

UNIVERSITY OF SOUTHAMPTON

Improving Motion Segmentation with Combined Classifiers

by

Ahmad H. Almazeed

A thesis submitted in partial fulfillment for the
degree of Doctor of Philosophy

in the
Faculty of Engineering, Science and Mathematics
School of Electronics and Computer Science

March 2006

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING, SCIENCE AND MATHEMATICS
SCHOOL OF ELECTRONICS AND COMPUTER SCIENCE

Doctor of Philosophy

by Ahmad H. Almazeed

Foreground/background segmentation is an active research area for moving object analysis. Many applications in machine vision depend on high quality and robust extraction of moving objects. Established and popular methods are mixture modelling and a threshold based technique (Horprasert et al., 2000). To find a better motion classifier, a new technique is developed here, a modified Unary classifier approach that uses the bases of SVM theory. As neither the mixture modelling nor the Unary approach had implicit shadow detection, this is achieved by including colour invariant colour models. The threshold based technique has the ability to detect shadow but with the consequences of mislabelling part of the foreground. The shadow detection criterion was improved by adding a statistical constraint to the shadow detection process. In order to further extend the performance, we formed different classifiers by combining base classifiers with a Bayesian approach. The observed performance advantages are associated with the fusion of operators with complementary properties. Tests on outdoor and indoor sequences confirm the efficacy of this approach. The new algorithms can successfully identify and remove shadows and highlights with improved moving-object segmentation. A particular advantage of our evaluation is that it is the first approach that compares foreground/background labelling with results obtained from labelling by broadcast techniques, comparing a computer vision technique with an established baseline.

Contents

Nomenclature	xii
Acknowledgements	xiii
1 Introduction	1
1.1 Motivation	1
1.2 Motion estimation	1
1.2.1 Background subtraction	2
1.2.2 Statistical methods	2
1.2.3 Temporal differencing	3
1.2.4 Stereo techniques	3
1.3 Applications	4
1.4 Contributions	5
1.5 Assumptions	6
1.6 Structure	7
2 Statistical Background Disturbance	9
2.1 Disturbance parameters	10
2.2 Automatic Thresholding	12
2.3 Experimental Results	13
2.3.1 Data and Assessment	13
2.3.2 SBD results	18
2.4 Improvements on shadow extraction process	24
2.5 Conclusions	27
3 Mixture of Gaussians	30
3.1 Introduction	30
3.2 Gaussian Density Function	31
3.3 Mixture Model	32
3.4 Stauffer mixture of Gaussians algorithm (MOG)	32
3.4.1 Parameter Settings	36
3.4.2 Experimental Results	44
3.5 Conclusions	49
4 Unary Classifiers	50
4.1 Introduction	50
4.2 Kernel Functions	50
4.3 Hypersphere Method	51

4.4	UC Results	53
4.5	Improved UC	59
4.6	Improved UC Results	60
4.7	Conclusions	67
5	Shadow Suppression using Invariant Colour Models	68
5.1	Motivation	68
5.2	Colour space	69
5.2.1	RGB	69
5.2.2	Normalised rgb	70
5.2.3	HSV	71
5.2.4	$l_1l_2l_3$	72
5.2.5	$c_1c_2c_3$	72
5.3	Colour Model Evaluation	73
5.3.1	Parameterising Unary Classifier	73
5.3.2	Indoor Evaluation	75
5.3.3	Outdoor Evaluation	79
5.4	MOG and the $c_1c_2c_3$ colour model	83
5.4.1	Extraction Results for the MOG using $c_1c_2c_3$ colour model	83
5.5	Conclusions	88
6	Combining Classifiers	90
6.1	Introduction	90
6.2	Bayesian Classification	92
6.3	Probability Estimation	92
6.4	Combination Rules	93
6.5	Weight averaging the classifiers decisions	94
6.6	Experimental results	95
6.6.1	Two Classifiers Combination	95
6.6.2	Three Classifiers Combination	104
6.7	Conclusions	125
7	Conclusions and Future Work	126
7.1	Conclusions	126
7.2	Future Work	128
A	Unary Classifier sphere radius varying on motion sequences	130
A.1	Indoor motion sequences	130
A.2	Outdoor motion sequences	135
B	MOG parameters setting procedure using the $c_1c_2c_3$ colour model	140
C	Combined Classifiers Detailed Assessment results	151
C.1	Two Classifiers	151
C.1.1	Indoor motion sequences	151
C.1.2	Outdoor motion sequences	153
C.2	Three Classifiers with a Non-weighted Combination	154
C.2.1	Indoor motion sequences	154

C.2.2	Outdoor motion sequences	156
C.3	Three Classifiers with a Weighted Combination	158
C.3.1	Indoor motion sequences	158
C.3.2	Outdoor motion sequences	160
References		163

List of Figures

2.1	The flow chart describes the way the Statistical Background Disturbance technique functions.	10
2.2	The proposed colour model to separate the brightness (β) from the chromaticity component (CD). E_i is the expected chromaticity line for the background image	11
2.3	A sample database image showing different categories available in an indoor environment	13
2.4	A sample of an outdoor image showing different categories that might appear in an outdoor sequences.	14
2.5	Two different classes with the TP, TN, FP and FN identified	15
2.6	Examples of indoor images extracted using the SBD algorithm	20
2.7	Examples of outdoor images extracted using the SBD algorithm	23
2.8	SBD algorithm region labelling	25
2.9	Modified SBD algorithm region labelling	26
2.10	Examples of indoor images extracted using the modified SBD algorithm .	27
2.11	Examples of outdoor images extracted using the modified SBD algorithm	28
3.1	An overall view of how the mixture of Gaussians algorithm operates.	33
3.2	Indoor image shows an input frame and an extracted moving subject with Stauffer mixture of Gaussians algorithm.	35
3.3	Stauffer mixture of Gaussians algorithm used to extract an outdoor frame from the HumanID database (Phillips et al., 2002)	36
3.4	The effect of varying the learning rate on background adaption	37
3.5	Examples of indoor images extracted using the MOG algorithm	45
3.6	Examples of outdoor images extracted using the MOG algorithm	47
4.1	A Hypersphere with clustered data. Data in the sphere will be given a positive sign while negative signs will be set for data outside the sphere .	52
4.2	Examples of indoor and outdoor images extracted using the Unary Classifier algorithm using the RGB colour model	54
4.3	Examples of indoor images extracted using the UC algorithm	55
4.4	Examples of outdoor images extracted using the UC algorithm	57
4.5	Examples of indoor images extracted using the improved UC algorithm .	63
4.6	Examples of outdoor images extracted using the improved UC algorithm .	66
5.1	RGB colour cube	70
5.2	Indoor images extracted using different colour models.	76
5.3	The averaged error of ten indoor sequences in different regions along with the overall percentage error	78

5.4	Outdoor images extracted using different colour models.	80
5.5	The averaged error of ten outdoor sequences in different regions along with the overall percentage error	82
5.6	Examples of indoor images extracted using the MOG algorithm with the $c_1c_2c_3$ colour model.	84
5.7	Examples of outdoor images extracted using the MOG algorithm with the $c_1c_2c_3$ colour model.	87
6.1	Indoor images extracted using the combined UC and SBD classifiers . .	95
6.2	Indoor images extracted using the combined UC and MOG classifiers . .	96
6.3	Indoor images extracted using the combined MOG and SBD classifiers .	98
6.4	Outdoor images extracted using the combined UC and SBD classifiers .	100
6.5	Outdoor images extracted using the combined UC and MOG classifiers .	101
6.6	Outdoor images extracted using the combined MOG and SBD classifiers .	103
6.7	Indoor images extracted by combining the UC, the SBD, and the MOG classifiers using the Maximum rule	105
6.8	Indoor images extracted by combining the UC, the SBD, and the MOG classifiers using the Median rule	106
6.9	Indoor images extracted by combining the UC, the SBD, and the MOG classifiers using the Product rule	107
6.10	Indoor images extracted by combining the UC, the SBD, and the MOG classifiers using the Sum rule	108
6.11	Indoor images extracted by combining the UC, the SBD, and the MOG classifiers using the weighted Maximum rule	109
6.12	Indoor images extracted by combining the UC, the SBD, and the MOG classifiers using the weighted Median rule	110
6.13	Indoor images extracted by combining the UC, the SBD, and the MOG classifiers using the weighted Product rule	111
6.14	Indoor images extracted by combining the UC, the SBD, and the MOG classifiers using the weighted Sum rule	112
6.15	Outdoor images extracted by combining the UC, the SBD, and the MOG classifiers using the Maximum rule	115
6.16	Outdoor images extracted by combining the UC, the SBD, and the MOG classifiers using the Median rule	116
6.17	Outdoor images extracted by combining the UC, the SBD, and the MOG classifiers using the Product rule	117
6.18	Outdoor images extracted by combining the UC, the SBD, and the MOG classifiers using the Sum rule	118
6.19	Outdoor images extracted by combining the UC, the SBD, and the MOG classifiers using the weighted Maximum rule	119
6.20	Outdoor images extracted by combining the UC, the SBD, and the MOG classifiers using the weighted Median rule	120
6.21	Outdoor images extracted by combining the UC, the SBD, and the MOG classifiers using the weighted Product rule	121
6.22	Outdoor images extracted by combining the UC, the SBD, and the MOG classifiers using the weighted Sum rule	122
B.1	The effect of varying the learning rate on background adaption for MOG with a $c_1c_2c_3$ colour model using an indoor motion sequence	141

B.2 The effect of varying the learning rate on background adaption for MOG
with a $c_1c_2c_3$ colour model using an outdoor motion sequence 143

List of Tables

2.1	Indoor sequences list	14
2.2	Outdoor sequences list	15
2.3	Detailed variance assessment for all the used measurements in Table 2.5 .	18
2.4	Overall assessment on a number of motion indoor sequences using the SBD algorithm	19
2.5	Overall assessment on a number of motion indoor sequences using the SBD algorithm	21
2.6	An overall assessment on a number of motion outdoor sequences using the SBD algorithm	24
2.7	Overall assessment on a number of motion indoor sequences using the improved SBD algorithm	29
2.8	Overall assessment on a number of motion outdoor sequences using the modified SBD algorithm	29
3.1	Assessment tests on the MOG using different learning rates on an indoor motion sequence	38
3.2	Assessment tests on the MOG using different learning rates on an outdoor sequence	39
3.3	Assessment tests on the MOG using different initial weights for an indoor motion sequence	40
3.4	Assessment tests on the MOG using different initial weights on an outdoor motion sequence	40
3.5	Assessment tests on the MOG using different background thresholds for an indoor motion sequence	41
3.6	Assessment tests on the MOG using different background thresholds . . .	41
3.7	Assessment tests on the MOG using different Initial Variances for an indoor motion sequence	42
3.8	Assessment tests on the MOG using different initial variances on an outdoor motion sequence	42
3.9	Assessment tests on the MOG algorithm using different number of Gaussians per pixel for an indoor motion sequence	43
3.10	Assessment tests on the MOG algorithm using different number of Gaussians per pixel on outdoor motion sequence	44
3.11	Overall assessment on a number of motion indoor sequences using the MOG algorithm	46
3.12	Overall assessment on a number of motion outdoor sequences using the MOG algorithm	48

4.1	Overall assessment on a number of motion indoor sequences using the UC algorithm	56
4.2	Overall assessment on a number of motion outdoor sequences using the UC algorithm	58
4.3	Assessment on the indoor motion sequences 018a059s00L using the improved UC algorithm	60
4.4	Overall assessment on a number of motion indoor sequences using the UC algorithm	61
4.5	Assessment on the outdoor motion sequences 018e059s01L using the improved UC algorithm	64
4.6	Overall assessment on a number of motion outdoor sequences using the UC algorithm	64
5.1	The effect of changing the standard deviation on an indoor sequence using different colour models	74
5.2	The effect of changing σ on a outdoor sequence using different colour models	75
5.3	Averaged overall assessment of different colour models on 10 motion indoor sequences using the UC algorithm	77
5.4	Averaged overall assessment of different colour models on 10 motion outdoor sequences using the UC algorithm	81
5.5	Overall assessment on a number of motion indoor sequences using the MOG algorithm with the $c_1c_2c_3$ colour model.	85
5.6	Overall assessment on a number of motion outdoor sequences using the MOG algorithm with the $c_1c_2c_3$ colour model.	86
6.1	The overall assessment of combining three sets of two combined classifiers using indoor motion sequences	98
6.2	The overall assessment of combining three sets of two combined classifiers using outdoor motion sequences	102
6.3	The overall indoor motion sequences assessment of combining three classifiers using the Maximum, Median, Sum, and product using two different principle of combination	113
6.4	The overall outdoor motion sequences assessment of combining three classifiers using the Maximum, Median, Sum, and product using two different principle of combination	124
A.1	Assessment on the indoor motion sequences 008a013s00L using the improved UC algorithm	130
A.2	Assessment on the indoor motion sequences 009a017s00L using the improved UC algorithm	131
A.3	Assessment on the indoor motion sequences 010a024s08L using the improved UC algorithm	131
A.4	Assessment on the indoor motion sequences 013a037s00L using the improved UC algorithm	132
A.5	Assessment on the indoor motion sequences 013a040s00L using the improved UC algorithm	132
A.6	Assessment on the indoor motion sequences 017a054s00L using the improved UC algorithm	133

A.7	Assessment on the indoor motion sequences 017a055s00R using the improved UC algorithm	133
A.8	Assessment on the indoor motion sequences 018a060s00L using the improved UC algorithm	134
A.9	Assessment on the indoor motion sequences 019a063s00L using the improved UC algorithm	134
A.10	Assessment on the outdoor motion sequences 008e013s00L using the improved UC algorithm	135
A.11	Assessment on the outdoor motion sequences 009e017s01L using the improved UC algorithm	135
A.12	Assessment on the outdoor motion sequences 010e024s00L using the improved UC algorithm	136
A.13	Assessment on the outdoor motion sequences 013e037s00L using the improved UC algorithm	136
A.14	Assessment on the outdoor motion sequences 013e040s00L using the improved UC algorithm	137
A.15	Assessment on the outdoor motion sequences 017e054s00L using the improved UC algorithm	137
A.16	Assessment on the outdoor motion sequences 017e055s00R using the improved UC algorithm	138
A.17	Assessment on the outdoor motion sequences 018e060s00L using the improved UC algorithm	138
A.18	Assessment on the outdoor motion sequences 019e063s05L using the improved UC algorithm	139
B.1	Assessment tests on the MOG using different learning rates on an indoor motion sequence with a $c_1c_2c_3$ colour model	142
B.2	Assessment tests on the MOG using different learning rates on an outdoor motion sequence with a $c_1c_2c_3$ colour model	144
B.3	Assessment tests on the MOG using different Initial Weigh on an indoor motion sequence with a $c_1c_2c_3$ colour model	145
B.4	Assessment tests on the MOG using different Initial Weigh on an outdoor motion sequence with a $c_1c_2c_3$ colour model	145
B.5	Assessment tests on the MOG using different background thresholds for an indoor motion sequence with a $c_1c_2c_3$ colour model	146
B.6	Assessment tests on the MOG using different background thresholds for an outdoor motion sequence with a $c_1c_2c_3$ colour model	147
B.7	Assessment tests on the MOG using different Initial Variances for an indoor motion sequence with a $c_1c_2c_3$ colour model	147
B.8	Assessment tests on the MOG using different Initial Variances for an outdoor motion sequence with a $c_1c_2c_3$ colour model	148
B.9	Assessment tests on the MOG using different number of Gaussians per pixel for an indoor motion sequence with a $c_1c_2c_3$ colour model	150
B.10	Assessment tests on the MOG using different number of Gaussians per pixel for an outdoor motion sequence with a $c_1c_2c_3$ colour model	150
C.1	Combining the UC and the MOG classifiers for indoor motion sequences .	151
C.2	Combining the UC and the SBD classifiers for indoor motion sequences .	152
C.3	Combining the MOG and the SBD classifiers for indoor motion sequences	152

C.4	Combining the UC and the MOG classifiers for outdoor motion sequences	153
C.5	Combining the UC and the SBD classifiers for outdoor motion sequences	153
C.6	Combining the MOG and the SBD classifiers for outdoor motion sequences	154
C.7	Combining the UC, the SBD, and the MOG classifiers for indoor motion sequences using the Maximum combination rule	154
C.8	Combining the UC, the SBD, and the MOG classifiers for indoor motion sequences using the Median combination rule	155
C.9	Combining the UC, the SBD, and the MOG classifiers for indoor motion sequences using the Product combination rule	155
C.10	Combining the UC, the SBD, and the MOG classifiers for indoor motion sequences using the Sum combination rule	156
C.11	Combining the UC, the SBD, and the MOG classifiers for outdoor motion sequences using the Maximum combination rule	156
C.12	Combining the UC, the SBD, and the MOG classifiers for outdoor motion sequences using the Median combination rule	157
C.13	Combining the UC, the SBD, and the MOG classifiers for outdoor motion sequences using the Product combination rule	157
C.14	Combining the UC, the SBD, and the MOG classifiers for outdoor motion sequences using the Sum combination rule	158
C.15	Combining the UC, the SBD, and the MOG classifiers for indoor motion sequences using a weighted Maximum combination rule	158
C.16	Combining the UC, the SBD, and the MOG classifiers for indoor motion sequences using a weighted Median combination rule	159
C.17	Combining the UC, the SBD, and the MOG classifiers for indoor motion sequences using a weighted Product combination rule	159
C.18	Combining the UC, the SBD, and the MOG classifiers for indoor motion sequences using a weighted Sum combination rule	160
C.19	Combining the UC, the SBD, and the MOG classifiers for outdoor motion sequences using a weighted Maximum combination rule	160
C.20	Combining the UC, the SBD, and the MOG classifiers for outdoor motion sequences using a weighted Median combination rule	161
C.21	Combining the UC, the SBD, and the MOG classifiers for outdoor motion sequences using a weighted Product combination rule	161
C.22	Combining the UC, the SBD, and the MOG classifiers for outdoor motion sequences using a weighted Sum combination rule	162

Nomenclature

<i>MOG</i>	Mixture of Gaussians algorithm
<i>SBD</i>	Statistical Background Disturbance algorithm
<i>UC</i>	Unary Classifier algorithm
z	Cutoff point for an ROC
\mathbf{x}	Input data
<i>CD</i>	Chrominance distortion
β	Brightness distortion
μ	The mean
σ	The standard deviation
τ_{CD}	One of foreground thresholds for the SBD
$\tau_{\beta lo}$	One of foreground thresholds for the SBD
$\tau_{\beta 1}$	One of background thresholds for the SBD
$\tau_{\beta 2}$	One of background thresholds for the SBD
$P(\cdot)$	Probability
$p(\cdot)$	Probability density function
Σ	Covariance matrix
K	Number of distributions
w	The weight estimate for a distribution
d	The input dimension
M	Matching distributions variable
η	Gaussian probability density function
B	The background model distributions
T	The background threshold
Φ	A feature mapping function
k	The kernel function
R	The hypersphere radius in a kernel space
c	The hypersphere centre in a kernel space
ℓ	The number of the trained samples
ν	A margin coefficient that controls the amount of training vectors in a sphere
ξ_i	The error cost corresponding to each training vector
α	The Lagrange multiplier

Acknowledgements

I am very grateful to my supervisors Prof. Mark Nixon and Dr. Steve Gunn whom without their guidance and valuable support, all along my thesis, the end result would have not been the same.

I would like to show my gratitude to my parents, Noura and Hathal, my wife, Muzah, and my children, Mohammad and Moneerah, for all their endless support and all what they have to go through for me to achieve this work.

Chapter 1

Introduction

1.1 Motivation

Detecting moving objects is an initial step of information extraction in many computer vision applications including: video surveillance, people tracking and traffic monitoring. In these applications, reliable moving object extraction is required. Such a procedure should feature:

- High accuracy in shape detection.
- Flexibility to handle diverse scenarios (indoor, outdoor) and weather conditions (clouds, rain, etc).
- Ability to operate at video rate (in some applications).

In the following sections of this chapter some of the potential applications will be presented as a motivation to this work on motion estimation. This will be followed by a section that will highlight the contributions of this work. The next section will state the assumptions that will be used as a base to produce new motion extraction techniques. The final section will introduce the layout of the whole thesis.

1.2 Motion estimation

The problem of motion estimation has been approached using different methodologies. For each there are advantages and disadvantages. The final application, the requirements, speed and quality of the method usually dictate the approach and the criteria to be chosen. In this chapter we will give a brief overview of some of the approaches to motion extraction.

1.2.1 Background subtraction

In the most basic approaches, moving regions are detected through a pixel by pixel differencing between a current image and a reference background image (a background model). Then, the resulting image is usually thresholded. The final result is a binary segmented image of the objects existing in the new frames and not in the reference image (i.e. the moving segments).

There are many methods available to produce the background model. Simple ways include taking an image of the background without moving objects. Methods using this background model suffer from many problems, such as the requirement of a training period where moving objects are absent. In addition, the background objects are assumed to be static and any movement of a background object (e.g. removal of an object from the background) would disturb the result of the tested frames. This requires the whole background model to be produced again. Also the system cannot handle gradual change in the background illumination (such as illumination change between daylight and dusk or change due to clouds and rain). Rosin (2002) used the simple way of background subtraction and implemented four different thresholding techniques to overcome the noise resulting from the differencing process.

Other approaches differ in forming the background model and in the procedure used to update the model. The simplest geometric background model is produced by averaging a sequence of images temporally (Dagless et al., 1993). A more robust form is to use the median (Arseneau and Cooperstock, 1999; Cucchiara, Grana, Piccardi, and Prati, 2000; Cutler and Davis, 1998) to form the background model. In Arseneau and Cooperstock (1999) a simple temporal median over N frames was used. Cutler and Davis (1998) used the same method for forming the background model for grey images. For colour images the background model was formed by computing each colour channel (RGB) separately. Cucchiara et al. (2000) used the median filter of N frames joined with a previously computed background (with the median as well) to form the model.

Other methods were also used to make the background model adaptive to background illumination changes. Kilger (1992) used a Kalman filter to adapt to the temporal dynamic changes of the weather and illumination.

1.2.2 Statistical methods

Many statistical techniques for motion extraction were inspired by background differencing methods and can be seen as improvements to the differencing methods (notice that some of the techniques mentioned on the background difference section could also be described as statistical methods).

Some statistical approaches provided a better background model (compared to the basic differencing technique) by using the characteristics of each individual pixel or groups of pixels (like illumination, colour models, etc). Each pixel can be labelled as foreground, background or even shadow by comparing the tested pixel with the available statistics of the different regions (the background, the foreground and the shadow).

Horprasert et al. (1999, 2000) introduced a new computational colour model which separates the brightness from the chromaticity component. This separation was used to distinguish between shading background from the ordinary background or moving foreground objects.

Gaussian models were also used as methods to obtain a motion extractor. A single Gaussian per pixel was used to model the background by Wren et al. (1997). While Stauffer and Grimson (1999) and Friedman and Russell (1997) have used a mixture of Gaussians to model the moving object and the background, Friedman and Russell (1997) modelled the shadow as well.

1.2.3 Temporal differencing

To extract moving regions, the temporal differencing approach applies a pixel-wise differencing between two or three consecutive frames in an image sequence. Even though temporal differencing can adapt to change in background, the method can suffer from poor quality whereby holes may exist in the extracted body. Holes are due to differencing a foreground pixel that appears on the same pixel in the successive frames (this happens if the moving object stops or if a foreground surface of a similar colour covers many consecutive pixels spatially).

Lipton et al. (1998) performed a temporal difference between consecutive video frames (two frames temporal differencing) followed by a thresholding process to determine change (i.e. moving pixels). A double differencing system was presented by Kameda and Minoh (1996), where three successive frames are used in a differencing operation. Each two frames are differenced separately and then an AND operator is applied on the resulting differenced images. The method was applied as a day time traffic monitoring system (Cucchiara, M. Piccardi, and Mello, 2000).

1.2.4 Stereo techniques

Stereo methods were used as another geometric method to form a background model (Jones and Malik, 1992; Polat et al., 2003), especially after the recent development of real time depth computation from stereo cameras. In (Polat et al., 2003) a stereo system is used for 3D feature extraction. The features are extracted from 2 camera sequences and matched to obtain the 3D coordinates. Stereo methods based on depth alone can form unreliable results in substantial parts of the

scene, and can fail to extract moving objects in close proximity to the background, such as feet touching a floor (Eveland et al., 1998).

