieved 15th May, 2013 last updated 5th Apr 2010. http://www.mathworks.com/matlabcentral/fileexchange/22354-glcmfeatures4-m-vectorized-version-of-glcmfeatures1-m-with-code-char

[97] Paul Viola and Michael J Jones. Robust real-time face detection. *International journal of computer vision*, 57(2):137–154, 2004.

[98] Chi-Chen Raxle Wang and J.-J.J. Lien. Automatic vehicle detection using local features;a statistical approach. *Intelligent Transportation Systems, IEEE Transactions on*, 9(1):83–96, 2008.

[99] Weihong Wang, Jian Zhang, and Chunhua Shen. Improved human detection and classification in thermal images. In *Image Processing (ICIP), 2010 17th IEEE International Conference on*, pages 2313–2316. IEEE, 2010.

[100] Xiaoyu Wang, Tony X Han, and Shuicheng Yan. An hog-lbp human detector with partial occlusion handling. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 32–39. IEEE, 2009.

[101] Contributors Wikipedia. Surf, Accessed 9 June 2014. http://en.wikipedia.org/wiki/SURF.

[102] Jianxin Wu, Christopher Geyer, and James M Rehg. Real-time human detection using contour cues. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 860–867. IEEE, 2011.

[103] Ziyan Wu, Yang Li, and Richard J Radke. Viewpoint invariant human re-identification in camera networks using pose priors and subject-discriminative features. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 37(5):1095–1108, 2015.

[104] Ming Yin, Hao Zhang, Huadong Meng, and Xiqin Wang. An hmm-based algorithm for vehicle detection in congested traffic situations. In *Intelligent Transportation Systems Conference, 2007. ITSC 2007. IEEE*, pages 736–741. IEEE, 2007.

[105] Lei Yu and Huan Liu. Feature selection for high-dimensional data: A fast correlation-based filter solution. In *ICML*, volume 3, pages 856–863, 2003.

[106] Ting-Jin Yun, Yong-Cai Guo, and Gao Chao. Human detection in far-infrared images based on histograms of maximal oriented energy map. In

Wavelet Analysis and Pattern Recognition, 2007. ICWAPR'07. International Conference on, volume 2, pages 933–938. IEEE, 2007.

[107] Ramin Zabih and John Woodfill. Non-parametric local transforms for computing visual correspondence. In *Computer VisionECCV'94*, pages 151–158. Springer, 1994.

[108] Xiaobin Zhu, Jing Liu, Jinqiao Wang, Yikai Fang, and Hanqing Lu. Anomaly detection in crowded scene via appearance and dynamics joint modeling. In *Image Processing (ICIP), 2012 19th IEEE International Conference on*, pages 2705–2708. IEEE, 2012.

[109] Zoran Zivkovic. Improved adaptive gaussian mixture model for background subtraction. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, volume 2, pages 28–31. IEEE, 2004.