1.3 Applications

This area of development is powered by the potential applications of motion extraction:

Surveillance (Kim and Kim, 2003; McKenna et al., 2000; Collins et al., 2001) is a major beneficiary of motion estimation. In this area subjects are monitored and possibly tracked over time. The security of the monitored premises is related to the robustness and efficiency of the surveillance system.

Traffic Control (Smith et al., 1996; Michalopoulos, 1991) where moving vehicles with different speeds and different sizes are extracted and monitored for traffic control.

Gait Recognition (Grant et al., 2002; Hayfron-Acquah et al., 2001) is used as a biometric concerning recognising people by the way they walk. In gait recognition, a moving object is extracted from a sequence of frames (one cycle). The extracted moving objects are then used to find a gait signature for that person.

Motion Capture Controlled Devices - this concerns devices that are controlled and directed through extracted motion. In many industries, especially where conveyor belts are used, a high number of repeated operations occur. These are mainly carried out by machines/robots to ensure similarity in production and speed. As the operation becomes more complex it becomes more complicated to program them to carry out the operations. To tackle this problem, a human performs the operation while a computer captures his/her motion. Afterwards, the computer may use the captured motion data to control the same machine performing the operation (Moeslund, 2000).

Video Coding (Jing et al., 2003; Boinovic and Konrad, 2005)- motion estimation plays an important role in this field (Su et al., 1999). The high temporal redundancy between successive frames is exploited to achieve high compression efficiency.

Virtual Reality Systems (Ohya et al., 1999; Davis and Bobick, 1997) and **Video Games** (Wren et al., 1997) - Motion extraction presents a powerful business tool for recreation and virtual reality games. In the ALIVE system (Maes et al., 1997), the motion extraction method is utilised to place the user in a scene with some artificial life forms, in real-time.

Medical Applications (Patias, 2002) motion extraction can be utilised for moving disabilities monitoring, clinical studies of orthopaedic patients, and X-rays of moving body parts.

Motion estimation can be accomplished through either active or passive sensing. In active sensing, devices are placed on the subject (to transmit the generated signals) and in the surroundings (to receive the generated signals). Active sensing is widely used when the applications are situated in well-controlled environment (laboratories). In passive sensing, natural signal sources are used, e.g. visual light. In applications where mounting devices on the subject is not an option, passive sensing is mainly used (e.g. surveillance). Computer vision is used as a tool for implementing passive sensing. In this work passive sensing will be used with no markers placed on moving subjects.

1.4 Contributions

The primary objective of this thesis is to provide a robust technique to extract moving objects. The following techniques were developed and implemented to accomplish this:

- Improved a statistical shadow extraction method by further evaluating the shadow labelling process, Section 2.4.
- Analysed a mixture of Gaussian technique on indoor and outdoor environment and provided a simple way to optimise it, Chapter 3.
- Introduced a new theory of Unary Classification to the area of motion extraction. We provided a parameter optimising procedure of the Unary Classifier motion segmenter along with an assessment of performance, Chapter 4.
- Improved the performance of the Unary Classifier by improving its decision theory. The improvement resulted in an increase in the efficiency of motion extraction for both indoor and outdoor environments, Section 4.6.
- Accomplished an assessment of five different colour models when applied on the area of motion estimation. This process helped us in finding a suitable colour invariant model that can help in suppressing shadows when the motion extraction process is implemented, Chapter 5.
- Improved evaluation process in which it is the first approach that compares foreground/background labelling with results obtained from labelling by broadcast techniques. In addition different methodologies of assessments were utilised. In which each method gave a prospective with a deferent angle providing collectively a better evaluation.
- Provided a simple way of combining motion extraction classifiers and proved that such process can improve the final extracted output. The followed procedure resulted in a novel production of multiple efficient motion extractors, Chapter 6.

- Tested different ways of data fusing, MAX, SUM, PRODUCT and MEDIAN. Also gave a comparison to their overall performance, Chapter 6.
- Provided a new way to estimate the probability for the used classifiers. The results supported the competence of such estimation.

In addition the systems are designed to be model-independent. Model-independent means that it can handle different testing model types (humans, vehicles, bicycles, etc). The models can be detected over different speeds and different motion trajectories.

A comparison of the implemented techniques with two-state-of the art algorithms (Horprasert et al., 2000; Stauffer and Grimson, 2000) is provided in the process of analysing the produced techniques.

This work has so far resulted in the following publications:

- Al-Mazeed, A., M. Nixon, and S. Gunn (2003). Fusing complementary operators to enhance foreground/background segmentation. In the British Machine Vision Conference (BMVC 2003), pp. 501-510. (Al-Mazeed et al., 2003) BMVC is the main UK conference on machine vision.
- Al-Mazeed, A. H., M. S. Nixon, and S. R. Gunn (2004). Classifiers combination for improved motion segmentation. In A. Campilho and M. Kamel (Eds.), Proceedings of International Conference on Image Analysis and Recognition, Lecture Notes in Computer Science, Volume LNCS 3212, pp. 363-371. Springer. (Al-Mazeed et al., 2004).

1.5 Assumptions

In this research we have developed novel approaches for extracting moving objects in successive frames in unconstrained indoor and outdoor video scenes. Certain assumptions are taken into consideration in our approaches. These assumptions will define among others, the background model and the foreground (or the moving subject):

[**single camera**] A single camera will be used (no stereo algorithms).

[**static camera**] The camera is assumed to be static.

[**background definition**] The background consists of all non-moving objects and objects that have repetitive motion (including computer monitor flicker, tree motion, a flag, sea waves, etc.).

[**foreground definition**] Everything that is not background is considered as foreground (such as humans, animals, vehicles, etc).

[**scene independence**] Object segmentation is model-independent i.e. the system detects different model types (humans, vehicles, bicycles, etc) whatever their speed, motion and trajectory.

1.6 Structure

This chapter has given a brief motivation for motion estimation. We have presented the applications, the contributions of this thesis, and the assumptions that were used as a basis to accomplish the target motion extractor. Also we clarified the data sets used along with the results assessment measures. In this section we will give an outline of the remaining chapters:

- **Chapter 2**

This chapter discusses a motion extraction algorithm called the Statistical Background Disturbance technique, (SBD) (Horprasert et al., 2000, 1999). The chapter presents an improvement made to enhance shadow extraction in this technique. An assessment on the technique is also presented using indoor and outdoor motion sequences.

- **Chapter 3**

This chapter presents a Mixture of Gaussians (MOG) technique as a motion extractor (Stauffer and Grimson, 2000, 1999; Grimson et al., 1998). The Mixture of Gaussians parameters are optimised. The technique is assessed using indoor and outdoor sequences.

- **Chapter 4**

In this chapter we present a new SVM technique called the Unary Classifier. An assessment is pursued on the technique performance in motion extraction. Also it is shown how the technique can be modified to give better performance in motion extraction. The technique is tested on indoor/outdoor sequences.

- **Chapter 5**

Since the MOG algorithm and the Unary Classifier method originally lacked the ability to extract shadows, different invariant colour models are tested to find a colour model that can suppress shadows. The following colour models are tested and compared: RGB, normalised rgb, HSV, and two more colour models titled $l_1l_2l_3$ and $c_1c_2c_3$. Prelabelled images for the shadow, the background and the motion pixels are used to test the colour models performance on the three different regions.

- **Chapter 6**

In this chapter the three different techniques presented earlier (MOG, SBD and

the Unary Classifier) are combined using Bayes theorem. The performance of each combination is assessed using indoor and outdoor sequences. A comparison of the performance of all the combinations are presented.

- **Chapter 7**

The chapter presents conclusions on the overall outcome of this thesis. This chapter also discusses the possible future paths that can be taken to carry on in this work.

Chapter 2

Statistical Background Disturbance

The Statistical Background Disturbance (SBD) algorithm (Horprasert et al., 2000, 1999) decomposes the colour space using prior knowledge established on a statistical computational model to separate the chromaticity from the brightness component. The algorithm is able to cope with shadow and highlights. The outcome of the algorithm is labelled pixels of one of four groups: motion pixels, background pixels, shadow and highlights.

In (Prati et al., 2003) this algorithm was tested against three different shadow extraction algorithms. The algorithm was found to be robust to noise detection. The study also concluded that this algorithm appears to be the best choice (among the tested algorithms) if there are different planes onto which the shadows are cast. Also the algorithm was found to be highly efficient in detecting shadows.

The algorithm was used in a system (Chang and Huang, 2000) to analyse the human walking motion. The system starts with the SBD method to find the moving object. Then the process of analysing the human motion uses a skeleton-based method followed by a Hidden Markov Model (HMM) to describe the motion type. The SBD was also used to segment the head of a system user to form a 3D virtual reality representation of the same user (Rajan et al., 2002).

In the tracking area, the algorithm was used to form silhouettes for articulated body model acquisition and tracking from voxel data (Mikic et al., 2003, 2001). The algorithm was also used in a tracking system to handle occlusion (Senior et al., 2001). Furthermore, a method derived from the basics of the SBD algorithm was used in a surveillance system (KaewTrakulPong and Bowden, 2003).

The SBD algorithm has also been applied in gait recognition. In (Bobick and Johnson, 2001) a gait recognition technique used the SBD technique (brightness distortion) for

shadow extraction. Also the technique was used to find the body contour in a method to analyse and extract the human gait motion (Yoo et al., 2002).

An overview of the algorithm is presented in Figure 2.1.

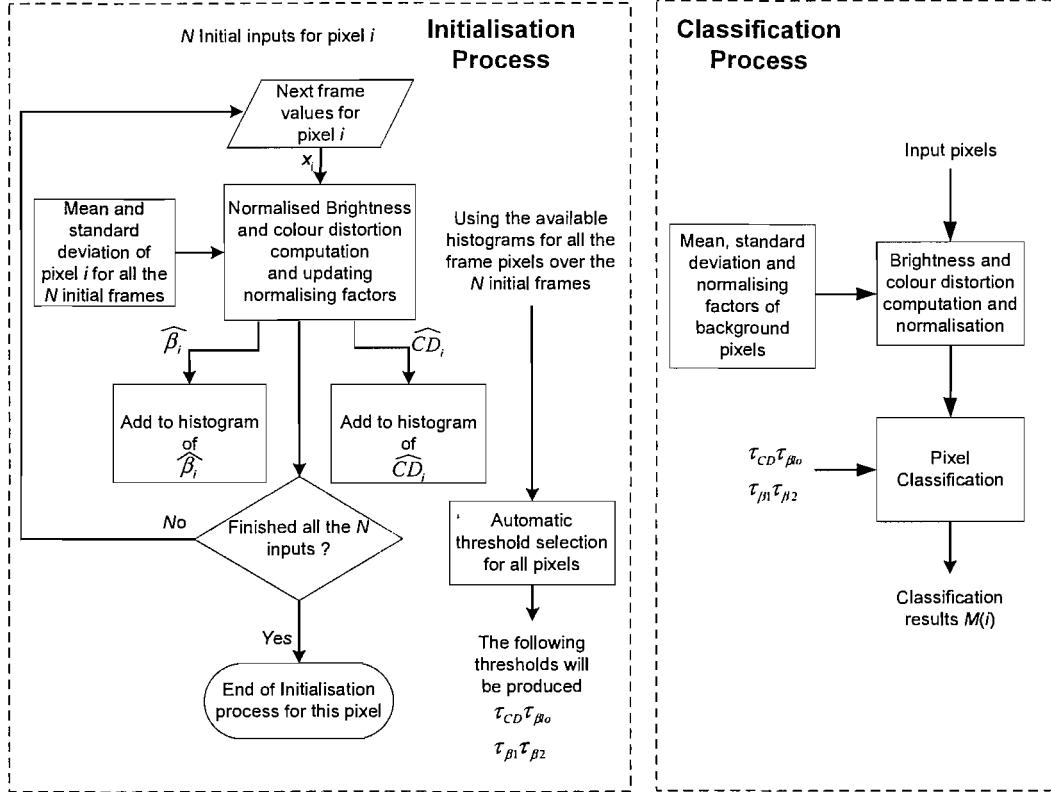


FIGURE 2.1: The flow chart describes the way the Statistical Background Disturbance technique functions.

2.1 Disturbance parameters

Horprasert et al. (1999, 2000) proposed a colour model in a 3 dimensional RGB colour space to separate the brightness from the chromaticity component. Figure 2.2 demonstrates the proposed colour model.

The algorithm initially uses N frames to form the background model. From these frames, the mean and the standard deviation is computed for each colour band (RGB) in each pixel. The chrominance distortion CD and the brightness distortion β between a background pixel and a new pixel x are computed as

$$CD = \sqrt{\sum_{c \in \{R, G, B\}} \left(\frac{x_c - \beta \mu_c}{\sigma_c} \right)^2} \quad (2.1)$$

and

$$\beta = \frac{\sum_{c \in \{R, G, B\}} \left(\frac{x_c \mu_c}{\sigma_c^2} \right)}{\sum_{c \in \{R, G, B\}} \left(\frac{\mu_c}{\sigma_c} \right)^2} \quad (2.2)$$

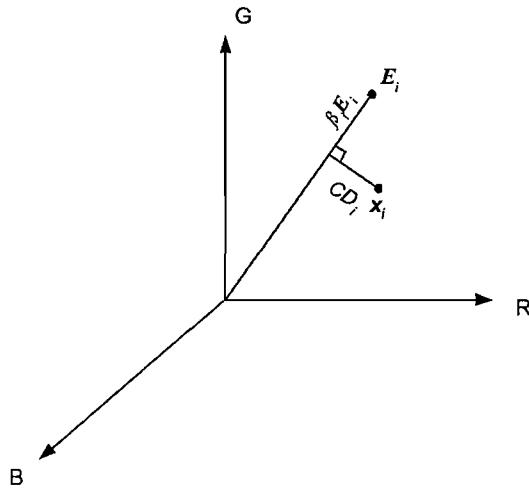


FIGURE 2.2: The proposed colour model to separate the brightness (β) from the chromaticity component (CD). E_i is the expected chromaticity line for the background image

where μ_c and σ_c represent the mean and the standard deviation for each background pixel colour band, respectively. The normalised chrominance distortion \widehat{CD} and the brightness distortion $\widehat{\beta}$ are used to classify each new pixel in frame i

$$\widehat{\beta}_i = \frac{\beta_i - 1}{\sqrt{\sum_{i=1}^N (\beta_i - 1)^2 / N}} \quad (2.3)$$

$$\widehat{CD}_i = \frac{CD_i}{\sqrt{\sum_{i=1}^N (CD_i)^2 / N}} \quad (2.4)$$

The following decision criteria are used to classify each pixel according to

$$M(i) = \begin{cases} \text{Foreground} & : \widehat{CD}_i > \tau_{CD} \text{ or } \widehat{\beta}_i < \tau_{\beta lo}, \\ \text{Background} & : \widehat{\beta}_i < \tau_{\beta 1} \text{ and } \widehat{\beta}_i > \tau_{\beta 2}, \\ \text{Shadow} & : \widehat{\beta}_i < 0 \\ \text{Highlight} & : \text{otherwise} \end{cases} \quad (2.5)$$

where τ_{CD} and $\tau_{\beta lo}$ are thresholds used to specify the borders of the foreground. $\tau_{\beta 1}$ and $\tau_{\beta 2}$ are thresholds used to identify the borders of the background. These thresholds are determined automatically through a statistical learning procedure (Horprasert et al., 1999, 2000).

This algorithm suffered in the design process from classifying motion pixels with low RGB values as shadows (Horprasert et al., 1999). Therefore the authors added the test $\widehat{\beta}_i < \tau_{\beta lo}$ to the foreground decision to solve this problem. Setting the threshold $\tau_{\beta lo}$ is a trade off between the shadow extraction and the motion detection criteria in this algorithm. Setting it to be high will result in better quality motion extraction (moving object extraction) at the expense of a decrease in quality in shadow extraction and vice versa for when it is low. In setting this threshold, the overall error will be considered in order to compensate between the shadow and the quality of the extracted motion subject.

One problem with the way the background is modelled is that there is no learning (the background model is not updated, i.e. not adaptive). The stored background parameters are not updated, which may lead to a deterioration in the quality of extraction if a background change occurs (illumination change or addition or removal of objects to the background).

2.2 Automatic Thresholding

Through the background building process a histogram is constructed for \widehat{CD}_i and $\widehat{\beta}_i$. The thresholds are then computed after fixing a detection rate which fixes the expected proportions of the image contents. The detection rate can be set to any value between 0 and 1 where a large value will result in a better overall motion accuracy and vice versa where it is low. Using a high detection rate will result in including most of the background histogram in consideration when setting the thresholds. On the contrary, having a small detection rate will include only the more frequent pixel values apparent in the tested background frames. This will allow more freedom for the foreground enabled colour values (more colours can be used for the foreground and will not be detected as background) but at the same time increase the background noise. τ_{CD} is the normalised chromaticity distortion value at the detection rate. $\tau_{\beta 1}$ is the value of $\widehat{\beta}_i$ at the detection rate while $\tau_{\beta 2}$ is the value of $\widehat{\beta}_i$ at (1 - detection rate). $\tau_{\beta lo}$ is determined by varying the parameter on a set of training samples until a minimum false detection error is reached.

2.3 Experimental Results

We will start by introducing the data that will be used for our assessment along with the evaluation measures which will be used for this chapter and the chapters that follows. The SBD testing results will then be introduced.

2.3.1 Data and Assessment

10 indoor sequences from the University of Southampton gait database (Shutler et al., 2002) were used in the testing of each method in this thesis. In each one 50-52 frames are used to form a background model (these frames were specifically recorded so as not to contain any foreground moving objects). Each frame is of a dimension of 720 x 367 pixels. Figure 2.3 shows an indoor database sample. Table 2.1 gives a list of the indoor sequences. To analyse the data, silhouettes provided with the database were used as an approximation to ground truth. The silhouettes were generated by chroma-key extraction via the green background (Shutler et al., 2002).

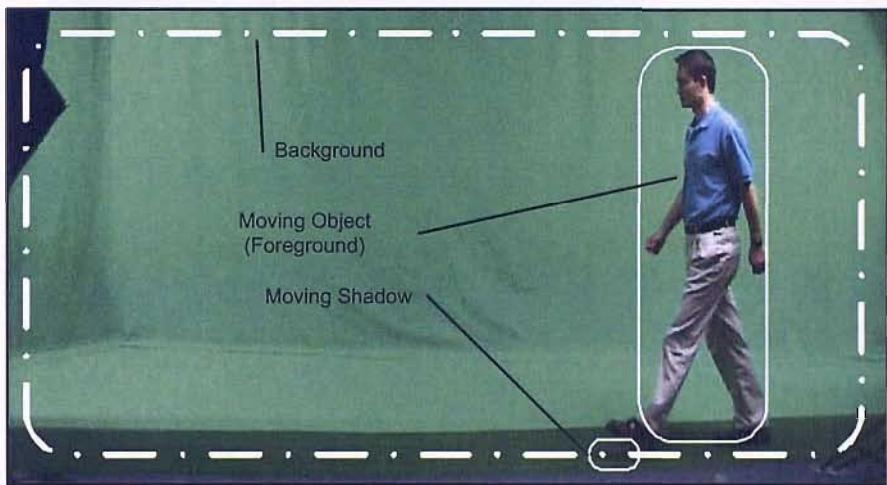


FIGURE 2.3: A sample database image showing different categories available in an indoor environment

To further test the robustness of an algorithm 10 outdoor sequences were used of the form shown in Figure 2.4. These sequences have different weather conditions (windy, cloudy, sunny, etc.) but image the same subjects walking in a non-laboratory environment with a similar geometry. Each frame is of a 220 x 220 pixels dimension. 48-55 background frames were used for background adaption. For each outdoor sequence, silhouettes for moving subjects were produced by manual labelling to provide a form of ground truth. Table 2.2 lists the outdoor sequences.

Throughout the thesis we present samples of extracted sequences for each presented classifier. The same samples are used in each evaluation for cross comparison.

Sequence Number	Number of Frames	Background Frames
008a013s00L	178	51
009a017s00L	169	52
010a024s08L	187	51
013a037s00L	114	50
013a040s00L	184	50
017a054s00L	188	50
017a055s00R	162	50
018a059s00L	188	50
018a060s00L	179	50
019a063s00L	186	50

TABLE 2.1: Indoor sequences list

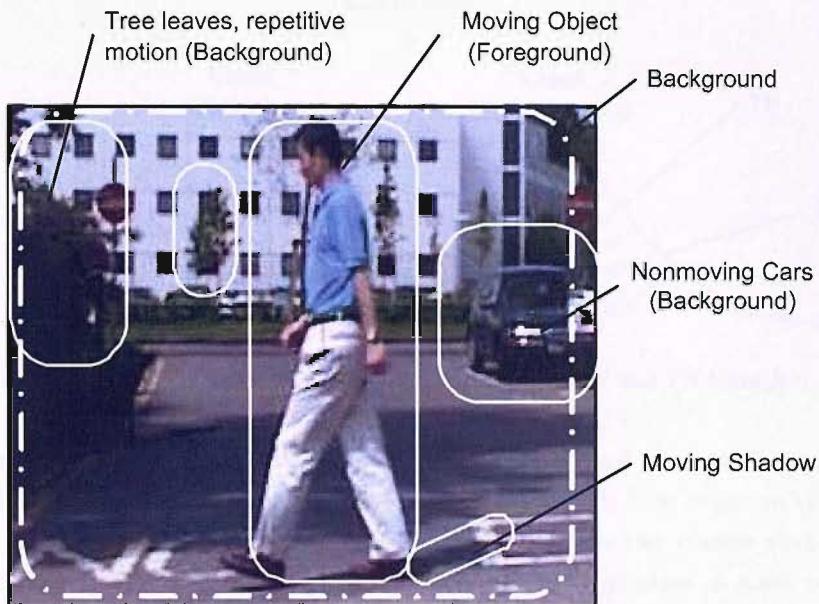


FIGURE 2.4: A sample of an outdoor image showing different categories that might appear in an outdoor sequences.

The following assessment measures were used to evaluate the performance of an algorithm: the Receiver Operating Characteristics curve (ROC) technique, the Root Mean Square Error (RMSE), the Peak Signal to Noise Ratio (PSNR), the percentage of error, and the regional percentage of error.

- **Receiver Operating Characteristics curve (ROC) technique**

The ROC curve is used as an assessment measure to find the optimal settings for a technique parameter. The ROC plot is a graphical representation of points defined by sensitivity and $(1 - \text{specificity})$.

Sequence Number	Number of Frames	Background Frames
008e013s00L	100	48
009e017s01L	96	50
010e024s00L	94	53
013e037s00L	158	55
013e040s00L	151	49
017e054s00L	112	52
017e055s00R	88	50
018e059s01L	104	49
018e060s00L	88	50
019e063s05L	112	50

TABLE 2.2: Outdoor sequences list

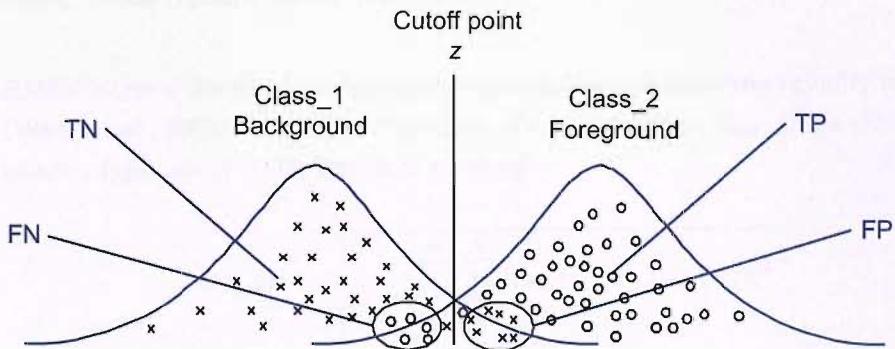


FIGURE 2.5: Two different classes with the TP, TN, FP and FN identified

Given a set of cutoffs (a cutoff is a parameter value) and test results, each observation can be classified into one of: true positive (TP); false negative (FN); true negative (TN); or false positive (FP). Figure 2.5 shows two classes with a cutoff z . *Class₁* (the x's) can be visualised as the background class in a set of motion sequences and *class₂* (the circles) as the foreground class. The cutoff point will classify all the points to its left as part of *class₁* (background) and any pixels to the right as part of *class₂* (foreground). Any background pixel classified on the left of the cutoff point is regarded as a TN. A background pixel found on the right of the cutoff point (the circled group of x's on the right) will be regarded as FP. For the foreground pixels, if they are on the right of the cutoff point then they are TP otherwise they are FN (the circled circles on the left). The Specificity, S_p , and the Sensitivity, S_n , are calculated for each cutoff point

$$S_p = \frac{TN}{TN + FP} \quad (2.6)$$

$$S_n = \frac{TP}{TP + FN} \quad (2.7)$$

Finding the best cutoff point (or the optimal cutoff point) which gives the best setting for a parameter (i.e. the best threshold) is the main concern for using this technique. This goal can be accomplished through finding the maximum value for the ROC optimal cutoff measure (Grzymala-Busse et al., 2003, 2002), which is a measure of the difference between the probability of true positive and false positive probability as

$$\begin{aligned} ROC_{optimal_cutoff} &= \arg \max_z (S_{n_z} - (1 - S_{p_z})) \\ &= \arg \max_z (S_{n_z} + S_{p_z} - 1) \end{aligned} \quad (2.8)$$

- **Root Mean Square Error (RMSE)**

RMSE is one of the widely used assessment techniques in distortion/quality metrics (Wang et al., 2003). When the dimension of a tested image, I_{test} , and a silhouette image, I_{Silh} , are $N \times M$, RMSE is given by

$$RMSE = \sqrt{\frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N [I_{test}(i, j) - I_{Silh}(i, j)]^2} \quad (2.9)$$

- **Peak Signal to Noise Ratio (PSNR)**

The PSNR in (dB) is computed by

$$PSNR = 20 \log_{10} \frac{L}{RMSE} \quad (2.10)$$

where L represents the dynamic range of a pixel (Wang et al., 2003).

Although RMSE, and PSNR have been widely criticised (Toet and Lucassen, 2004; Lai and Kuo, 2000; Winkler, 1999a; Wang and Bovik, 2002; Eckert and Bradley, 1998; Eskicioglu and Fisher, 1995; Winkler, 1999b; Teo and Heeger, 1994), they are widely used because they are simple to calculate. Also they have a clear physical meaning, and are mathematically easy to deal with for optimisation purposes.

- **Overall percentage error**

The overall percentage error is compared to number of motion pixels in the silhouettes instead of comparing it to the whole image dimension (as in the RMSE).

The overall percentage error will be calculated as follows

$$Overall\ Percentage\ Error = \frac{Misclassified\ Pixels}{Silhouette's\ Motion\ Pixels} \times 100 \quad (2.11)$$

The main concern in our results is extracting the motion pixels. Therefore having a measure describing the percentage of error compared to the motion pixels will give us a measure of how much that error compares to our moving object.

This measure might be affected by bias to the background if the moving object is relatively very small compared with the background, or vice versa. Knowing the average relative size of the foreground to the background can ameliorate this problem. For the motion sequences used, the foreground represents around 10% of the background area.

- **Regional percentage error**

The regional percentage error is used to measure the percentage error for a specific region. The region in our tests will be either the background or foreground (motion pixels). Also it will be used for shadow region assessment in the colour model chapter, Chapter 5.

$$\text{Regional Percentage Error} = \frac{\text{Misclassified Regional Pixels}}{\text{Region's Pixels}} \times 100 \quad (2.12)$$

The RMSE and the PSNR can give false indications to the performance of an algorithm. For instance if we have a system that performs well in suppressing background pixels but has a bad performance in motion extraction. The RMSE can be minimised by using images with a background larger than the (small) moving object. In this case the RMSE will give a smaller error value even if more than half of the moving object is not labelled correctly. A better way to measure the performance is to measure the performance of a method in each region (background, foreground), i.e. regional percentage error. Now in the regional percentage error, if a region was more proportionally dominant than another, the regional error will still indicate the correct percentage of error in that region.

- **The Variance**

To test the consistency of our results we will use the variance of one of the presented errors. The variance of an error measure, ERM , of N samples and a mean of, μ_{ERM} , is calculated by

$$\sigma^2 = \frac{\sum_{i=1}^N (ERM_i - \mu_{ERM})^2}{N - 1} \quad (2.13)$$

The error measure should cover the whole image area and not part of it (which excludes the foreground and the background error measures). The overall error is a relative measure comparing the error to a changing value of the size of the silhouette's motion pixels. Thus to measure the consistency for each tested sequence result we are left with the RMSE and the PSNR.

Seq. No.	σ_{RMSE}^2	σ_{PSNR}^2	σ_{FG}^2	σ_{BG}^2	$\sigma_{Overall}^2$
008a013s00L	4.886E-05	0.661	0.344	0.004	0.854
009a017s00L	3.893E-05	0.468	1.236	0.002	1.194
010a024s08L	6.261E-05	0.733	2.101	0.003	3.157
013a037s00L	3.822E-05	0.728	0.985	0.002	1.034
013a040s00L	6.092E-05	0.762	0.626	0.012	2.571
017a054s00L	5.722E-05	0.325	0.518	0.032	5.991
017a055s00R	2.682E-05	0.399	0.502	0.003	1.576
018a059s00L	2.826E-05	0.442	0.374	0.001	0.858
018a060s00L	2.736E-05	0.328	1.097	0.007	0.860
019a063s00L	7.372E-05	0.996	3.636	0.003	2.792
Average	4.629E-05	0.584	1.142	0.007	2.089

TABLE 2.3: Detailed variance assessment for all the used measurements in Table 2.5

The PSNR function is based on finding the log of the reciprocal of the RMSE. The PSNR results were processed using the logarithmic function while the RMSE preserve the error evaluation without such processing. Therefore we will use the variance of the RMSE as a consistency measure in our implemented evaluations.

Despite this, we have included a tabulation of the variances of all measures in Table 2.3. The tabulated variances are for the results of the SBD indoor extraction that will be shown and discussed later in this chapter. An illustrative copy of the extraction table is also given below , Table 2.4.

We notice in Table 2.3, that the variance of the foreground error is proportionally higher than for the other measures. This is because the foreground concerns the silhouette of the walking subject and the variance can be much higher as it depends on the gait of the walking subject, relative to the sequence. The overall error is proportionally high as well since, as mentioned before, this error compares the whole error in each frame to the motion pixels in a silhouette. The size of the motion pixels in silhouette changes from one frame to another. Therefore the results of those two measures will be considered with caution when drawing conclusions.

The small RMSE variance, in Table 2.3, justifies the use of three digits after the decimal point for the RMSE. Those digits are significant with such small variance. Further, we can see the the other variances are proportionate. For these reasons, only the RMSE variance will be quoted in later tables.

2.3.2 SBD results

We tested the method on indoor and outdoor sequences with a detection rate set to 0.9999 to accommodate most of the background pixel values within the background model giving it a better representation of the scene. In the following parts of this section we present the indoor and the outdoor tests along with assessment measures on each. Samples of indoor and outdoor extractions are also presented.

Sequence Number	Number of Frames	RMSE	PSNR (dB)	FG Error ¹	BG Error ²	Overall Error ³
008a013s00L	178	0.077	22.275	3.810	0.346	8.139
009a017s00L	169	0.080	21.952	4.613	0.372	10.044
010a024s08L	187	0.080	21.959	4.707	0.375	10.316
013a037s00L	114	0.065	23.733	4.671	0.192	8.114
013a040s00L	184	0.080	21.938	4.169	0.427	10.908
017a054s00L	188	0.119	18.529	1.160	1.432	21.215
017a055s00R	162	0.072	22.879	3.297	0.332	8.133
018a059s00L	188	0.071	23.047	3.846	0.266	7.640
018a060s00L	179	0.081	21.830	3.928	0.427	9.840
019a063s00L	186	0.075	22.550	4.302	0.319	9.138
Average		0.080	22.069	3.850	0.449	10.349

1 FG Error Foreground Percentage Error

2 BG Error Background Percentage Error

3 The percentage of the overall error compared to the motion pixels only

TABLE 2.4: Overall assessment on a number of motion indoor sequences using the SBD algorithm

• Indoor Motion Sequences

10 indoor sequences are used in these tests. 50-52 frames from each were used for background modelling (these frames are pure background and do not contain any moving objects).

Samples of the extracted images are shown in Figure 2.6 presenting some of the problems in the SBD extraction. From the samples shown, the background pixels were mostly detected correctly though with some problem near the seam of the background cloth. The background is suppressed well in the samples 008a013s00L, 009a017s00L and 010a024s08L with few misclassified background pixels. The extracted background in sample 017a054s00L has some noise (there are few misclassified background pixels).

The motion pixels can be misclassified as reflected by the holes that appear in the moving objects (this is not always the case; we focus on examples that are instructive to performance analysis). The holes are most serious in 009a017s00L where the legs have many small isolated holes, and some larger connected ones. The holes are due to the statistical nature of the technique: the area in which the larger holes occur is consistent across the sequences, but the small isolated holes appear uncorrelated. Sequences 008a013s00L, 010a024s08L and 017a054s00L also have holes.

The noise is due to the large background intensity variation in these pixels that the algorithm cannot label correctly. The algorithm uses the same thresholds on all the pixels (global thresholding). This property leads the algorithm to try to compensate between having large thresholds that can accommodate variation in

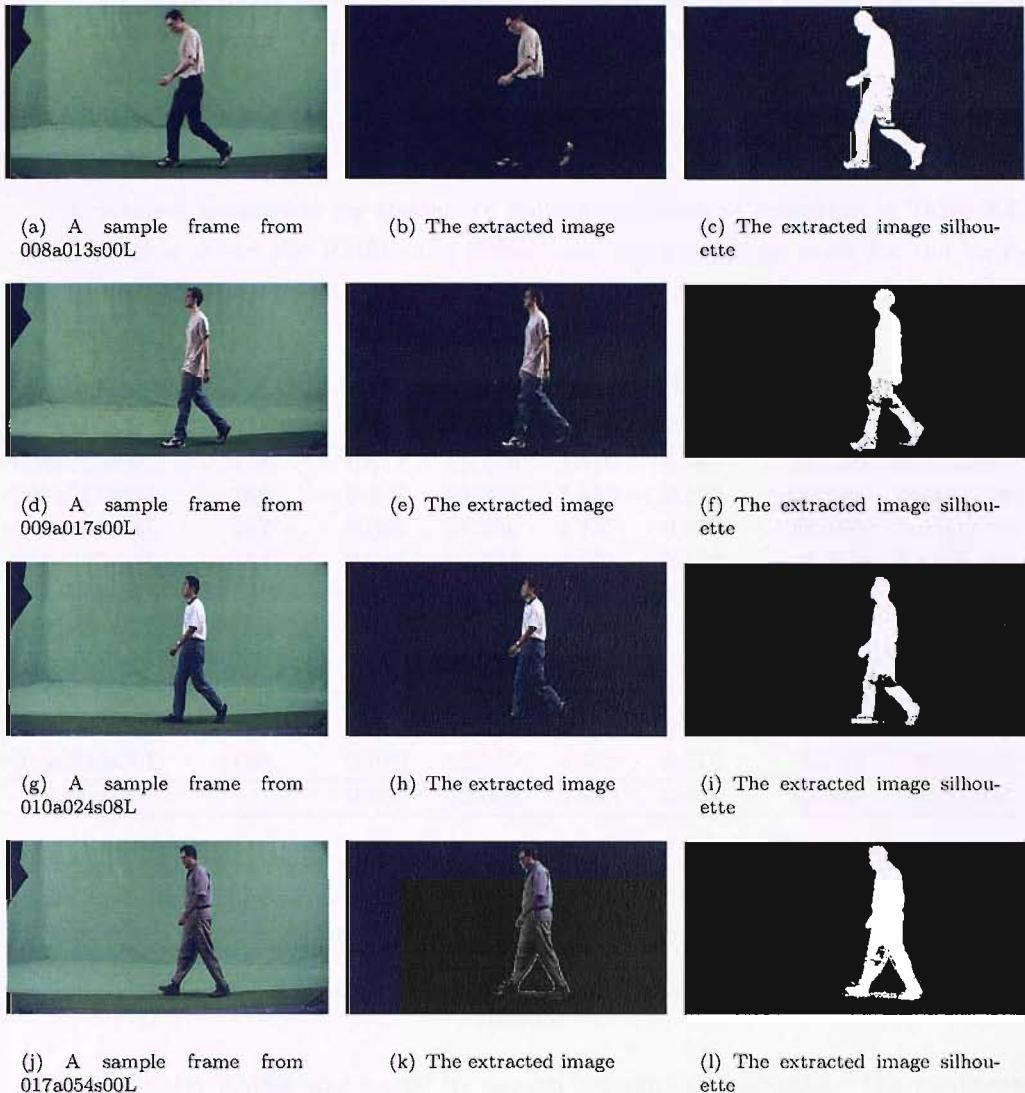


FIGURE 2.6: Examples of indoor images extracted using the SBD algorithm

intensity (the isolated noise), and between having small thresholds to cover only the most common background pixel values. This avoids a background pixel from covering more colour intensity which gives it more tendency to consider some of the moving object colours as background, as we have in the large holes in the moving objects samples.

The uncorrelated noise pixels and even the larger holes can be overcome if other motion segmenter's noise appears in another position. A combination can then result in a better overall output. Such combination requires a confidence measure for each decision showing poor confidence on the misclassified pixels and high confidence on the correctly labelled pixels, to be discussed in Chapter 4.

The shadow suppression performs well in the tested indoor images though some shadow can remain in the extracted images. The shadow has mostly disappeared in sequences 008a013s00L and 009a017s00L. Small traces of the shadow appear in sample 010a024s08L. The shadow is clearly apparent in sequence 017a054s00L sample.

A detailed assessment for the indoor motion sequences is presented in Table 2.5. The table shows the RMSE, the PSNR, and the percentage error for the background, foreground and the overall error compared to the frames silhouette motion pixels.

Sequence Number	Number of Frames	RMSE	PSNR (dB)	FG Error ¹	BG Error ²	Overall Error ³	σ_{RMSE}^2
008a013s00L	178	0.077	22.275	3.810	0.346	8.139	4.886E-05
009a017s00L	169	0.080	21.952	4.613	0.372	10.044	3.893E-05
010a024s08L	187	0.080	21.959	4.707	0.375	10.316	6.261E-05
013a037s00L	114	0.065	23.733	4.671	0.192	8.114	3.822E-05
013a040s00L	184	0.080	21.938	4.169	0.427	10.908	6.092E-05
017a054s00L	188	0.119	18.529	1.160	1.432	21.215	5.722E-05
017a055s00R	162	0.072	22.879	3.297	0.332	8.133	2.682E-05
018a059s00L	188	0.071	23.047	3.846	0.266	7.640	2.826E-05
018a060s00L	179	0.081	21.830	3.928	0.427	9.840	2.736E-05
019a063s00L	186	0.075	22.550	4.302	0.319	9.138	7.372E-05
Average		0.080	22.069	3.850	0.449	10.349	4.629E-05

¹ FG Error Foreground Percentage Error

² BG Error Background Percentage Error

³ The percentage of the overall error compared to the motion pixels only

TABLE 2.5: Overall assessment on a number of motion indoor sequences using the SBD algorithm

The highest RMSE was scored by motion sequence 017a054s00L. The minimum PSNR was also scored in this motion sequence. The lowest RMSE occurred with 013a037s00L which has also given the highest PSNR. For the foreground percentage the values were fairly similar (in the range of 3.30% to 4.71%) except for 017a054s00L where it gave a smaller value of 1.16%. The average foreground error was 3.85%. The maximum value for the background percentage error was 1.43% for the sequence 017a054s00L. The minimum background error is 0.19% and the average background error is 0.45%. The overall error compared to the silhouette's motion pixels ranged from 7.64%, by 018a059s00L, to 21.22% in 017a054s00L (21% is an abnormal maximum error where the second max is 10.91%). The overall average for the percentage error is 10.35%. The variance of the averaged RMSE measure shows that the error displacement is limited to a small range in all the tested motion sequences. This also shows the stability of the SBD algorithm in each tested motion sequence. The variance of the averaged RMSE measure shows that the error displacement is limited to a small range in all the tested motion

sequences. This also shows the stability of the SBD algorithm in each tested motion sequence. The variance of the averaged RMSE measure shows that the error displacement is limited to a small range in all the tested motion sequences with a maximum variance of 7.37E-05 in 019a063s00L. This also shows the stability of the SBD algorithm in each tested motion sequence.

The low background error percentage is due to the ability of this algorithm to identify and remove shadows along with efficiency in background detection. The algorithm performs well in indoor sequences background and shadow areas but the algorithm fails sometimes to fully extract the motion pixels, which reflect in the higher foreground error.

- **Outdoor Motion Sequences**

For the outdoor sequences, 10 were tested using the SBD algorithm. 48-55 background frames were used for background adaption.

Figure 2.7 shows samples of outdoor extracted images. From the samples shown the algorithm's performance was less successful than when applied to indoor sequences. Motion pixels are not always detected correctly. Holes can be seen on all the presented images in all the different sequences. The problem is more serious in motion sequences 008e013s00L and 009e017s01L where holes are more than in the other two sequences. In addition, the edges of the moving object are not finely extracted.

The shadow detection criteria does not perform as well as in indoor sequences. The shadow is not always removed. The shadow can be seen in the first three shown sequences, 008e013s00L, 009e017s01L and 010e024s00L.

The SBD still performs well in extracting the background: few background pixels erroneously labelled as motion pixels in all the shown samples.

The evaluation of the averaged error of each outdoor sequence is shown in table 2.6. The table shows the RMSE, the PSNR, and the percentage error of the foreground, background and the overall error over the silhouette motion pixels.

Motion sequence 017e055s00R gave the highest RMSE, and the minimum PSNR value. While 018e059s01L motion sequence provided the minimum RMSE and the highest PSNR. For the foreground percentage error, motion sequence 017e055s00R gave the highest percentage error recorded by all the sequences. The range of error here is in the interval of 6.88% to 30.79% with an average of 14.60%. The minimum foreground percentage error was given by 010e024s00L motion sequence. The background percentage error gave a lower average of 2.14% with a range of error between 1.17% to 4.10%. When looking at the percentage of the overall error over the silhouette's motion pixels, the maximum error was as high as 43.26% while the minimum was 16.91%. The average error here is 29.14%. The variance of the RMSE shows that the SBD error on these sequences is consistent since the

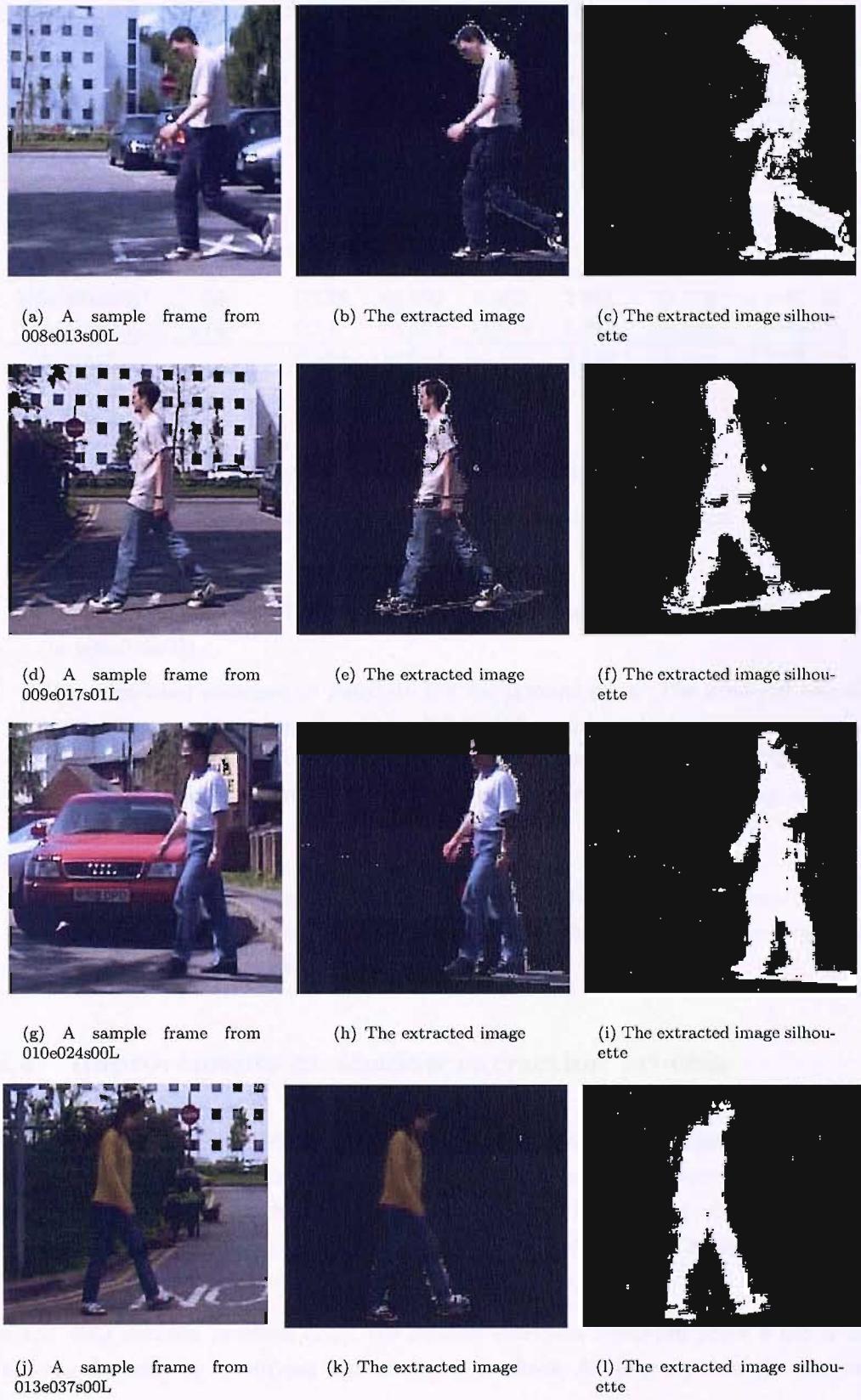


FIGURE 2.7: Examples of outdoor images extracted using the SBD algorithm

Sequence Number	Number of Frames	RMSE	PSNR (dB)	FG Error ¹	BG Error ²	Overall Error ³	σ_{RMSE}^2
008e013s00L	100	0.178	15.127	10.544	2.055	23.893	9.653E-04
009e017s01L	96	0.200	13.986	15.300	2.408	32.146	1.340E-04
010e024s00L	94	0.212	13.500	6.878	4.095	31.768	6.356E-05
013e037s00L	158	0.154	16.297	7.344	1.659	19.020	1.480E-04
013e040s00L	151	0.225	12.965	28.141	1.984	43.036	2.345E-05
017e054s00L	112	0.221	13.164	18.455	2.674	34.840	5.689E-04
017e055s00R	88	0.234	12.638	30.792	1.810	43.257	1.118E-04
018e059s01L	104	0.146	16.700	8.878	1.165	16.909	3.484E-05
018e060s00L	88	0.178	14.999	8.580	2.311	23.174	2.968E-05
019e063s05L	112	0.147	16.663	11.010	1.269	23.398	5.618E-05
Average		0.189	14.604	14.592	2.143	29.144	2.136E-04

1 FG Error Foreground Percentage Error

2 BG Error Background Percentage Error

3 The percentage of the overall error compared to the motion pixels only

TABLE 2.6: An overall assessment on a number of motion outdoor sequences using the SBD algorithm

variance is small in all the motion sequences. The maximum variance is 9.653E-04 in 008e013s00L.

The algorithm managed to maintain low background error. The averaged overall percentage error reaches almost one third of the motion pixels suggests that the algorithm has room for improvement to reduce the foreground labelling error. The foreground region remain the main error of concern with an average error of 14.59%.

The performance analysis clearly suggests that the technique can be improved to reduce errors and remove the defects, especially in the foreground region for both indoor and outdoor motion sequences.

2.4 Improvements on shadow extraction process

To improve the SBD algorithm we tested the performance of each region condition in this algorithm. Figure 2.8 shows the pixel labelling for each region where each region's pixels are labelled with a different colour. Notice that the moving object's holes in the indoor samples are mainly caused by mislabelling motion pixels as shadow. In the outdoor sample small parts of the holes are caused by the shadow.

In the SBD decision function (2.5), the shadow detection condition plays a factor in reducing the error by identifying and extracting shadows. At the same time the shadow

detection criterion mislabels some motion pixels as shadows which causes holes to appear in the moving object.

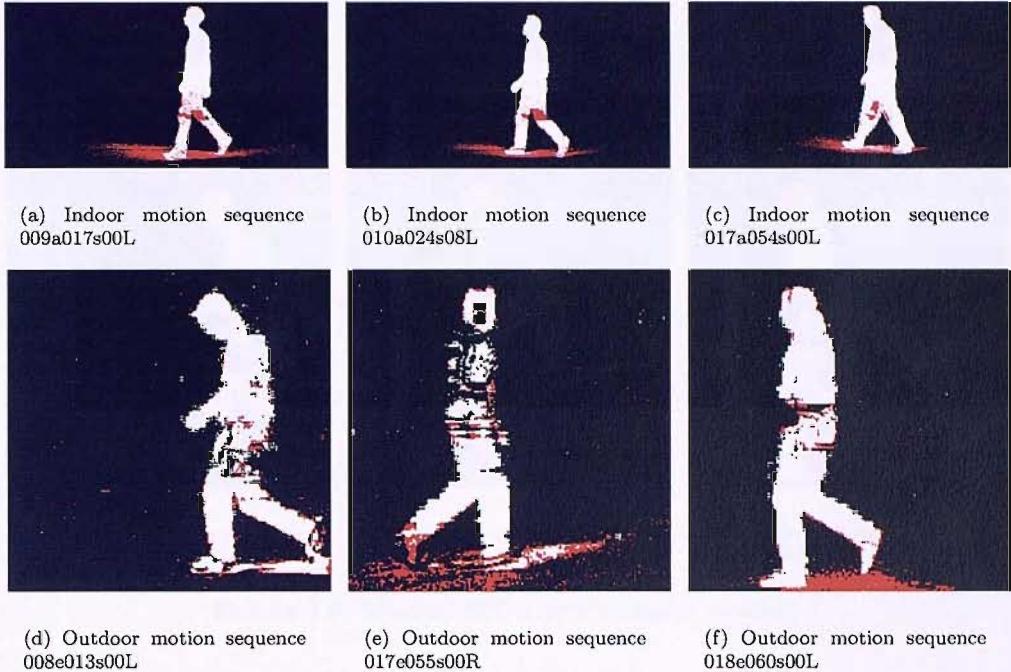


FIGURE 2.8: SBD algorithm region labelling

To optimise this process a threshold distance between the background mean and a virtual border for the shadow class is determined. The border is drawn at the point where the distance from the tested pixels to the mean gives the minimum overall error. In this process a training set of N frames is used. Then a search is performed by incrementing a parameter, $MULTIP$ in the border function

$$\text{Shadow.Border} = \text{MULTIP} \times \sqrt{\sigma_R^2 + \sigma_G^2 + \sigma_B^2}. \quad (2.14)$$

Any shadow pixel with a distance exceeding the shadow border will be considered as a motion pixel.

The result of this improvement is shown in the images of Figure 2.9. The improvements are marked in green and the shadow in red. In the indoor sequences the holes are partially filled in the sequences. The improvement has filled half of the large holes and most of the small holes in the sample of 009a017s00L and 010a024s08L. Part of the large hole in 017a054s00L is also filled. In sample 010a024s08L, in reducing the errors of the moving object the new condition has added more error in the shadow region. The outdoor sequences shows only slight improvement on the shown sequences (only small green pixels can be seen filled in the moving object).

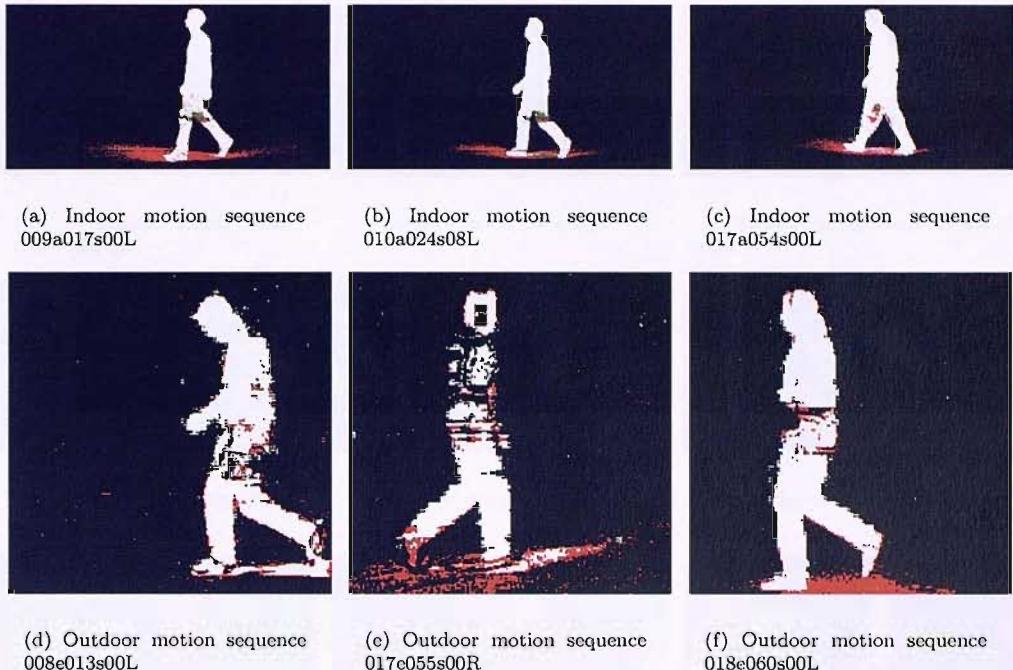


FIGURE 2.9: Modified SBD algorithm region labelling

Samples of extracted indoor motion sequences are shown in Figure 2.10. comparing these samples with the original SBD samples, Figure 2.6, motion sequences 008a013s00L, 009a017s00L, and 010a024s08L results improved (the large holes are smaller also some of the scattered small holes had disappeared). The outdoor motion sequences are shown in Figure 2.11 without much noticeable improvement when comparing to the outdoor results of the original SBD, Figure 2.7.

Table 2.7 shows the effect of function (2.14) on indoor sequences errors. From the table we can see that the modification implemented has reduced the error slightly. The averaged RMSE, the foreground percentage , and overall percentage errors have dropped slightly. The averaged PSNR has also increased slightly. The averaged percentage background error is the only result were the error has slightly increased. The RMSE variance is small in all tested sequences with a maximum value of 6.859E-05 in 013a040s00L. This means that the algorithm is fairly consistent in its performance in each motion sequence.

Table 2.8 shows the result of the modifications on outdoor sequences. When comparing the result with the original SBD algorithm outdoor table, the RMSE is similar. The PSNR values have slightly increased. The averaged foreground error have slightly dropped from 14.59% to 14.45% while the averaged background error has slightly increased. The overall averaged error has slightly dropped from 29.14% to 29.08%. Though we are testing outdoor sequences, the algorithm shows a steady performance error wise with a range of variance between 2.272E-05 and 9.485E-04.

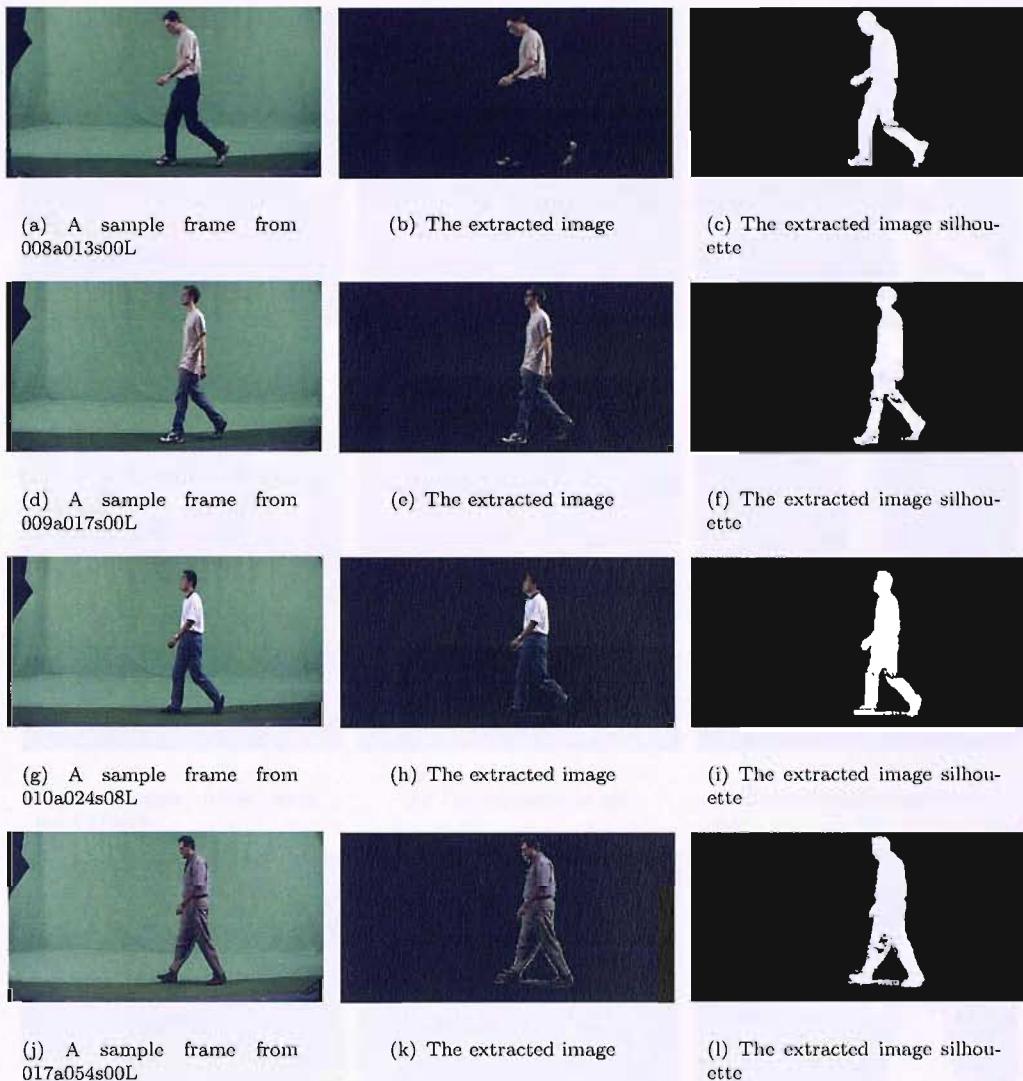


FIGURE 2.10: Examples of indoor images extracted using the modified SBD algorithm

As an overall assessment we can say that the effect of this modification is more apparent on the indoor sequences while on outdoor sequence such modification has only a slight positive effect.

2.5 Conclusions

This chapter has presented the Statistical Background Disturbance (SBD) algorithm. The algorithm was tested on indoor and outdoor motion sequences. The assessment of the motion sequences showed the problems of the algorithm in indoor and outdoor sequences. The SBD suffers from holes which appear erroneously in the moving subjects. On the outdoor data the algorithm fails sometimes to extract the shadows. The shadow

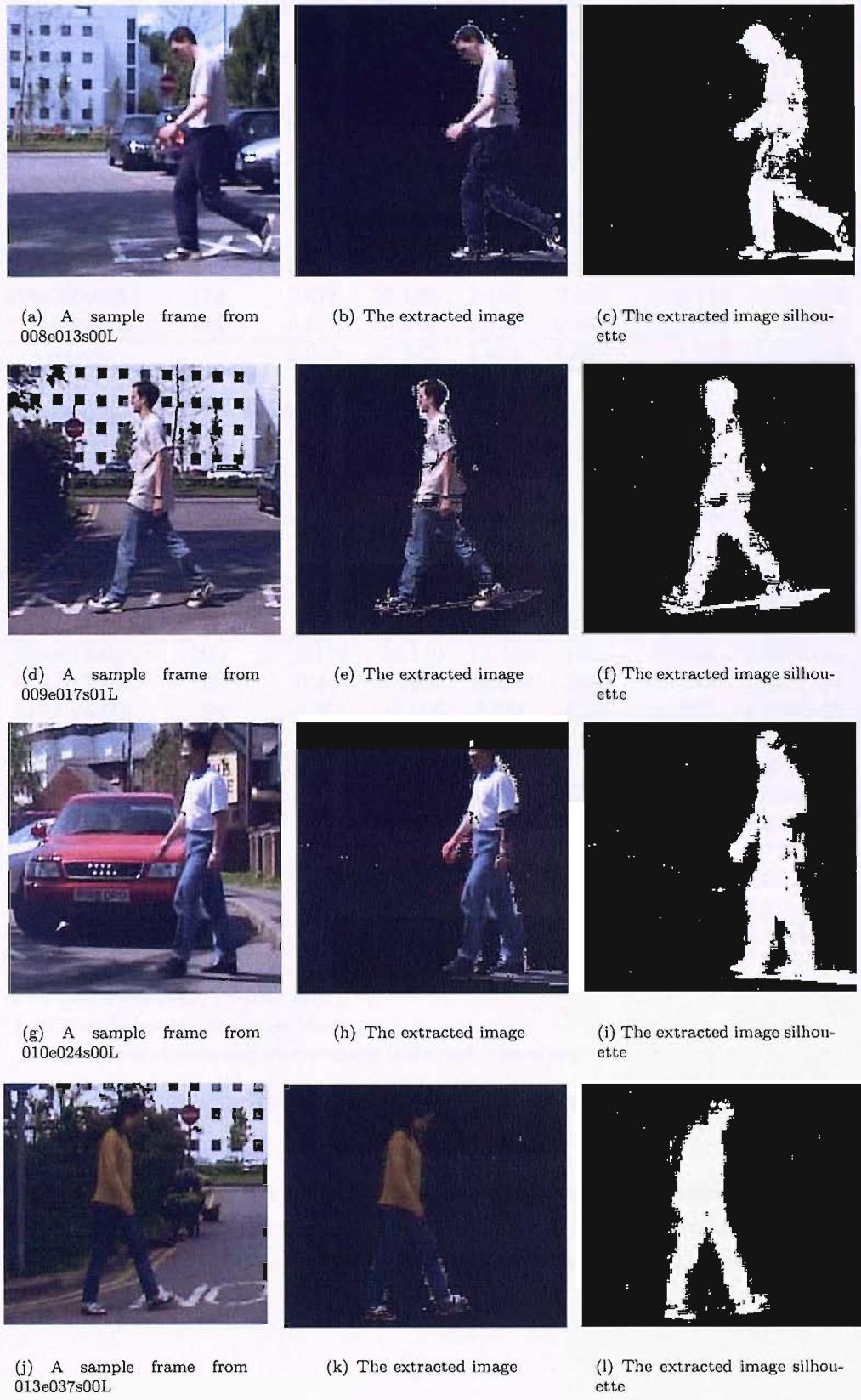


FIGURE 2.11: Examples of outdoor images extracted using the modified SBD algorithm

Sequence Number	Number of Frames	RMSE	PSNR (dB)	FG Error ¹	BG Error ²	Overall Error ³	σ_{RMSE}^2
008a013s00L	178	0.076	22.430	2.943	0.392	7.853	4.860E-05
009a017s00L	169	0.076	22.468	2.383	0.448	8.923	3.325E-05
010a024s08L	187	0.077	22.285	2.581	0.464	9.520	3.487E-05
013a037s00L	114	0.065	23.738	4.648	0.193	8.104	3.824E-05
013a040s00L	184	0.078	22.167	2.745	0.484	10.395	6.859E-05
017a054s00L	188	0.118	18.553	0.955	1.438	21.095	5.438E-05
017a055s00R	162	0.070	23.151	2.419	0.357	7.629	2.648E-05
018a059s00L	188	0.069	23.317	3.125	0.284	7.190	2.854E-05
018a060s00L	179	0.078	22.149	2.450	0.483	9.149	2.880E-05
019a063s00L	186	0.070	23.162	2.345	0.367	7.894	4.293E-05
Average		0.078	22.342	2.659	0.491	9.775	4.047E-05

1 FG Error Foreground Percentage Error

2 BG Error Background Percentage Error

3 The percentage of the overall error compared to the motion pixels only

TABLE 2.7: Overall assessment on a number of motion indoor sequences using the improved SBD algorithm

Sequence Number	Number of Frames	RMSE	PSNR (dB)	FG Error ¹	BG Error ²	Overall Error ³	σ_{RMSE}^2
008e013s00L	100	0.178	15.110	11.539	1.913	23.969	9.485E-04
009e017s01L	96	0.200	13.986	15.300	2.408	32.146	1.340E-04
010e024s00L	94	0.211	13.506	6.757	4.107	31.721	6.167E-05
013e037s00L	158	0.154	16.306	7.211	1.674	18.991	1.544E-04
013e040s00L	151	0.225	12.969	27.911	2.009	42.992	2.272E-05
017e054s00L	112	0.221	13.170	17.953	2.749	34.797	5.676E-04
017e055s00R	88	0.233	12.645	30.655	1.821	43.192	1.110E-04
018e059s01L	104	0.145	16.808	7.937	1.243	16.502	3.771E-05
018e060s00L	88	0.178	15.014	8.409	2.326	23.095	3.045E-05
019e063s05L	112	0.147	16.669	10.860	1.281	23.363	5.770E-05
Average		0.189	14.618	14.453	2.153	29.077	2.126E-04

1 FG Error Foreground Percentage Error

2 BG Error Background Percentage Error

3 The percentage of the overall error compared to the motion pixels only

TABLE 2.8: Overall assessment on a number of motion outdoor sequences using the modified SBD algorithm

decision criteria in the SBD algorithm was improved by adding a second condition to test the distance of the shadow labelled pixel to the background mean of the tested pixel. The improvement in error reduction was more apparent in the indoor rather than the outdoor motion sequences.

Chapter 3

Mixture of Gaussians

3.1 Introduction

In density estimation a commonly used approach is to represent the probability density in a functional form, which consists of a number of adjustable parameters. Then the values of the parameters are optimised to model the density of the data. A simple density estimation model is the *Gaussian* distribution (also called the *normal distribution*).

Univariate Gaussian distributions can be used for univariate data (grey levels of images for instance), for colour images (multi-value data) multivariate Gaussians can be used.

Gaussian distributions are useful in practice for two reasons: first, the normal distribution serves as a bona fide population model in some instances; second, the sampling distributions of many multivariate statistics are approximately normal, regardless of the form of the parent population (Johnson and Wichern, 2002).

A Mixture of Gaussians can be used to model the density of single or multivariate data when the density is more complex, e.g. bimodal. In addition, a mixture of Gaussians model is a very appealing approach to data fitting as it scales favourably with dimensionality of the data, has good analytic properties and many data sets form clusters which are approximately Gaussian in nature (Roberts et al., 1998).

Mixture models in general provide powerful techniques for density estimation. In the remainder of this section we will provide a recent history of the use of Gaussians models in the motion estimation area.

Pfinder (Wren et al., 1997; Maes et al., 1997) used a multi-scale statistical model of colour and shape with a single Gaussian per pixel to model the background. The algorithm succeeded in finding a 2-D representation of head, hands and feet locations of a moving human subject. In contrast, Friedman and Russell (1997) took a simpler approach to modelling the statistical nature of the image by using a mixture of Gaussians with a

single distribution to model the whole of the background and two other distributions to model the variability in shadows and moving objects.

In real situations the background is typically multi-modal. A single Gaussian would suffice to approximate the background if each pixel resulted from a single surface under fixed lighting. Often multiple surfaces appear on a particular background pixel and the lighting conditions change. Multiple adaptive Gaussians can cope with such situations and thus they are a suitable solution to model multi-modal backgrounds.

Traven (1991) introduced a stochastic on-line technique to optimise the parameters of a Gaussian mixture model. An extensive simulation study was presented in (Cwik, 1996). Stauffer and Grimson (2000, 1999); Grimson et al. (1998) used the online mixture of Gaussians technique for motion estimation and tracking. The persistence and the variance of each of the Gaussians is used to identify background distributions. The approach was designed to deal robustly with bimodal backgrounds, lighting changes, repetitive motions of scene elements. The method lacks the capability to remove shadows and highlights. The method was extended to be used for a moving camera in (Mittal and Huttenlocher, 2000). Javed et al. (2002); Javed and Shah (2002) used the same updating parameters along with a gradient based modelling scheme. The same principle of Stauffer was applied in an on-line EM algorithm (KaewTraKulPong and Bowden, 2001a,b) and combined with Horprasert et al. (1999) for shadow extraction. McKenna et al. (1999, 1998) used also an adaptive EM algorithm with an HSI colour representation to track a multi-colour moving object. Elgammal et al. (2002, 2000) used a Gaussian density estimator as a kernel in the process of background modelling. The final background model is updated by combining a short and a long term model of the background. A colour representation was used to suppress shadows.

3.2 Gaussian Density Function

The Gaussian density function for a single variable is presented in the form

$$p(x) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad (3.1)$$

where μ is the mean, and σ^2 is the variance, and the function is normalised such that

$$\int_{-\infty}^{\infty} p(x)dx = 1 \quad (3.2)$$

For d -dimensions the general multivariate Gaussian density function is written

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp \left(-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right) \quad (3.3)$$

where $\boldsymbol{\mu}$ is a d -dimensional vector and Σ is a $d \times d$ covariance matrix

3.3 Mixture Model

The principle here is to model the distribution of a candidate as a mixture of Gaussian densities. The density of the input data is modelled as follows:

$$p(\mathbf{x}) = \sum_{j=1}^K p(\mathbf{x} | j) P(j) \quad (3.4)$$

where $p(\mathbf{x} | j)$ are the component densities and $P(j)$ are the priors. The priors are chosen to satisfy the constraints

$$\sum_{j=1}^K P(j) = 1 \quad (3.5)$$

and

$$0 \leq P(j) \leq 1 \quad (3.6)$$

The component density functions, $p(\mathbf{x} | j)$, are normalised so that

$$\int p(\mathbf{x} | j) d\mathbf{x} = 1, \quad j = 1 \dots K \quad (3.7)$$

An important property of such mixture models is that, for many choices of component density function, they can approximate any continuous density to arbitrary accuracy provided the model has a sufficiently large number of components, and provided the parameter of the model are chosen correctly (Bishop, 1996).

3.4 Stauffer mixture of Gaussians algorithm (MOG)

This approach (Stauffer and Grimson, 2000, 1999; Grimson et al., 1998) models the background with independent distributions that are updated by an efficient on-line method. Figure 3.1 gives an overall view of how this algorithm operates. The model applies a

pixel-by-pixel process. A pixel is scalar for a grey pixel and a vector for a colour pixel. The recent history of each pixel is modelled as a mixture of K Gaussian distributions. For coloured (RGB) pixels, the probability of a pixel intensity, $\mathbf{x} = (x_R, x_G, x_B)$,

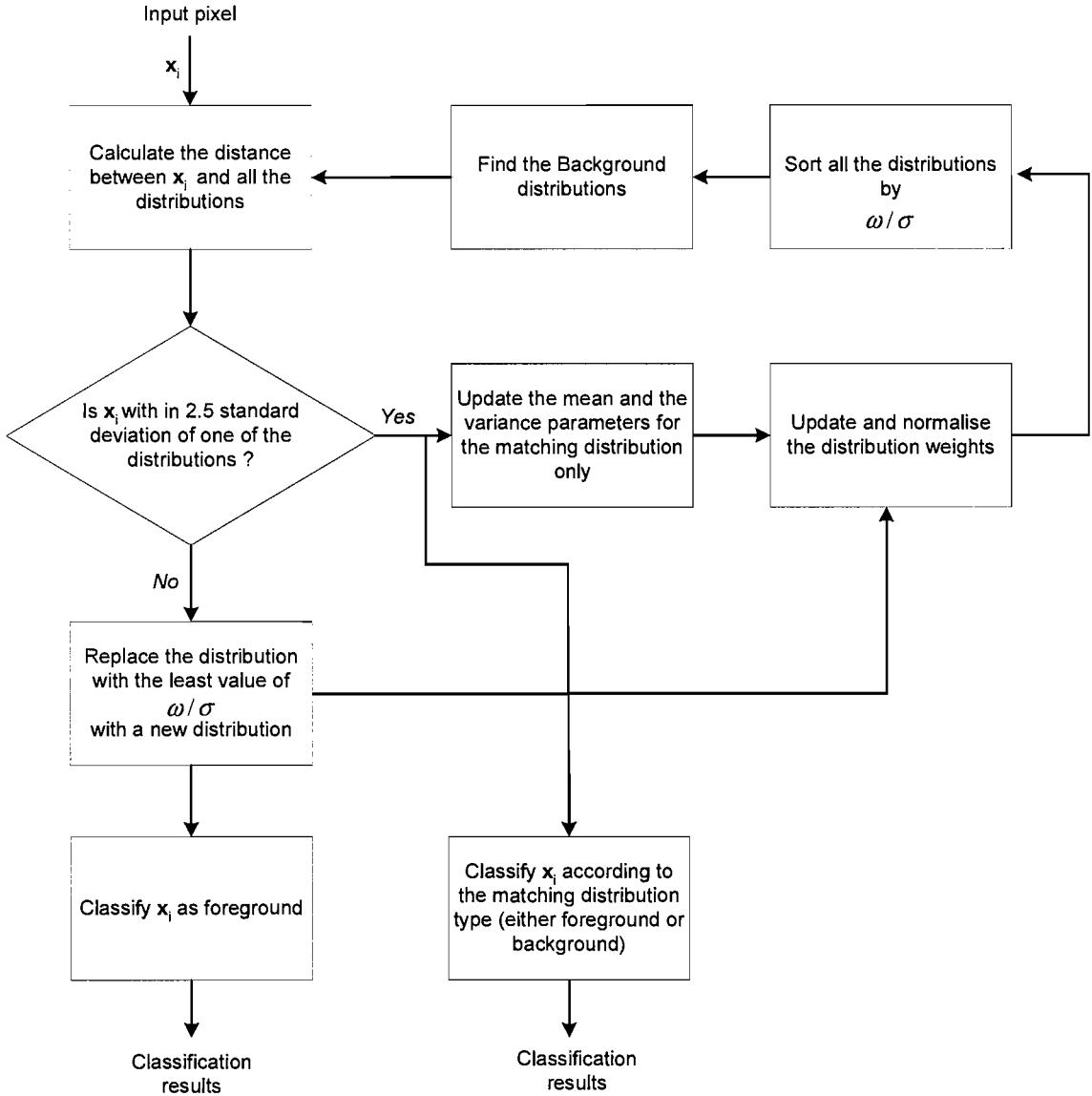


FIGURE 3.1: An overall view of how the mixture of Gaussians algorithm operates.

$$p(\mathbf{x}) = \sum_{j=1}^K w_j \eta(\mathbf{x}, \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) \quad (3.8)$$

where K is the number of distributions (a value 3 to 5 is often used). The number of distributions is supposed to be chosen so as to give a reasonable fit to the true density.

w_j is the weight estimate for the j^{th} distribution, μ_j is the mean value for the j^{th} distribution, Σ_j is the covariance matrix for the j^{th} distribution, and it is assumed to be of the form $\Sigma_j = \sigma_j^2 \mathbf{I}$ (to simplify computation). η is a Gaussian probability density function formed from the multivariate Gaussian

$$\eta(\mathbf{x}, \mu_j, \sigma_j) = \frac{1}{(2\pi\sigma_j^2)^{d/2}} \exp\left(-\frac{\|\mathbf{x} - \mu_j\|^2}{2\sigma_j^2}\right) \quad (3.9)$$

where d is the input dimension which is 3 for the (RGB) colour model.

The algorithm implements an on-line K -means approximation method. Every new pixel value, \mathbf{x} , is compared to the existing K Gaussian distributions. The pixel is classified to be in a particular distribution if the pixel is within 2.5 times the standard deviation of the distribution (this number is chosen to make the probability of detection for a distribution data equals 0.99, i.e 99%). The pixel is checked against the background distributions first and then to the foreground distributions.

The distributions are ordered according to the ratio of the weight over the standard deviation of each distribution, w_j/σ_j . This process will rank the most probable (those with high weight and low variance) to the least probable background distributions (those with low weight and high variance). The background model is formed from a number of background distributions

$$B = \arg \min_b \left(\sum_{j=1}^b w_j > T \right) \quad (3.10)$$

where $T \in [0, 1]$ controls the number of modes of variations in the background. A small value for T will result in a strict background where only unimodal background surfaces are accepted. In contrast, a large value for T will enable the system to allow bimodal backgrounds (sea surface, moving trees, etc) to be considered as part of the background.

If a pixel does not match any of the K distributions, the pixel will be considered as a new distribution replacing the distribution with the smallest w_j/σ_j . The new distribution mean, $\mu_{j,t}$, will be the pixel value. The distributions with small w_j/σ_j are presumed to be non-background pixels. The distributions are meant to model the background. Allocating a Gaussian for non-background pixels makes the model adaptive to any change in the background model (like adding an item to the scene). Now if a non-background pixel (part of a moving object) does not move over a period of time, its distribution weight over time will increase and its variance will decrease until this distribution becomes part of the background model.

The prior weight of a new distribution will be set to a low weight and the variance to a high variance. After evaluating a new pixel, the K distributions prior weights are updated at time t

$$w_{j,t} = (1 - \alpha)w_{j,t-1} + \alpha M_{j,t} \quad (3.11)$$

where α is the learning rate. $M_{j,t}$ is 1 for the matching distribution, and 0 for the remaining distributions. The weights are normalised after this process. The values of $\mu_{j,t}$ and $\sigma_{j,t}^2$ are updated only for the matching distribution

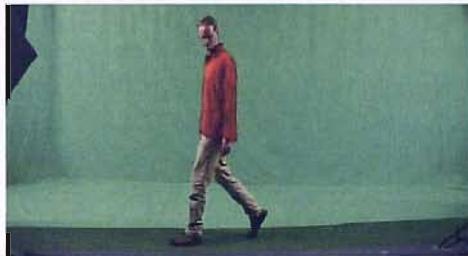
$$\mu_{j,t} = (1 - \rho)\mu_{j,t-1} + \rho \mathbf{x}_t \quad (3.12)$$

$$\sigma_{j,t}^2 = (1 - \rho)\sigma_{j,t-1}^2 + \rho(\mathbf{x}_t - \mu_{j,t})^T(\mathbf{x}_t - \mu_{j,t}) \quad (3.13)$$

where

$$\rho = \alpha \eta(\mathbf{x}_t, \mu_{j,t-1}, \sigma_{j,t-1}) \quad (3.14)$$

A pixel is identified as a motion pixel if it matches a non-background distribution or if it does not match any of the available distributions of a pixel. All the motion pixels are tested using a connected component algorithm (Horn, 1986) to remove the noise pixels. The Stauffer Mixture of Gaussians algorithm was implemented on images of indoor and outdoor scenes. Samples of indoor and outdoor extracted images are shown in Figures (3.2 - 3.3)(the connected components process is not applied on these frames).



(a) An input frame



(b) Motion extraction result

FIGURE 3.2: Indoor image shows an input frame and an extracted moving subject with Stauffer mixture of Gaussians algorithm.

The shadow in the extracted moving object is a concern that has to be rectified to improve the extraction of this method. Noise pixels exist in the background area. Stauffer and Grimson (2000, 1999) used the connected components method to remove components of two or even single pixels but this process does not assure the removal of all the background noise.

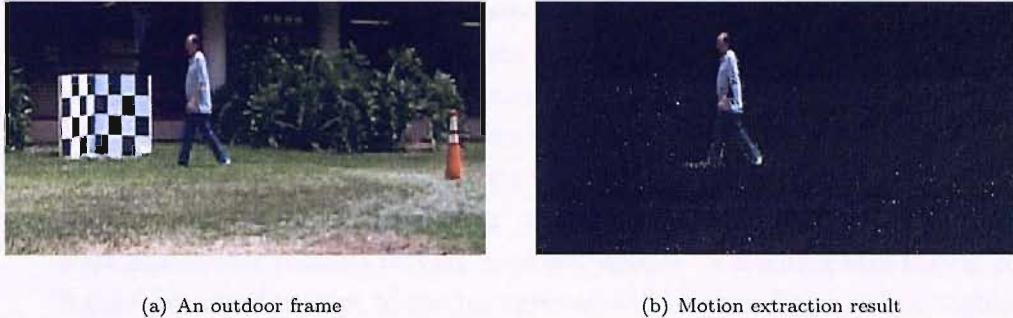


FIGURE 3.3: Stauffer mixture of Gaussians algorithm used to extract an outdoor frame from the HumanID database (Phillips et al., 2002)

3.4.1 Parameter Settings

The setting of the parameters in this algorithm is very important. The effects of each parameter will be illustrated using an indoor sequence. Then the same process will be repeated for an outdoor sequence. This is because parameters settings are dependent on the motion sequence scene, i.e. if another indoor/outdoor sequences were used with a different background scene the parameters should be reoptimised for the new scene. The parameters are as follows: the learning rate, α ; the initial weight; the background threshold, T ; the initial variance; and different number of Gaussians per mixture. The same initial parameter setting is used for indoor and outdoor data unless stated otherwise. Essentially the effect of each parameter is assessed independently aiming to determine an optimal set for later use. The assessment was performed without including the shadow part of the images. This was done due to the fact that the algorithm does not support shadow extraction and including such pixels will give wrong indications by adding more pixels to the false positive region. Also for some parameters varying its values while including shadow pixels will give mistakenly better results in the overall extraction result while in fact the motion extraction deteriorates (such an effect was noticed clearly in initial setting of the variance parameter).

- **The learning rate (α)**

The initial learning rate can be set between [0-1]. The effect of changing the learning rate will be shown by testing a variable learning rate on an indoor motion sequence. The test was performed with the initial weight set to 0.05, the background threshold set to 0.6, the initial variance set to 50 and 5 Gaussians per mixture.

Using different learning rates, we will be testing how long the mixture of Gaussians algorithm takes to adapt to the background. Also the performance of the algorithm in motion detection will be measured.

For the tested sequences, having fast adaptation of (2 - 5) frames will not be suitable since a moving object surface (with a human movement speed) might take such time to finish passing through a pixel i.e. the background mixtures with fast learning rates might replace the background with the moving object surface colour and use it as a background. Also we need more frames to allow more time for a moving object, e.g. a person, to stop for a short time or at least to consider persons moving in slower speeds. A learning rate taking 20 frames or more to adapt to the background will be considered as reasonable.

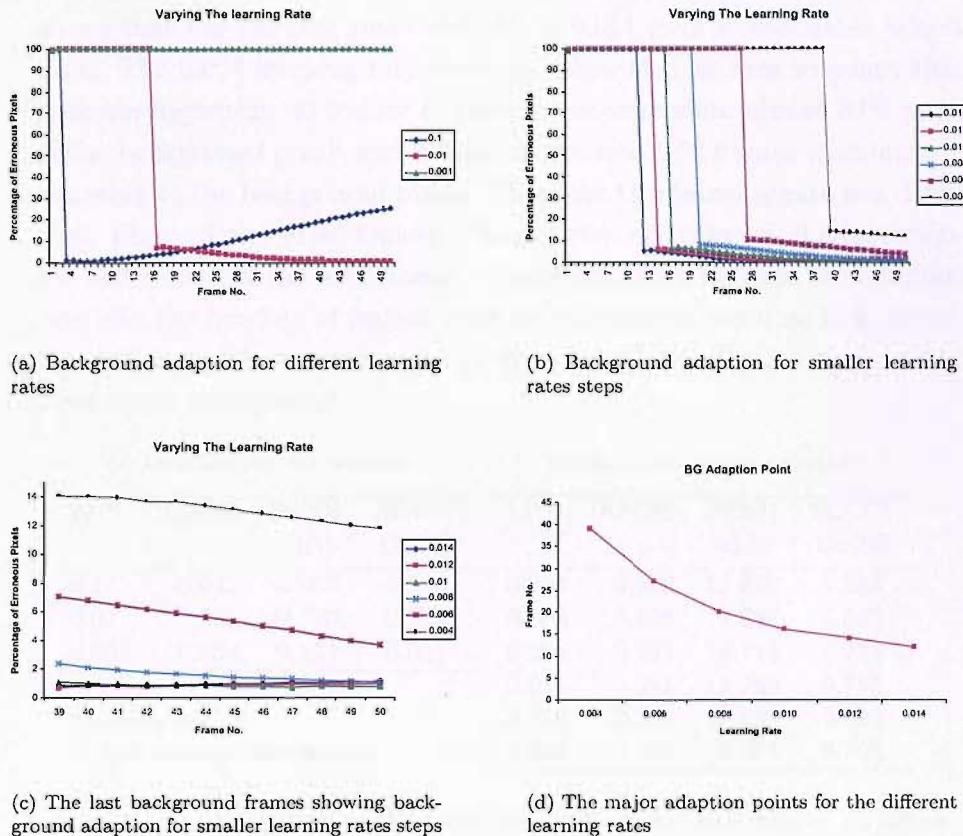


FIGURE 3.4: The effect of varying the learning rate on background adaption

In the beginning the learning rates: 0.1, 0.01 and 0.001 were used. 50 background frames were used to test how fast the algorithm will adapt to the background. Figure 3.4(a) shows the effect on using different learning rates on background adaptation. As it shows using 0.1 gave a very fast adaptation in four frames and the system was not robust in maintaining the background model since the error started to rise later. Therefore a learning rate with a 0.1 value is not suitable. On the other extreme using a very slow learning rate of 0.001 led to all 50 frames passing without the system adapting to the background. In this figure the optimum learning rate for our criteria is 0.01

which took 16 frames to adapt to more than 90% of the background and took another 15 frames to handle most of the remaining background pixels. The system with this algorithm continued robustly in containing the background pixels as we can see in frames 32 to frame 50.

Since the gaps between these learning rates (0.1, 0.01, and 0.001) are large, the effect of changing the learning rate in smaller steps around the 0.01 learning rate was investigated. The same test was repeated on 50 background frames with the following values for the learning rate: 0.004, 0.006, 0.008, 0.010, 0.012 and 0.014. The result is shown in Figure 3.4(b). The figure shows that the learning rates of 0.006 to 0.014 gave a reasonable adaption time. The 0.004 learning rate made the algorithm so slow to adapt that it took the algorithm 40 frames to start to accommodate almost 84% percent of the background pixels and it finished the whole 50 frames without totally adapting to the background pixels. Therefore this learning rate was disqualified. Figure 3.4(c) gives a more enlarged view of background adaptation on the last frames of the background. The relationship between the adaptation rate and the number of frames gave an exponential trend as it is shown in Figure 3.4(d). The figure is a plot of the point where adaption accommodates most of the background.

(a) Large learning rate variations				(b) Small learning rate variations			
LR*	RMSE	PSNR (dB)	ROC** Cutoff	LR*	RMSE	PSNR (dB)	ROC** Cutoff
0.1	0.642	3.861	0.328	0.004	0.238	12.496	0.938
0.01	0.261	11.732	0.753	0.006	0.196	14.267	0.845
0.001	0.964	0.323	0.001	0.008	0.233	12.714	0.772
				0.01	0.261	11.732	0.753
* Learning Rate				0.012	0.286	10.932	0.730
** ROC Optimal Cutoff Measure				0.014	0.307	10.332	0.705

TABLE 3.1: Assessment tests on the MOG using different learning rates on an indoor motion sequence

To choose a specific value of adaptation among the values shown in Figure 3.4(b), the algorithm was tested with an indoor motion sequence after the first 50 background frames. The motion sequence with the different learning rates was analysed with three different assessment tests the RMSE, the PSNR and the ROC optimal cutoff measure (Grzymala-Busse et al., 2003, 2002), Table 3.1. Table 3.1(a) shows the large variation of learning rates. The 0.01 learning rates showed better results in all the three measurements, where it gives the minimum error for the RMSE, the largest PSNR and maximum ROC cutoff value. Looking on smaller variation close to 0.01, Table 3.1(b), 0.006 gives the smallest RMSE, the largest PSNR value and the second best

cutoff value after the excluded 0.004 rate. Therefore 0.006 will be used as the learning rate for the MOG algorithm on indoor sequences.

(a) Large learning rate variations				(b) Small learning rate variations			
LR*	RMSE	PSNR (dB)	ROC** Cutoff	LR*	RMSE	PSNR (dB)	ROC** Cutoff
0.1	0.561	5.028	0.329	0.004	0.454	6.857	0.683
0.01	0.349	9.155	0.623	0.006	0.379	8.444	0.725
0.001	0.921	0.714	0.002	0.008	0.364	8.788	0.625
* Learning Rate				0.01	0.349	9.155	0.623
** ROC Optimal Cutoff Measure				0.012	0.345	9.261	0.621
				0.014	0.345	9.271	0.613

TABLE 3.2: Assessment tests on the MOG using different learning rates on an outdoor sequence

For outdoor sequences, testing the learning rate with the large variations resulted in Table 3.2(a). The result also shows the MOG with 0.01 as the learning rate performs better than the 0.1 and 0.001 in all the assessment measures used, RMSE, PSNR and ROC optimal cutoff measure. This result led us to test the values around 0.01 with small variation steps which resulted in Table 3.2(b). The assessment measures does not agree on a single value, but when searching for a rate that performed better in most of the measures we find that 0.012 and 0.014 performed better in the RMSE and the PSNR measures. These two values for the learning rate gave the same value for the RMSE. The 0.012 performs better in the ROC optimal cutoff measure while 0.014 performs better in the PSNR measure. Since 0.012 and 0.014 are almost equal in preference, we will choose 0.012 since it is slightly slower (slower learning rate means more persistent background model).

- **Initial weight**

The initial weight can be set to any value larger than zero. We will discover the effect of changing the initial weight through testing different initial weights on an indoor sequence. The test was performed with the learning rate set to 0.01, the background threshold set to 0.6, the initial variance set to 50 and 5 Gaussians per mixture.

The indoor motion sequence assessment result is shown in Table 3.3. The decrease of the initial weight resulted in a decrease in the RMSE and an increase in the PSNR value. 0.0005 gave a similar result to the values of 0.005 initial weight but slightly better. 0.0005 initial weight gave the best result in RMSE and in the PSNR. In the ROC optimal cutoff measure all the three 0.05, 0.005 and 0.0005 gave similar result with minor differences. Due to the good results accomplished by 0.0005 initial weight in the RMSE

the PSNR and a reasonable result in the ROC optimal cutoff measure, it was used as the initial weight.

Initial Weight	RMSE	PSNR (dB)	ROC* Cutoff
0.5	0.307	10.259	0.650
0.05	0.261	11.732	0.753
0.005	0.250	12.105	0.748
0.0005	0.249	12.152	0.749

* ROC Optimal Cutoff Measure

TABLE 3.3: Assessment tests on the MOG using different initial weights for an indoor motion sequence

The tests on the outdoor motion sequence is given in Table 3.4. The table shows that the variation of the initial weight is not significant on outdoor motion sequences (i.e. when the parameter is changed the output results is not affected). The initial weight will be set to be the same as the indoor sequence initial weight, 0.0005

Initial Weight	RMSE	PSNR (dB)	ROC* Cutoff
0.5	0.349	9.155	0.623
0.05	0.349	9.155	0.623
0.005	0.349	9.155	0.623
0.0005	0.349	9.155	0.623

* ROC Optimal Cutoff Measure

TABLE 3.4: Assessment tests on the MOG using different initial weights on an outdoor motion sequence

- **The background threshold (T)**

The background threshold, T , can have a value of $0 < T \leq 1$. The smaller the value of T , the more chance for more Gaussians to be considered as part of the background model. A larger value of T makes the system being able to accommodate multi-modal backgrounds, such as a waving flag, the sea tide or tree leaves moving with the wind. Smaller values of T will allow only few Gaussians to be in the background model. The effect of changing the background threshold will be shown through testing different threshold values on an indoor and an outdoor motion sequence. The test was performed with the learning rate set to 0.01, the initial weight set to 0.05, the initial variance set to 50 and 5 Gaussians per mixture. Even though T can be varied between 0 to 1, we will use 0.2 as a minimum value for our tests. Since we are using

5 Gaussians, if we assume a similar priority in the beginning a persistent background Gaussian should be of a value above 0.2.

T^*	RMSE	PSNR	ROC**
	(dB)		Cutoff
0.2	0.115	18.774	0.982
0.4	0.175	15.212	0.950
0.6	0.261	11.732	0.753
0.8	0.268	11.507	0.736

* Background Threshold

** ROC Optimal Cutoff Measure

TABLE 3.5: Assessment tests on the MOG using different background thresholds for an indoor motion sequence

Table 3.5 shows an assessment for the background threshold parameter variation on the MOG algorithm using an indoor sequence. The performance is inversely proportional to the value of the background threshold. We can see as we decrease the value of T the error decreases. 0.2 threshold gives the smallest RMSE, the largest PSNR and the largest ROC optimal cutoff measure. Thus 0.2 was used as a value for T .

T^*	RMSE	PSNR	ROC**
	(dB)		Cutoff
0.2	0.265	11.547	0.840
0.4	0.295	10.607	0.819
0.6	0.349	9.155	0.623
0.8	0.334	9.535	0.606

* Background Threshold

** ROC Optimal Cutoff Measure

TABLE 3.6: Assessment tests on the MOG using different background thresholds

Table 3.6 gives the assessment for varying the background threshold on an outdoor motion sequence. The minimum RMSE, the maximum PSNR and ROC optimal cutoff measure is scored by the 0.2 background threshold. This value will be used as a background threshold for the outdoor motion sequences.

• Initial variance

This parameter will be used as an initialisation variance for any new Gaussian. The value is supposed to be large enough to accommodate a normal background pixel variation. The effect of changing the initial variance will be shown through testing variable initial variances on an indoor and outdoor motion sequences. The test is performed with the learning rate set to 0.01,

the initial weight set to 0.05, the background threshold set to 0.6 and 5 Gaussians per mixture.

(a) Large Initial Variance variations				(b) Small Initial Variance variations			
Init. Var.*	RMSE (dB)	PSNR	ROC**	Init. Var.*	RMSE (dB)	PSNR	ROC**
5	0.817	1.754	0.278	10	0.462	6.725	0.744
50	0.261	11.732	0.753	30	0.293	10.731	0.793
500	0.247	12.189	0.418	50	0.261	11.732	0.753
				70	0.248	12.173	0.718
				90	0.238	12.500	0.690
				110	0.229	12.823	0.666

* Initial Variance
** ROC Optimal Cutoff Measure

TABLE 3.7: Assessment tests on the MOG using different Initial Variances for an indoor motion sequence

The assessment table for the initial variance, Table 3.7, gives two assessment stages on the same parameter. The first stage is done with large increments in initial variance starting with the following variances 5, 50, 500. The 500 initial variance gave the best result in two of the the three assessment measures (RMSE and PSNR). Such a value might make the acceptable background range too wide which might endanger the accuracy of the algorithm (i.e. if a moving object with a colour similar to the background might be labelled as background). This will be the case until the background distribution converges to its proper size but with such a large variance this might take a long time. So the test was done again starting from variance 10 onwards in steps of 20. The test aimed to find a value smaller than the 500 variance with comparable performance. The test was successful where the initial variance, 110 gave even better performance in all the assessment measures than the 500 initial variance. Therefore 110 was used as an initial variance for the indoor sequences.

(a) Large Initial Variance variations				(b) Small Initial Variance variations			
Init. Var.*	RMSE (dB)	PSNR	ROC**	Init. Var.*	RMSE (dB)	PSNR	ROC**
5	0.761	2.370	0.283	10	0.624	4.098	0.470
50	0.349	9.155	0.623	30	0.403	7.906	0.641
500	0.376	8.496	0.304	50	0.349	9.155	0.623
				70	0.337	9.467	0.587
				90	0.335	9.500	0.549
				110	0.337	9.469	0.517

* Initial Variance
** ROC Optimal Cutoff Measure

TABLE 3.8: Assessment tests on the MOG using different initial variances on an outdoor motion sequence

Table 3.8 gives the evaluation tables for two different tests for changing the initial variance for an outdoor motion sequence one using the large steps of 5, 50 and 500 on Table 3.8(a), and the other using small steps (20 each) starting from 10 and reaching up to 110, Table 3.8(b). In Table 3.8(a) we can see the initial variance of 50 giving the best RMSE, PSNR and ROC optimal cutoff measure. In Table 3.8(b) smaller steps were used near the best performing initial variance, 50, to find an optimal initial variance. The 90 initial variance managed to score the best rate in two of the three evaluation measures, the RMSE and PSNR, and therefore it was used as the initial variance for the outdoor motion sequences.

- **Number of Gaussians per mixture**

The number of Gaussians was varied from 2 to 9 Gaussians per pixel to test the effect on the system performance. Bearing in mind that the more Gaussians used, the more the speed performance of the system will degrade. Also we started with two Gaussians (not one) where one will be used for the background model and the other is used for motion pixels. The test was performed with the learning rate set to 0.01, the initial weight set to 0.05, the initial variance set to 50 and the background threshold set to 0.2. Here we used a small background threshold, 0.2, because it was chosen as a threshold for indoor and outdoor sequences. Also by testing using this threshold we will make sure that using such a small threshold will not hinder multiple of Gaussians from building the background model and decrease the error.

(a) Changing number of Gaussians (2 - 5)				(b) Changing number of Gaussians (5 - 9)			
No. of Gauss.*	RMSE	PSNR (dB)	ROC**	No. of Gauss.*	RMSE	PSNR (dB)	ROC**
2	0.474	6.484	0.576	5	0.115	18.774	0.982
3	0.161	15.861	0.971	7	0.124	18.141	0.958
4	0.119	18.521	0.983	9	0.193	14.366	0.784
5	0.115	18.774	0.982				

* Number of Gaussians

** ROC Optimal Cutoff Measure

TABLE 3.9: Assessment tests on the MOG algorithm using different number of Gaussians per pixel for an indoor motion sequence

The evaluation of using different number of Gaussians on an indoor motion sequence is shown in Table 3.9. Using 5 mixture of Gaussians gives the best result in the RMSE and the PSNR. Also in the ROC optimal cutoff measure 5 Gaussians scored very close to the highest value. Accordingly, 5 Gaussians will be used for extracting indoor motion sequences.

(a) Changing number of Gaussians (2 - 5)				(b) Changing number of Gaussians (5 - 9)			
No. of Gauss.*	RMSE	PSNR (dB)	ROC** Cutoff	No. of Gauss.*	RMSE	PSNR (dB)	ROC** Cutoff
2	0.570	4.896	0.367	5	0.265	11.547	0.840
3	0.383	8.353	0.757	7	0.220	13.169	0.855
4	0.312	10.130	0.812	9	0.248	12.126	0.747
5	0.265	11.547	0.840				

* Number of Gaussians

** ROC Optimal Cutoff Measure

TABLE 3.10: Assessment tests on the MOG algorithm using different number of Gaussians per pixel on outdoor motion sequence

Table 3.10 shows the evaluation of changing the number of Gaussians for the MOG algorithm using an outdoor motion sequence. An MOG with 7 Gaussians gives the best result for the outdoor sequence giving the smallest RMSE, the highest PSNR and the highest ROC optimal cutoff measure as well. Therefore 7 Gaussians was used for outdoor motion sequences.

3.4.2 Experimental Results

The final settings for the indoor and the outdoor motion sequences were used to further test more motion sequences.

• Indoor Motion Sequences

The 10 indoor sequences were processed using the MOG algorithm with the optimised settings (in each one, 50-52 frames were used for background modelling). The optimised settings are: learning rate 0.006, the initial weight 0.0005, the initial variance 110, 0.2 background threshold and 5 Gaussians per mixture.

Samples of the extracted images are shown in Figure 3.5. From the samples shown, the motion pixels were mostly detected correctly in all the shown sequences. In the background region, some background pixels were labelled erroneously as motion pixels. The algorithm as stated does not suppress shadows. The shadow pixels are often erroneously labelled as motion pixels.

The evaluation of the averaged error of each sequence is shown in Table 3.11. The table shows the RMSE, the PSNR, and three different percentage errors, the background, foreground and the overall percentage error compared to the motion pixels.

Motion sequence 017a055s00R gave the highest RMSE, and the minimum PSNR value. While 019e063s00L motion sequence provided the minimum RMSE and the

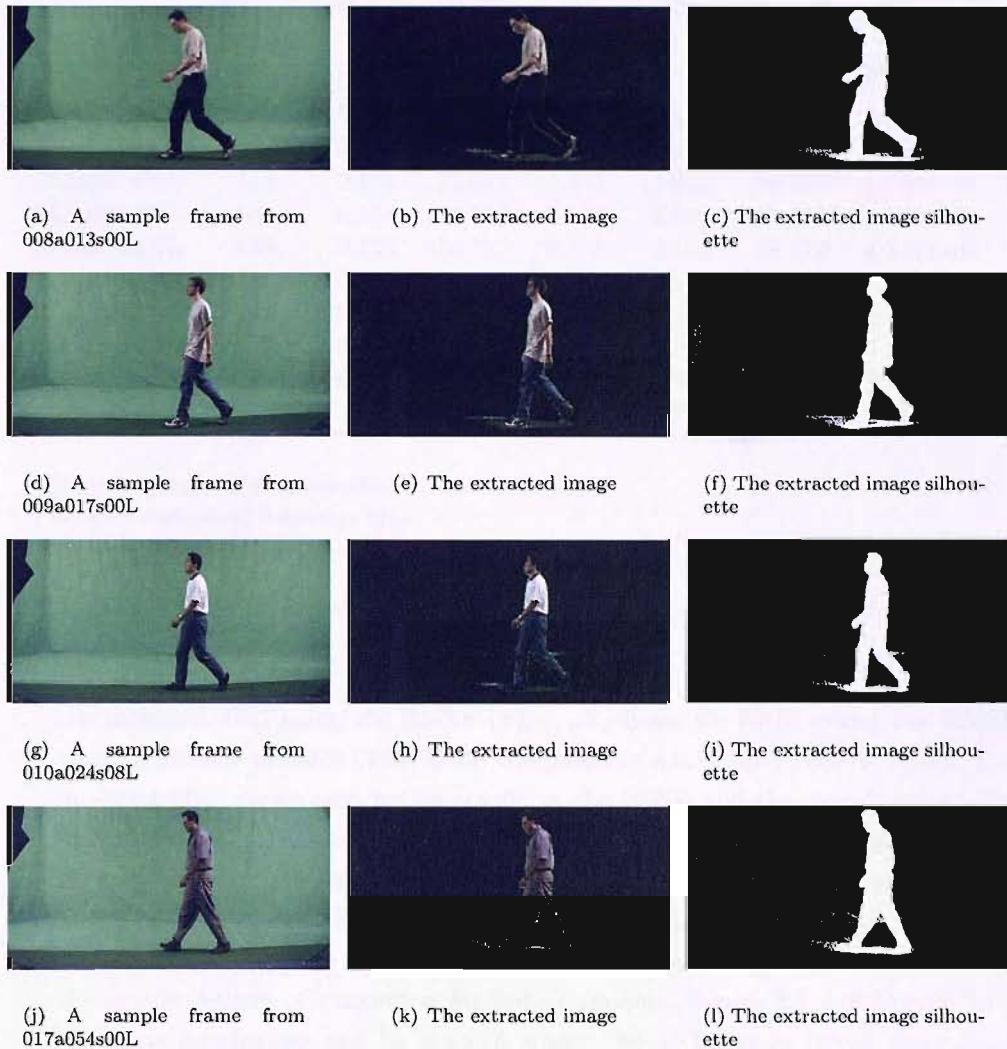


FIGURE 3.5: Examples of indoor images extracted using the MOG algorithm

highest PSNR. The percentage error for the foreground and the background pixels are low. But when comparing the total erroneous pixels to the motion pixels the result is quite different where the minimum error can reach one quarter of the motion pixels, 26.48% in 019a063s00L. The maximum percentage error reaches 63.51% in 017a055s00R (the error is more than half the actual motion pixels). The overall percentage error gives an indication that improvements is needed to reduce the overall detection error by removing the shadow and improving the shadow detection process. The recorded RMSE variance is small for all the tested sequences. The largest variance in motion sequence 008a013s00L is 7.361E-04. The small variance values shows the MOG consistency in its performance in each motion sequence.

When comparing the results of the MOG indoor sequences with the results of

Sequence Number	No. of Frames	RMSE	PSNR (dB)	FG Error ¹	BG Error ²	Overall Error ³	σ_{RMSE}^2
008a013s00L	178	0.172	15.389	0.482	3.236	41.557	7.361E-04
009a017s00L	169	0.148	16.670	0.605	2.331	34.925	3.564E-04
010a024s08L	187	0.142	17.006	0.912	2.083	32.267	9.206E-05
013a037s00L	114	0.135	17.518	1.371	1.913	36.270	5.746E-04
013a040s00L	184	0.154	16.378	0.963	2.500	41.115	5.545E-04
017a054s00L	188	0.195	14.232	0.395	4.102	58.479	4.337E-04
017a055s00R	162	0.201	13.979	0.661	4.284	63.510	2.778E-04
018a059s00L	188	0.138	17.246	1.181	1.983	29.788	2.419E-04
018a060s00L	179	0.174	15.299	0.729	3.237	46.140	6.101E-04
019a063s00L	186	0.128	17.896	1.094	1.665	26.483	3.693E-05
Average		0.159	16.161	0.839	2.733	41.053	3.914E-04

1 FG Error Foreground Percentage Error

2 BG Error Background Percentage Error

3 The percentage of the overall error compared to the motion pixels only

TABLE 3.11: Overall assessment on a number of motion indoor sequences using the MOG algorithm

the modified SBD using the RMSE (σ_{RMSE}^2) values, the SBD scored less RMSE with an average of 0.078 (4.047E-05) compared to a 0.159 (3.914E-04) MOG. The modified SBD also scored better results in the PSNR and the overall error. The foreground region extraction for the MOG is better with less foreground error and no holes in the moving subject. On the other hand, the background extraction is better in the modified SBD with less background error (0.49% in the modified SBD and 2.73% in the MOG) and less noise in the background region. When comparing the sample figures of extraction for both classifiers, Figure 3.5 and Figure 2.10, the same conclusions can be reached where the MOG gives better foreground extraction with less holes and the SBD gives better background suppression. The shadow suppression is a clear advantage for the SBD where most of the shadow (if not all in some of the samples) had disappeared while for the MOG the shadow is apparent on the extracted sequences. This difference in performance when utilised properly by using each classifier's strengths, justifies the fusion of both classifiers aiming for a better classifier, to be discussed later in Chapter 6 (notice also that the isolated mislabelled pixels mostly appear in different positions in the scene).

• Outdoor Motion Sequences

In the process of evaluating the MOG, 10 outdoor sequences were tested using this algorithm. 48-55 background frames were used for background adaption. The MOG algorithm with the optimised outdoor parameters were used. The parameters are: learning rate of 0.012, an initial weight of 0.0005, an initial variance of 90, 0.2 for the background threshold and 7 Gaussians per mixture.

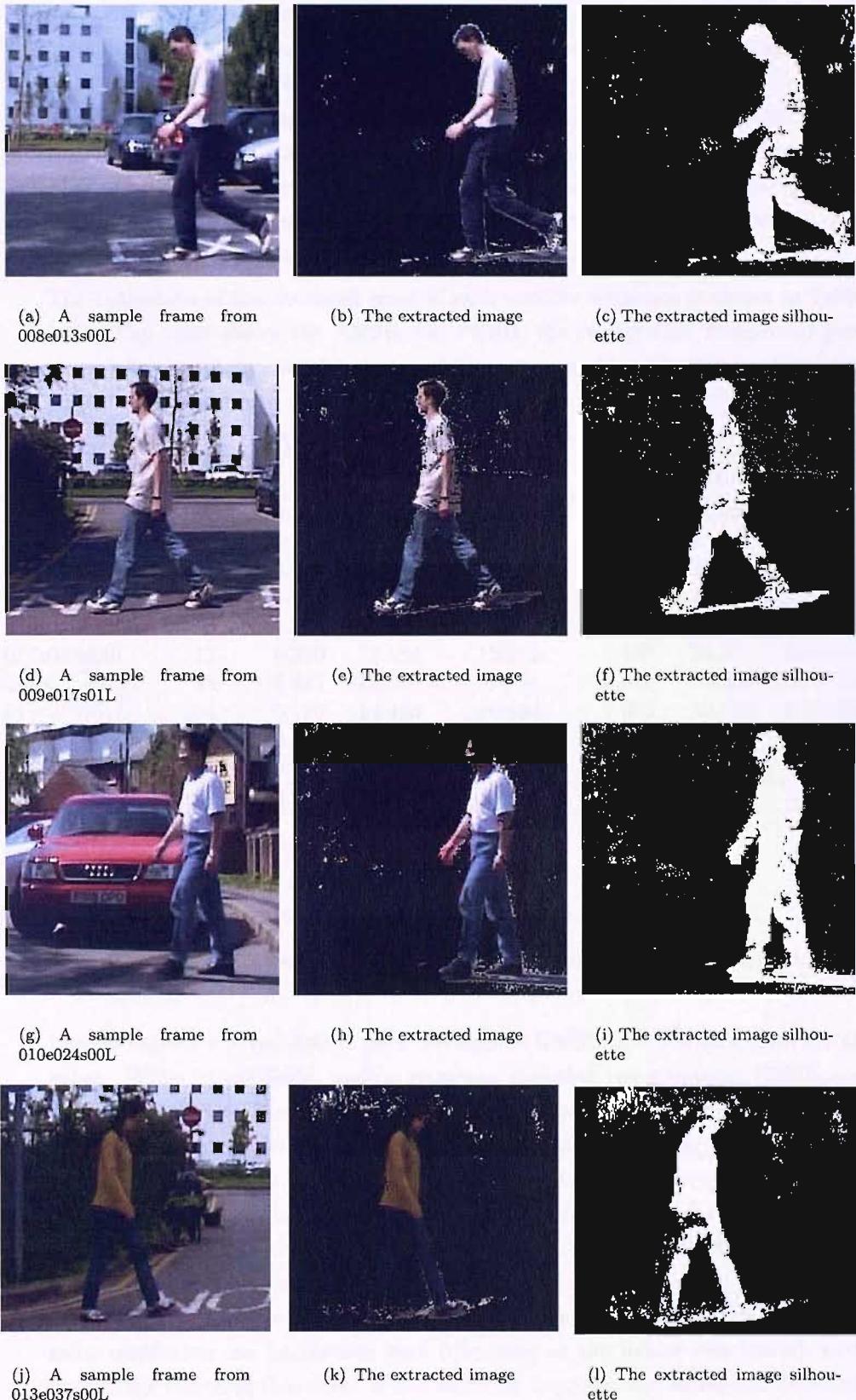


FIGURE 3.6: Examples of outdoor images extracted using the MOG algorithm

Figure 3.6 shows samples of outdoor extracted images. From the shown samples the motion pixels are not always detected correctly. In motion sequences 008e013s00L, 009e017s01L and 013e037s00L the problem is clearly illustrated with big holes appearing on the moving object. Sample of motion sequence 010e024s00L shows small holes in the moving object. Also, the background is not detected precisely where many groups of points can be observed in many parts of the background. Sample of motion sequence 013e037s00L has the worst background extraction. The shadow is labelled as motion pixels in all the extracted images.

The evaluation of the averaged error of each outdoor sequence is shown in Table 3.12. The table shows the RMSE, the PSNR, the background/foreground percentage error and the overall percentage error compared to the silhouette's motion pixels.

Sequence Number	No. of Frames	RMSE	PSNR (dB)	FG Error ¹	BG Error ²	Overall Error ³	σ_{RMSE}^2
008e013s00L	100	0.207	13.719	8.634	3.587	31.932	2.964E-04
009e017s01L	96	0.226	12.923	16.551	3.479	41.119	1.151E-04
010e024s00L	94	0.237	12.533	8.216	5.176	39.679	1.395E-04
013e037s00L	158	0.216	13.365	11.520	3.747	37.868	6.922E-04
013e040s00L	151	0.240	12.403	8.370	5.408	48.967	4.123E-05
017e054s00L	112	0.220	13.184	12.949	3.498	34.476	2.398E-04
017e055s00R	88	0.251	12.036	8.304	6.025	50.033	3.709E-04
018e059s01L	104	0.190	14.426	10.194	2.663	28.540	6.936E-05
018e060s00L	88	0.207	13.691	7.171	3.846	31.445	2.294E-04
019e063s05L	112	0.169	15.458	8.184	2.331	30.935	1.763E-04
Average		0.216	13.374	10.009	3.976	37.499	2.370E-04

¹ FG Error Foreground Percentage Error

² BG Error Background Percentage Error

³ The percentage of the overall error compared to the motion pixels only

TABLE 3.12: Overall assessment on a number of motion outdoor sequences using the MOG algorithm

Motion sequence 017e055s00R gave the highest RMSE, and the minimum PSNR value. While 019e063s05L motion sequence provided the minimum RMSE and the highest PSNR. The foreground percentage error is higher than in the indoor sequences (expected since outdoor sequences are more challenging) where the minimum error recorded is 7.17% and the maximum 16.55%. The average foreground error is 10.01%. For the background percentage error the values ranged between 2.33% to 6.03%. The average error for the background region is 3.98%. Finally the percentage error compared to the silhouette motion pixels gives high results where the average is 37.50% with a minimum of 28.54% and a maximum of 50.03%. The same conclusion can be reached here (the same as the indoor conclusion), such percentage indicates that there is still room for improvement on motion detection and shadow suppression. The MOG error variance ,RMSE variance, gave a small

value which means large consistency in its performance in each sequence. The variance ranged from 4.123E-05 in 013e040s00L to 6.922E-04 in 013e037s00L.

Comparing the sample images for the MOG in Figure 3.6 with the modified SBD in Figure 2.11 shows that in motion sequences 009e017s01L and 010e024s00L MOG gave better extraction for the moving subject (than the modified SBD) but in the other two samples, 008e013s00L and 013e037s00L, the MOG result were worse than the modified SBD with large wholes in both motion sequences. On the other hand, the background region in the modified SBD has few noisy pixels compared with groups of connected noisy points in the MOG. In the shadow region, the modified SBD managed to suppress the shadow in 013e037s00L while the MOG does not support shadow suppression. We notice from the comparison that the classifiers perform differently in different regions. When comparing the results of the MOG classifier with the modified SBD for the outdoor sequences using the RMSE (σ_{RMSE}^2) values, the SBD scored less error with an averaged RMSE of 0.189 (2.126×10^{-4}) compared to 0.216 (2.370×10^{-4}) for the MOG. The modified SBD also scored better results in the PSNR and the overall error. The MOG gave better results for the the foreground region (considering also the moving object holes shown in the SBD samples figure). In the background region, the modified SBD performed better with an error of 2.15% compared to an error of 3.98% in the MOG classifier. Proper fusion of those different performing classifiers can result in improving the overall performance.

3.5 Conclusions

In this chapter the MOG algorithm was presented. Optimised performance was achieved by experimenting with different settings of the algorithm's parameters. The best values were those which optimised performance figures developed earlier for moving object extraction analysis. When optimised values had been selected, the algorithm was tested on indoor and outdoor motion sequences. Also different analytical assessment methods were used to clarify the algorithm performance. The algorithm lacks the ability to extract shadows. The algorithm's background suppression can still be optimised further and the algorithm's motion extraction showed some shortcomings on outdoor sequences

Chapter 4

Unary Classifiers

4.1 Introduction

Unary classification (UC) is concerned with a single class with a decision function that states the likelihood of a given data being a member to such class. The method determines a class boundary using the data given for a class.

Support Vector Machines (SVMs) can be used to implement unary classification (Schölkopf et al., 2001; Chen et al., 2001; Manevitz and Yousif, 2001). There exist two SVM based methods that can be used for unary classification, Hyperplane (Schölkopf et al., 2001) and Hypersphere (Tax and Duin, 1999) methods. In a Hyperplane method, data is bounded using a hyperplane in a feature space. In a Hypersphere the data is bounded using a hypersphere in a feature space. The strategy we use is to map the data into a feature space and then to use a hypersphere to determine membership of the class.

In the following sections we will present our new motion classifier using the Hypersphere UC method. We will start with a section on kernel functions followed by a section on the Hypersphere method in which we will explain the details of our classifier. After that we will present the results of using the UC on indoor and outdoor motion sequences. This section will be followed by a novel improvement on the UC method to improve its performance. Finally, we will show the results of the improved UC on indoor and outdoor sequences.

4.2 Kernel Functions

Kernel functions are a way to represent an inner product. In many learning algorithms the only place the data appears is in an inner product between examples in feature space. Kernel functions are an implicit representation of this inner product in the input

space. Using the kernel functions data representation and then bounding the data using a hyperplane or a hypersphere increases the options in making class boundaries for the data.

The mapping function $\Phi : \chi \rightarrow F$ maps vectors from input space to feature space. The dot product between examples can be computed by evaluating some simple kernel.

$$k(\mathbf{x}, \mathbf{x}') = \Phi(\mathbf{x}) \cdot \Phi(\mathbf{x}') \quad (4.1)$$

Many kernels can be used

$$\text{Linear} : k(\mathbf{x}, \mathbf{x}') = \mathbf{x} \cdot \mathbf{x}' \quad (4.2)$$

$$\text{Polynomial} : k(\mathbf{x}, \mathbf{x}') = (\mathbf{x} \cdot \mathbf{x}' + 1)^p \quad (4.3)$$

$$\text{Gaussian Radial Bias} : k(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{x} \cdot \mathbf{x}'\|^2}{2\sigma^2}\right) \quad (4.4)$$

$$\text{Exponential Radial Bias} : k(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{x} \cdot \mathbf{x}'\|}{2\sigma^2}\right) \quad (4.5)$$

where \mathbf{x} is a point in input space, p is the polynomial degree, and σ^2 is the variance.

The Gaussian radial basis kernel will be used as a kernel which is known to be useful to approximate multivariate functions efficiently (Buhmann, 2000).

4.3 Hypersphere Method

For a training set of $\mathbf{x}_1, \dots, \mathbf{x}_\ell \in \chi$ where $\ell \in \mathbb{N}$ is the number of the trained samples and χ is some set. A hypersphere in a feature space is defined by:

$$\left\{ \Phi(\mathbf{x}) \mid \forall \mathbf{x} \in \mathbb{R}, R^2 - \|\Phi(\mathbf{x}) - \mathbf{c}\|^2 = 0 \right\} \quad (4.6)$$

Where $R \in \mathbb{R}$ is the hypersphere radius in a feature space, $\mathbf{c} \in \mathbb{R}^N$ is the hypersphere centre in a feature space. The decision function will be

$$f(\mathbf{x}) = \text{sgn}(R^2 - \|\Phi(\mathbf{x}) - \mathbf{c}\|^2) \quad (4.7)$$

If the function is positive then \mathbf{x} lies within the boundary of the sphere defined by the R and \mathbf{c} , if the function is negative then \mathbf{x} lies outside the sphere, Fig 4.1.

To encapsulate most of the data within a minimal sphere we solve the following equation

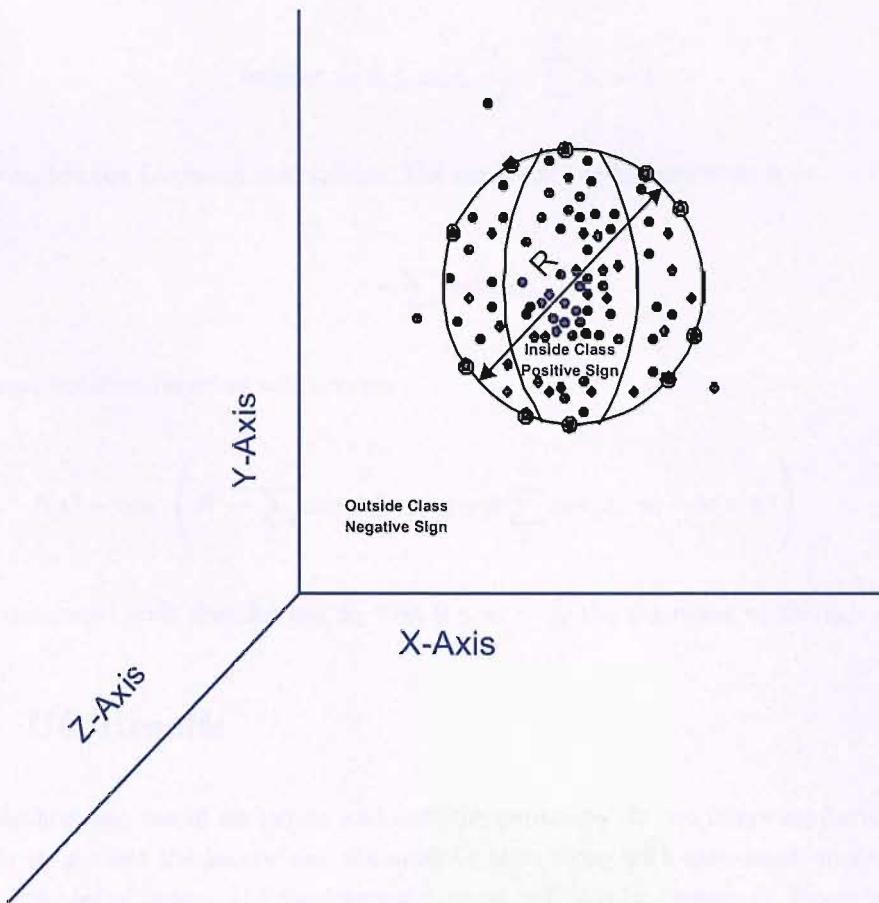


FIGURE 4.1: A Hypersphere with clustered data. Data in the sphere will be given a positive sign while negative signs will be set for data outside the sphere

$$\min_{R \in \mathbb{R}^+, \xi \in \mathbb{R}^\ell, c \in \mathbb{F}} R^2 + \frac{1}{\nu \ell} \sum_i^\ell \xi_i$$

$$\text{subject to } \|\Phi(\mathbf{x}) - \mathbf{c}\|^2 \leq R^2 + \xi_i, \xi_i \geq 0 \text{ for } i = 1, \dots, \ell. \quad (4.8)$$

Where ν is a margin coefficient that controls the amount of training vectors to be included within the hypersphere boundary, and ξ_i is the error cost corresponding to each training vector.

By using the Lagrangian multiplier method, the following dual optimisation problem will result

$$\min_{\alpha} \sum_{i,j}^\ell \alpha_i \alpha_j k(\mathbf{x}_i, \mathbf{x}_j) - \sum_i^\ell \alpha_i k(\mathbf{x}_i, \mathbf{x}_i) \quad (4.9)$$

$$\text{subject to } 0 \leq \alpha_i \leq \frac{1}{\nu\ell}, \quad \sum_i^{\ell} \alpha_i = 1 \quad (4.10)$$

where α_i are the Lagrange multipliers. The centre of the hypersphere is

$$c = \sum_i^{\ell} \alpha_i \Phi(\mathbf{x}_i), \quad (4.11)$$

The final decision function will become

$$f(\mathbf{x}) = \text{sgn} \left(R^2 - \sum_{i,j}^{\ell} \alpha_i \alpha_j k(\mathbf{x}_i, \mathbf{x}_j) + 2 \sum_i^{\ell} \alpha_i k(\mathbf{x}_i, \mathbf{x}) - k(\mathbf{x}, \mathbf{x}) \right) \quad (4.12)$$

R^2 is computed such that for any \mathbf{x}_i with $0 \leq \alpha_i \leq \frac{1}{\nu\ell}$ the argument of the sgn is zero.

4.4 UC Results

The method was tested on indoor and outdoor sequences. In the following parts of this section we present the indoor and the outdoor tests along with assessment measures on each. Samples of indoor and outdoor extractions will also be presented. Experimenting using the RGB colour model resulted in shadows appearing erroneously as part of the foreground both on indoor and on outdoor motion sequences as shown in Figure 4.2

A detailed analysis of the Unary Classifier using different colour models is shown in Chapter 5. The chapter concludes in choosing $c_1c_2c_3$ colour model, Section 5.2.5, to optimise the Unary Classifier performance. A detailed test for the Unary Classifier using $c_1c_2c_3$ colour model will be shown in this section using indoor and outdoor motion sequences.

To set up the UC parameters for indoor and outdoor motion sequences, the soft margin coefficient, ν , is set to 0.01 (to include 99% of the background pixels in the background model). Also the RBF's σ is set to a large value of 10 to enable the RBF kernel which is used with in the UC to include the wide range of background pixels variations. A detailed analysis and assessments of different settings for σ will be discussed later in Chapter 5.

- **Indoor Motion Sequences**

10 indoor sequences are used in these tests. 50-52 frames from each were used for background modelling (these frames are pure background and do not contain any moving objects).

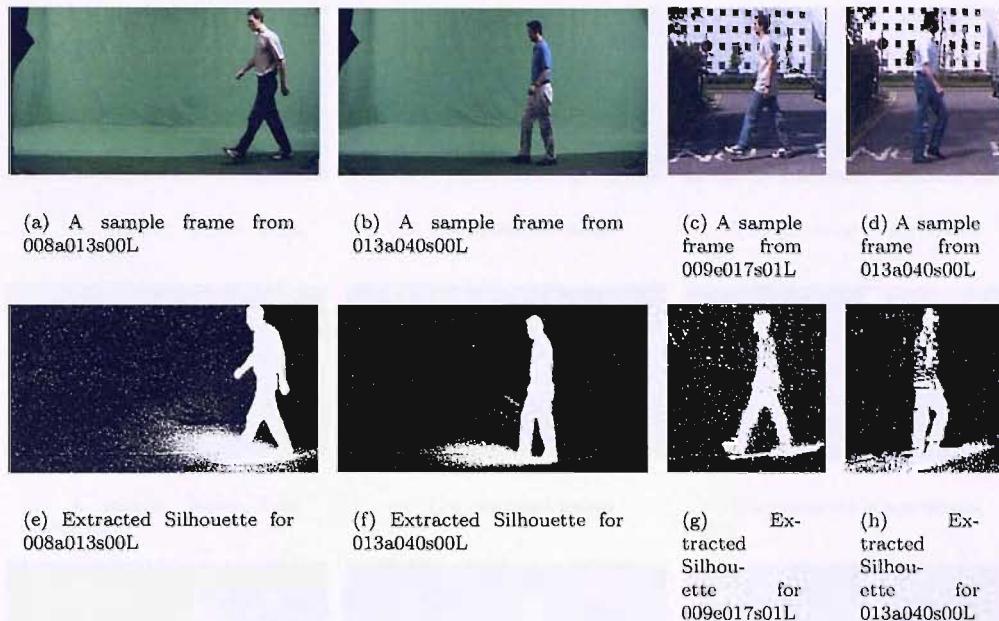


FIGURE 4.2: Examples of indoor and outdoor images extracted using the Unary Classifier algorithm using the RGB colour model

Figure 4.3 shows samples of extracted sequences using the Unary Classifier. From the samples shown, the background is noisy with background pixels misclassified as motion pixels. The sample of motion sequence 017a054s00L showed more noise than the other three motion sequence samples. The motion pixels are detected well in all the shown samples except for the moving object borders where it is not finely extracted. A reasonable part of the shadow has disappeared due to the use of the $c_1c_2c_3$ colour model instead of the RGB. Part of the shadow is still resident in all the shown samples.

A detailed assessment for the indoor motion sequences is presented in Table 4.1. The table shows the RMSE, the PSNR, and the percentage error for the background, foreground and the overall error compared to the frames silhouette's motion pixels.

The maximum RMSE was scored by motion sequence 017a054s00L which has also given the minimum PSNR. The minimum RMSE is scored by 018a059s00L. The maximum PSNR was given by the same motion sequence along with 013a040s00L. For the foreground percentage the error was small for all the motion sequences with a maximum of 0.33% scored by both 010a024s08L and 013a040s00L. The average foreground error was 0.22%. In the background percentage error, the maximum value was given by 017a054s00L with 11.72%. All the other sequences scored similar values in the range of 7.26-8.82%. The overall error compared to the silhouette motion pixels was very high in all the sequences exceeding even the size of the motion object. The values ranged from 104.87-164.87% with an average of

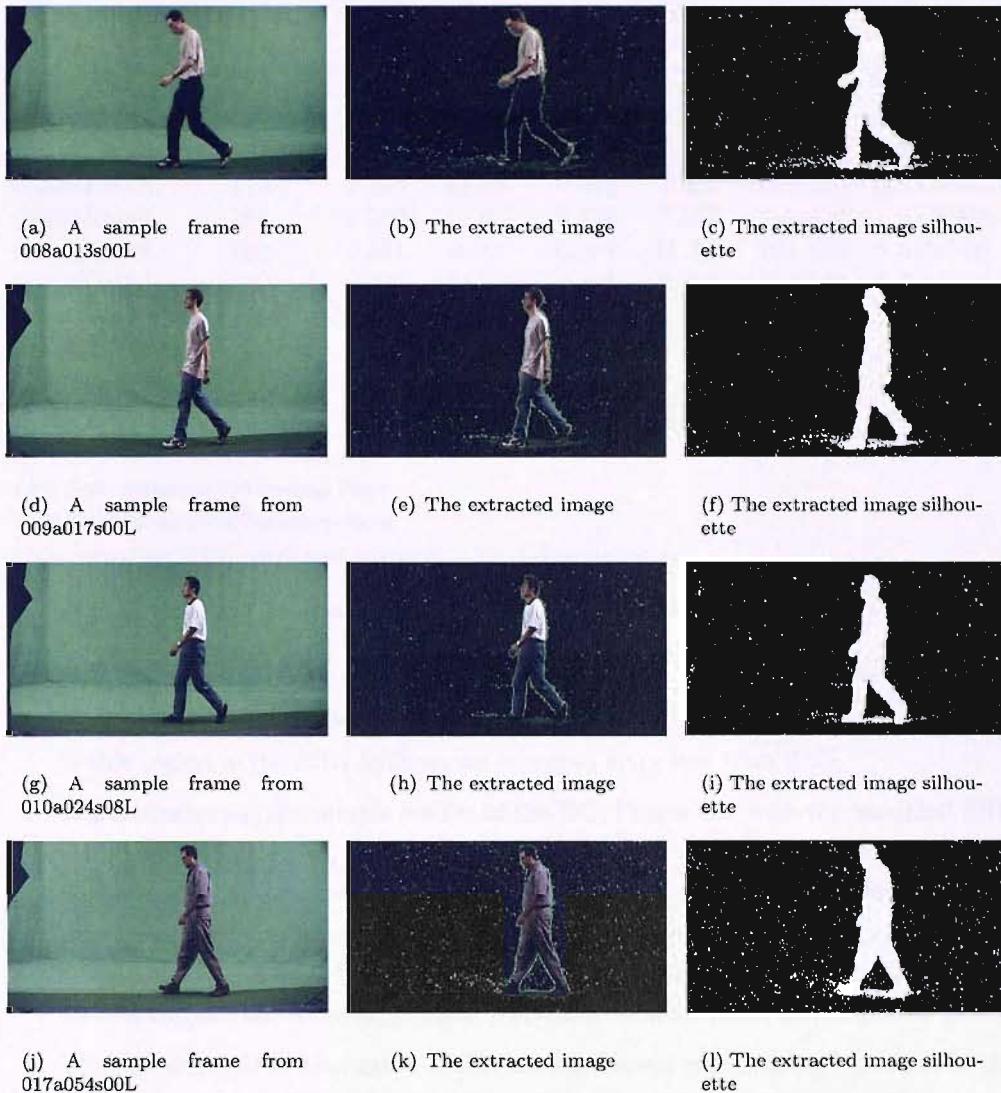


FIGURE 4.3: Examples of indoor images extracted using the UC algorithm

118.66%. The RMSE variance is small for all the sequences ranging from 7.279E-06 in 009a017s00L to 7.982E-05 in 008a013s00L. The small error variance means a large consistency performance for this algorithm.

Comparing the indoor results of the UC with the previously presented classifiers using the RMSE (σ_{RMSE}^2) results, the minimum error is scored by the modified SBD, 0.078 (4.047E-05), followed by the MOG, 0.159 (3.914E-04), and the UC scored the highest error, 0.273 (3.941E-05). The MOG and the modified SBD, the UC scored the worst overall error with an average of 118.66% (i.e. larger than the silhouette), followed by the MOG with an averaged overall error of 41.05%. The modified SBD gave the minimum overall error, 9.78%. When looking closely to the results the UC main weak point is in the background region with an averaged

Sequence Number	Number of Frames	RMSE	PSNR (dB)	FG Error ¹	BG Error ²	Overall Error ³	σ_{RMSE}^2
008a013s00L	178	0.286	10.879	0.212	8.816	112.038	7.982E-05
009a017s00L	169	0.265	11.547	0.117	7.475	109.847	7.279E-06
010a024s08L	187	0.262	11.626	0.326	7.329	110.363	6.062E-05
013a037s00L	114	0.263	11.597	0.498	7.291	132.105	4.966E-05
013a040s00L	184	0.262	11.655	0.326	7.258	115.937	7.540E-05
017a054s00L	188	0.331	9.607	0.094	11.724	164.872	8.615E-06
017a055s00R	162	0.270	11.373	0.181	7.782	114.147	2.264E-05
018a059s00L	188	0.261	11.655	0.207	7.298	104.866	1.357E-05
018a060s00L	179	0.273	11.290	0.172	7.966	111.480	6.465E-05
019a063s00L	186	0.262	11.651	0.104	7.284	110.984	1.188E-05
Average		0.273	11.288	0.224	8.022	118.664	3.941E-05

1 FG Error Foreground Percentage Error

2 BG Error Background Percentage Error

3 The percentage of the overall error compared to the motion pixels only

TABLE 4.1: Overall assessment on a number of motion indoor sequences using the UC algorithm

error of 8.02%. The MOG background error was less than 3%. The best performer in this region is the SBD with an averaged error less than 0.5%.

When comparing the sample results of the UC, Figure 4.3, with the modified SBD and the MOG samples, Figure 2.10 and Figure 3.5 respectively, the background region in the UC is excessively noisy compared with the other two classifiers. On the other hand, when comparing the moving subject extraction quality of the three classifiers, the UC and the MOG scored the best results while the modified SBD moving subject has some large holes in addition to some small isolated noisy pixels.

The overall result of extraction of this classifier is not encouraging especially in the background region. Therefore an effort was made to improve the overall results of this classifier and will be shown in the following section.

• Outdoor Motion Sequences

For the outdoor sequences, 10 sequences were tested using the UC algorithm. 48-55 background frames were used for background adaption.

Figure 4.4 shows samples of outdoor extracted images. From the shown samples, the algorithm performance did not perform as well as in indoor environment.

Motion pixels are mostly detected correctly but holes still show in the moving subject in all the presented sequences samples. The problem is more serious in motion sequence 009e017s01L. The borders of the extracted objects are again not finely extracted. The background region is very noisy in all the presented samples with background pixels mislabelled as motion pixels. Shadow disappeared from sequence 013e037s00L but still resident in all the sequences.

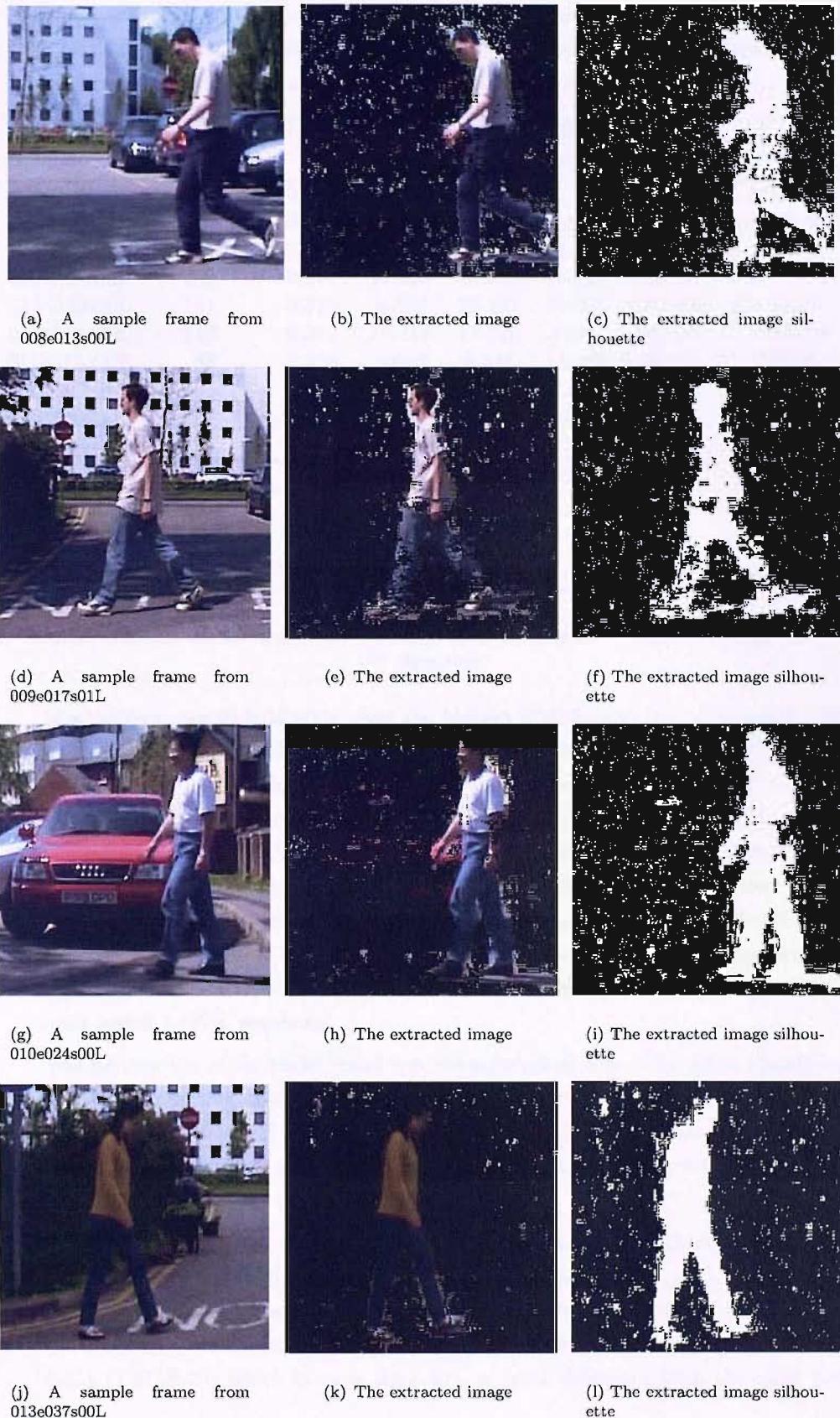


FIGURE 4.4: Examples of outdoor images extracted using the UC algorithm

The evaluation of the averaged error of each outdoor sequence is shown in table 4.2. The table shows the RMSE, the PSNR, and the percentage error of the foreground, background and the overall error over the silhouette's motion pixels.

Sequence Number	Number of Frames	RMSE	PSNR (dB)	FG Error ¹	BG Error ²	Overall Error ³	σ_{RMSE}^2
008e013s00L	100	0.330	9.625	6.418	11.618	81.768	1.683E-04
009e017s01L	96	0.314	10.086	16.709	8.882	79.158	3.189E-04
010e024s00L	94	0.343	9.298	5.610	12.800	83.350	3.243E-05
013e037s00L	158	0.275	11.216	3.981	8.096	60.877	1.339E-04
013e040s00L	151	0.376	8.506	20.397	13.335	120.390	3.501E-04
017e054s00L	112	0.312	10.118	17.726	8.454	69.696	1.621E-04
017e055s00R	88	0.329	9.653	8.528	11.199	85.951	1.552E-04
018e059s01L	104	0.267	11.492	3.412	7.644	56.100	9.876E-05
018e060s00L	88	0.294	10.623	4.960	9.256	63.491	3.101E-05
019e063s05L	112	0.266	11.512	4.914	7.284	76.572	2.061E-05
Average		0.311	10.213	9.265	9.857	77.735	1.471E-04

1 FG Error Foreground Percentage Error

2 BG Error Background Percentage Error

3 The percentage of the overall error compared to the motion pixels only

TABLE 4.2: Overall assessment on a number of motion outdoor sequences using the UC algorithm

Motion sequence 013e040s00L gave the highest RMSE, and the minimum PSNR value. While 019e063s05L motion sequence provided the minimum RMSE and the highest PSNR. The range of error for the foreground is in the interval of 3.41% to 20.40% with an average of 9.27%. The background percentage error gave a similar average of 9.86% with a range of error between 7.28% to 13.34%. When looking at the percentage of the overall error over the silhouette's motion pixels, the maximum error was as high as 120.39% while the minimum was 56.10%. The average error here is 77.74%. The small RMSE variance that ranges between 2.061E-05 and 3.501E-04 shows a large system consistency in error performance in each tested motion sequence.

The percentage of the background error is substantial especially when visualising the effect of such error in the samples presented previously, Figure 4.3 and Figure 4.4. Also, the score of the overall error is high with an average exceeding two thirds of the moving object size. This flags a problem in this algorithm and leaves room for improvement to reduce the error.

When comparing the outdoor results of the UC classifier in Table 4.2 with the results of the modified SBD and the MOG using the RMSE (σ_{RMSE}^2), the modified SBD scored the least error, 0.189 (2.126E - 04), followed by the MOG, 0.216 (2.370E-04). The UC scored the worst highest error among the three classifiers, 0.311 (1.471E-04) which is more than 40% of error difference than the other two

classifiers. The overall error for the UC is more than double the error in the other two classifiers (77.74% in UC, 29.08% in modified SBD, and 37.50% in MOG). In the background region, the UC again scored the worst result with 9.86% error while the MOG and the modified SBD error is only 3.98% and 2.15% respectively. The UC scored the least error in the foreground error region, 9.27%. The MOG scored almost the same error, 10.01%, while the modified SBD scored the worst with a 14.45% of error.

When comparing the UC samples, in Figure 4.4, with the samples of the modified SBD, Figure 2.11, and the MOG, Figure 3.6, the background region of the UC is excessively noisy while the SBD has few isolated noisy background pixels. The MOG background shown some noise. In the moving subject quality extraction (the foreground region), the modified SBD result was the worst with large holes appearing on the moving subject body. The MOG and the UC overall foreground extraction is better (than the SBD) in most of the shown samples.

Overall this algorithm, as the indoor and outdoor result indicates (figures and tables), has a high accuracy in identifying the foreground pixels. On the other hand, the algorithm's main problem is in misclassifying the background pixels. This is due to the fact that this algorithm has drawn a hypersphere around the training background pixels and even though it had left some margin of error, this margin is not sufficiently wide to tolerate the range of change experienced with background pixels. Thus the extracted samples with the error analysis suggest clearly that there is still work to be done to reduce such errors especially in the background region for both indoor and outdoor motion sequences.

4.5 Improved UC

As concluded in the last section that the main problem is the tightness of the sphere size encapsulating the training data. The value for ν was selected so as to include all the data i.e ξ was set to a maximum value as well but still the result as shown in the previous section was not satisfactory with an average error exceeding the moving object size in the indoor sequences and more than two thirds of the moving object size in the outdoor sequences. Enlarging the size of the sphere will enable the system to be more tolerant to changes in the background model. To accomplish this goal we modified the sphere radius in the decision function of the UC to be

$$R_{Improved.UC} = R_{UC} + M_R R_{UC} \quad (4.13)$$

where $M_R \in \mathbb{N}$ is the radius multiplier and R_{UC} is the unary classifier sphere radius. The final decision function becomes

$$f(\mathbf{x}) = \text{sgn} \left((R_{UC} + NR_{UC})^2 - \sum_{i,j}^{\ell} \alpha_i \alpha_j k(\mathbf{x}_i, \mathbf{x}_j) + 2 \sum_i^{\ell} \alpha_i k(\mathbf{x}_i, \mathbf{x}) - k(\mathbf{x}, \mathbf{x}) \right) \quad (4.14)$$

4.6 Improved UC Results

The improvement was tested on indoor and outdoor motion sequences with an incremental values for the radius multiplier starting from 0 with a step of 2 until the optimal value is reached for the tested sequence. A stopping criteria is adopted to stop the incrementing procedure when the error reduction does not exceed 0.5% in the overall percentage error. This condition will avoid enlarging the radius size excessively for a minimal error reduction.

- **Indoor Motion Sequences**

The same 10 indoor sequences used in the UC section are used in these tests. Each sequence is tested by changing the radius size.

Table 4.4 shows the optimising table for the same indoor motion sequence, 018a059s00L.

M_R	ROC ¹	RMSE	PSNR (dB)	FG Error ²	BG Error ³	Overall Error ⁴	σ_{RMSE}^2
0	0.925	0.261	11.655	0.207	7.298	104.866	1.357E-05
2	0.981	0.103	19.721	0.816	1.089	16.448	2.729E-05
4	0.981	0.081	21.847	1.334	0.612	10.106	3.255E-05
6	0.978	0.072	22.850	1.788	0.435	8.021	3.043E-05
8	0.974	0.068	23.422	2.223	0.335	7.020	2.461E-05
10	0.971	0.065	23.785	2.634	0.266	6.445	2.091E-05
12	0.967	0.063	23.984	3.047	0.217	6.149	1.879E-05
14	0.964	0.063	24.092	3.442	0.178	5.992	1.651E-05
16	0.960	0.063	24.082	3.868	0.149	5.999	1.528E-05

1 ROC Optimal Cutoff Measure

2 FG Error Foreground Percentage Error

3 BG Error Background Percentage Error

4 The percentage of the overall error compared to the motion pixels only

TABLE 4.3: Assessment on the indoor motion sequences 018a059s00L using the improved UC algorithm

In Table 4.4 the highest optimal cutoff measure for the ROC, 0.981, is scored by the radius multiplier 2 and 4. The best RMSE is scored by M_R 12, 14 and 16 with a value of 0.063. For the PSNR radius multiplier size 14 scored the maximum value.

In the foreground error column we notice that the minimum error is in the smallest radius, $M_R = 0$. The foreground error increases as we increase the radius size with a maximum error in $M_R = 16$. The background error behaves in the opposite manner. $M_R = 0$ gave the highest error with the error decreasing as we increased the radius size. The overall error gives a compromise between the foreground and the background error. The overall error compared to the motion pixels presented $M_R = 14$ with the minimum error. But to abide by the condition set (not to excessively increase the radius size) the error drops less than %0.5 after $M_R = 10$. Thus a radius multiplier size of 10 will be used for motion sequence 018a059s00L. All the detailed tables for the other sequences are presented in appendix A.1.

The final assessment for the indoor motion sequences with the selected radius multiplier are presented in Table 4.4. The table shows the radius multiplier M_R , RMSE, the PSNR, the background and foreground percentage error, and the overall error compared to the frames silhouette's motion pixels.

Sequence Number	Number of Frames	M_R	RMSE	PSNR (dB)	FG Error ¹	BG Error ²	Overall Error ³	σ_{RMSE}^2
008a013s00L	178	12	0.079	22.066	3.031	0.438	8.528	4.420E-05
009a017s00L	169	14	0.072	22.910	2.532	0.378	8.046	2.424E-05
010a024s08L	187	12	0.073	22.809	3.126	0.352	8.393	2.098E-05
013a037s00L	114	10	0.070	23.144	4.409	0.274	9.310	5.052E-05
013a040s00L	184	12	0.077	22.292	3.346	0.427	10.102	6.996E-05
017a054s00L	188	16	0.078	22.189	2.028	0.509	9.127	3.284E-05
017a055s00R	162	12	0.069	23.217	2.300	0.357	7.505	3.132E-05
018a059s00L	188	10	0.065	23.785	2.634	0.266	6.445	2.091E-05
018a060s00L	179	14	0.075	22.522	2.741	0.407	8.388	2.351E-05
019a063s00L	186	10	0.065	23.784	1.865	0.326	6.801	2.092E-05
Average		0.072	22.872	2.801	0.373	0.373	8.265	3.394E-05

1 FG Error Foreground Percentage Error

2 BG Error Background Percentage Error

3 The percentage of the overall error compared to the motion pixels only

TABLE 4.4: Overall assessment on a number of motion indoor sequences using the UC algorithm

The minimum RMSE is scored by 018a059s00L and 019a063s00L. The maximum PSNR was given by 018a059s00L. The maximum RMSE was scored by motion sequence 008a013s00L which has also given the minimum PSNR. For the foreground percentage error, the smallest error was 1.87% by motion sequence 019a063s00L. The maximum error was for motion sequence 013a037s00L with an error of 4.41%. The average foreground error was 2.80%. In the background percentage error column, the error was small for all the sequences with a maximum of 0.51% scored by 017a054s00L. The average background error over all the sequences is 0.38%. The overall error compared to the silhouette motion pixels was in the range of

6.45-10.10% with an average of 8.27%. The modified UC RMSE variance illustrates small error displacement when performing in indoor motion sequences. The largest error variance scored is 6.996E-05, in motion sequence 013a040s00L. The small RMSE variances demonstrates a large error consistency for the modified UC. When comparing the modified UC result, Table 4.4, with the original UC classifier, Table 4.1 the reduction in the error is quite substantial. For instance, the RMSE dropped almost 74% from 0.273 to 0.072. The PSNR improved from 11.288dB to 22.872dB (more than 100% improvement). The background error and the overall error dropped as well. The foreground error is the only error which appeared to increase. Given that there is less confidence in this measure, the modification overall would appear successful.

Comparing the performance of the modified UC on indoor sequences, Table 4.4, with the modified SBD, Table 2.7, and the MOG classifiers, Table 3.11, shows that the modified UC outperformed the other two in most of the measurements. For the RMSE (σ_{RMSE}^2) result, the modified UC also scored the best results with the least RMSE of 0.072 (3.394E-05), followed by the modified SBD, 0.078 RMSE and (4.047E-05) variance. The worst performer in this measure is the MOG with a RMSE of 0.159 (3.914E-04). Also in the background region, the modified UC gave the least error, 0.37%, followed by the modified SBD, 0.49%, and then the MOG, 2.73%. In addition, the modified UC results in the improved UC overall error, the PSNR are also better than the other two classifiers. However, the modified UC scored the highest error in the foreground region but with a value very close to the modified SBD error (2.80% for the modified UC and 2.66% for the modified SBD). The MOG scored the least error in this region, 0.84%.

Figure 4.5 shows samples of extracted sequences using the improved Unary Classifier. The samples of the extracted images in Figure 4.5 show the background is now much cleaner except for the small traces of shadow noticeable in 017a054s00L. The motion pixels are detected well in all the shown samples, though some small noise holes exist in all the extracted subjects. When comparing the extracted samples of the modified UC with the the modified SBD, Figure 2.10, and the MOG, Figure 3.5, extracted samples. The background of the UC is much cleaner than the other two classifiers background. The shadow extraction of the modified UC is also as good as the improved SBD (UC is slightly better). For the the quality of the extracted subject, the MOG is the best performer in this region while the improved SBD is the worst. The extraction quality of the improved UC is comparable to the MOG quality except for some small holes (mostly isolated noisy pixels).

- **Outdoor Motion Sequences**

The 10 outdoor sequences used in the UC section are used in these tests. The radius size was changed also on each sequence to determine its effect on performance.

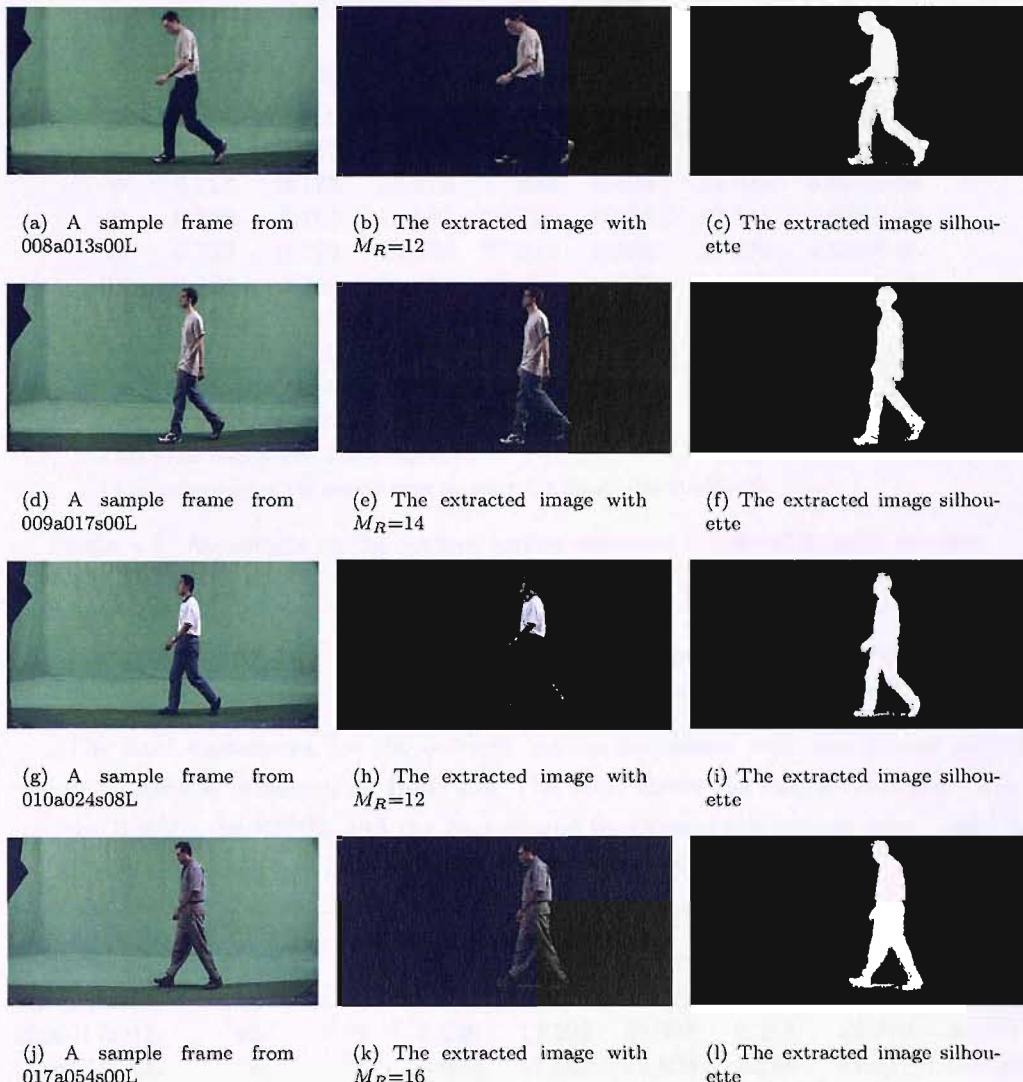


FIGURE 4.5: Examples of indoor images extracted using the improved UC algorithm

Table 4.5 shows the optimising table for the indoor motion sequence 018e059s01L. The highest optimal cutoff measure for the ROC, 0.901, is scored by $M_R = 2$. The least RMSE value 0.16 is scored by the sphere radius multiplier 2. For the PSNR, radius multiplier size 2 scored the maximum value of 16.01dB. In the foreground error column we notice that the minimum error is in the smallest radius, $M_R = 0$. The foreground error increases as we increase the radius size with a maximum error in $M_R = 16$. The background error behaves in the opposite manner. $M_R = 0$ gave the highest error with the error decreasing as the radius size was increased. The foreground and the background error are inversely proportional. The overall error over the silhouette's motion pixels gives a compromise between the foreground and the background error. The minimum overall error 19.94% is given by $M_R = 2$.

M_R	ROC ¹	RMSE	PSNR (dB)	FG Error ²	BG Error ³	Overall Error ⁴	σ_{RMSE}^2
0	0.889	0.267	11.492	3.412	7.644	56.100	9.876E-05
2	0.901	0.159	16.005	8.170	1.708	19.937	1.195E-04
4	0.860	0.161	15.901	12.938	1.084	20.410	1.127E-04
6	0.813	0.173	15.273	17.854	0.824	23.539	8.873E-05
8	0.766	0.187	14.581	22.750	0.695	27.541	5.903E-05
10	0.722	0.199	14.024	27.211	0.590	31.278	4.619E-05
12	0.686	0.209	13.613	30.868	0.509	34.372	4.089E-05
14	0.654	0.217	13.263	34.170	0.445	37.235	2.948E-05
16	0.625	0.225	12.967	37.111	0.399	39.855	2.746E-05

1 ROC Optimal Cutoff Measure

2 FG Error Foreground Percentage Error

3 BG Error Background Percentage Error

4 The percentage of the overall error compared to the motion pixels only

TABLE 4.5: Assessment on the outdoor motion sequences 018e059s01L using the improved UC algorithm

Thus radius multiplier size 2 will be used for motion sequence 018e059s01L. All the detailed tables for the other sequences are presented in appendix A.2.

The final assessment for the outdoor motion sequences with the chosen radius multipliers is presented in Table 4.6. The table shows the radius multiplier, M_R , the RMSE, the PSNR, and the background foreground percentage error, and the overall error compared to the frames silhouette's motion pixels.

Sequence Number	Number of Frames	M_R	RMSE	PSNR (dB)	FG Error ¹	BG Error ²	Overall Error ³	σ_{RMSE}^2
008e013s00L	100	4	0.202	13.915	19.138	1.740	30.414	1.774E-04
009e017s01L	96	2	0.248	12.102	33.998	2.218	49.529	9.127E-05
010e024s00L	94	2	0.236	12.567	13.808	4.228	39.512	9.564E-05
013e037s00L	158	4	0.153	16.307	12.249	0.933	18.815	6.633E-05
013e040s00L	151	4	0.259	11.731	43.961	1.763	57.197	6.092E-05
017e054s00L	112	2	0.257	11.820	34.688	2.041	47.166	3.969E-04
017e055s00R	88	4	0.213	13.433	24.040	1.720	35.875	1.188E-04
018e059s01L	104	2	0.159	16.005	8.170	1.708	19.937	1.195E-04
018e060s00L	88	4	0.188	14.545	15.117	1.677	25.728	3.786E-05
019e063s05L	112	4	0.156	16.134	15.103	1.165	26.511	1.322E-04
Average			0.207	13.856	22.027	1.919	35.068	1.297E-04

1 FG Error Foreground Percentage Error

2 BG Error Background Percentage Error

3 The percentage of the overall error compared to the motion pixels only

TABLE 4.6: Overall assessment on a number of motion outdoor sequences using the UC algorithm

Motion sequence 013e040s00L gave the highest RMSE, and the minimum PSNR value. While 013e037s00L motion sequence provided the minimum RMSE and the

highest PSNR. For the foreground percentage error, motion sequence 013e040s00L gave the highest percentage error recorded by all the sequences. The range of error here is in the interval of 8.17% to 43.96% with an average of 22.03%. The minimum foreground percentage error was given by 018e059s01L motion sequence. The background percentage error gave an average of 1.92% which is a much lower error than in the foreground error. The error ranged from 0.93% to 4.23%. When looking at the percentage of the overall error over the silhouette's motion pixels, the maximum error was 57.20% while the minimum was 18.82%. The average error here is 35.07%. The RMSE variance is small in the tested sequences which ranged from 3.786E-05 to 3.969E-04. These values illustrate that the improved UC error performance is consistent on outdoor motion sequences.

When comparing Table 4.6 with Table 4.2, the RMSE dropped more than 30% from 0.31 to 0.21. The PSNR also improved from 10.21dB to 13.86dB. The foreground error has increased in the modified UC, from 9.27% to 22.03%, but at the same time the background error decreased, from 9.86% to 1.92%. The overall error also decreased in the modified UC.

Comparing the performance of the improved UC on outdoor motion sequences, Table 4.6, with the performance of the improved SBD, Table 2.8, and the MOG, Table 3.12, yields that the improved UC gave the second best result in the RMSE, the PSNR, and the overall error. The best result was scored by the improved SBD. Though the improved UC has scored the worst result in the foreground region, but it scored the best result in the background region with an error of 1.92% compared to an error of 2.15% and 3.98% in the improved SBD and the MOG classifiers respectively.

Figure 4.6 shows samples of outdoor extracted images. From these samples the algorithm performance did not perform as well as in indoor environment. However, the background region is much cleaner than in the original UC algorithm. All the presented samples showed less error in the background region. Motion sequence 013e040s00L sample showed more error especially in the top left corner where this is the position of a tree leaves moving due to the wind in this sequence.

The shadow is still persistent in three of the four presented samples. Large part of the shadow is showing in motion sequence 008e013s00L. Smaller shadow parts are also showing in 013e040s00L and 017e055s00R. The shadow in 018e060s00L has mostly disappeared.

Foreground detection has deteriorated compared to the original UC. This is due to the enlargement of the background radius sphere which resulted in classifying motion pixels erroneously as background pixels. Foreground region of sequence 018e060s00L sample has less deterioration effects.

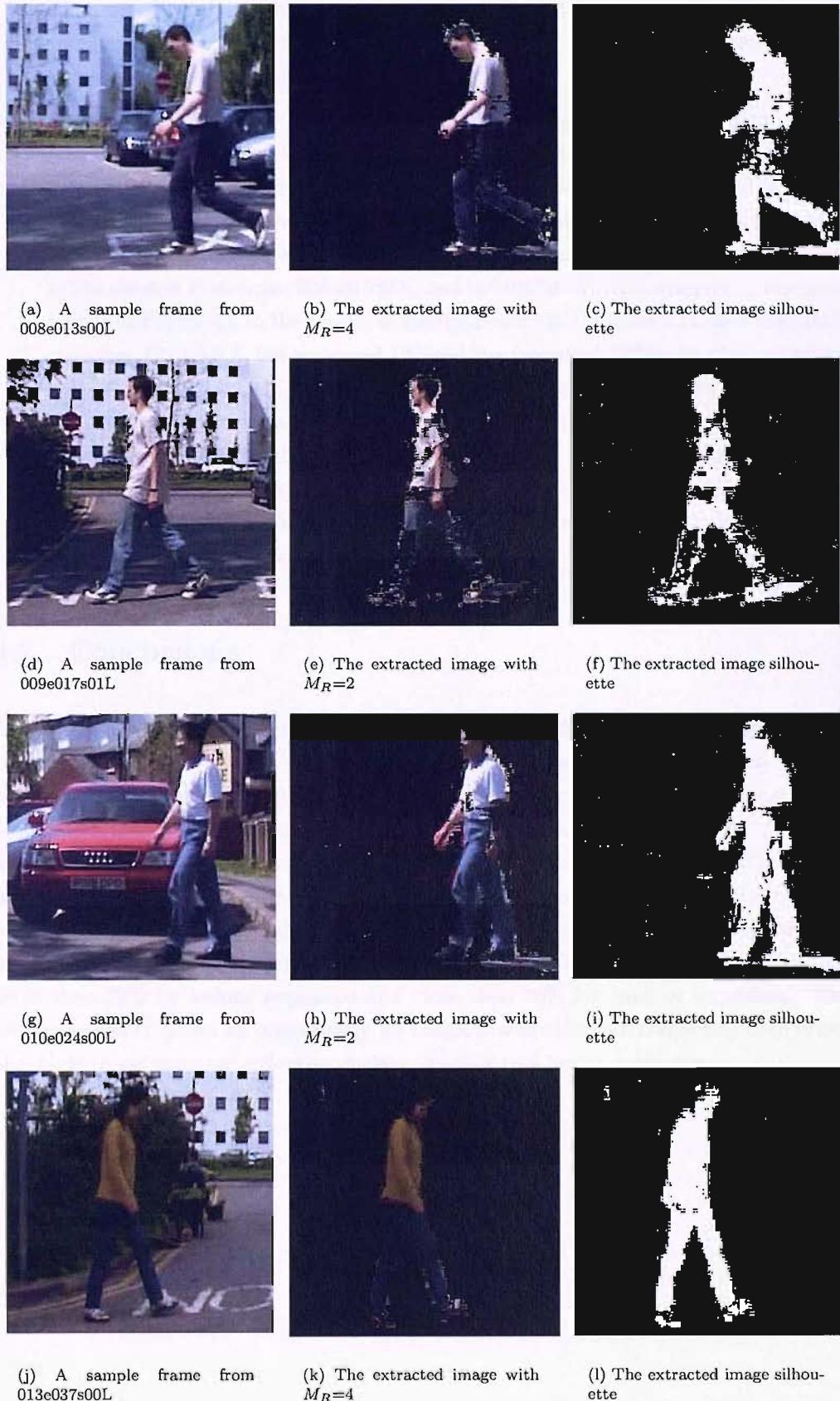


FIGURE 4.6: Examples of outdoor images extracted using the improved UC algorithm

The overall result on the outdoor motion sequences is positive. Overall a better extraction was achieved for the motion sequences but with a disadvantage of loosing accuracy in moving subject (foreground) extraction.

Comparing the outdoor extraction of the improved UC, Figure 4.6, with the extraction of the original UC, Figure 4.4, shows a large improvement in the overall extraction. Though, the moving subject extraction holes increased in all the samples, but the background noise had dropped substantially. We notice the background became cleaner with only few isolated noise in addition to a slight decrease in the shadow in samples 008e013s00L and 009e017s01L. Also comparing the same results of Figure 4.6 to the results of the improved SBD, Figure 2.11, and the MOG classifier, Figure 3.6, the improved UC and the improved SBD gave the best background result with only few isolated noise pixels. Some parts of the shadow were removed by the improved SBD and the improved UC (in sequences 008e013s00L and 013e037s00L). In the extraction of the moving subject's quality, the MOG gave the best results (in all the sequences except 013e037s00L sample) where the improved SBD and the improved UC extracted subjects had more holes (improved UC holes are larger than in the improved SBD).

4.7 Conclusions

In this chapter we have developed a new classifier algorithm for motion extraction, the Unary Classifier. The Unary Classifier performed well in extracting the motion region but failed in performing as well in the background region. The indoor/outdoor sequences shown excessive noise in the background with almost a 10% of error (9.86% for outdoor and 8.02 for indoor). The overall error exceeded 100% of the silhouette's pixels in the indoor sequences and over 70% for the outdoor sequences. We provided an improved version for the UC algorithm. The improvement reduced the RMSE noticeably in both the indoor and the outdoor motion sequences. The error dropped on average more than 70% for indoor sequences and more than 30% for outdoor sequences. The UC performance leaves an opportunity for enhancements through combining with other classifiers to get a better collective decision leading to a better extraction.

Chapter 5

Shadow Suppression using Invariant Colour Models

5.1 Motivation

In the process of distinguishing between motion pixels and background pixels, the pixels containing shadow can appear as a major area of misclassification for many motion extraction algorithms (the mixture of Gaussians algorithm and the unary classifier algorithm are clear examples for such a problem). The reason behind the difficulty of correctly classifying shadow pixels as background pixels is due to the fact that those pixels have suffered from a change in luminance resulting in a change in intensity. These changes led those pixels to be out of the domain of the background model. The problem of shadow extraction can be solved by statistical measures (Horprasert et al., 2000, 1999). Other research also aimed to accommodate shadow extraction by using colour models that minimise (if not remove) the effect of shadows luminance change to the background pixels (Gevers and Smeulders, 2000; Cheng et al., 2001; Gevers and Smeulders, 1999). Since in this work we consider using different classifiers to strengthen the final outcome of motion extraction, we will consider using different methods for shadow suppression for the same reason. So that techniques as yet without inherent shadow suppression (MOG and Unary Classifier) can have performance in this respect equally to that of the SBD algorithm. In this way, performance is then balanced for future fusion.

In the process of finding a suitable colour model we tried different colour models that were claimed to be effective in this arena. Accordingly, we will compare the RGB colour model with four different colour models to find empirically an effective model to enhance the overall labelling process.

5.2 Colour space

A colour model is an abstract mathematical model that describes a representation of colours as a set of numbers, typically as three or four values or colour components. There are several models used to describe the colour scheme: RGB (Red, Green, and Blue), normalised rgb , HSV (Hue, Saturation, and Value) etc. Each model is derived for specific purposes and has certain advantages over the others. Converting between the different models is generally achieved by a relatively simple mapping. Selecting the best colour space is still one of the difficulties in colour image segmentation (Gauch and Hsia, 1992). In this section we will consider five different colour models.

5.2.1 RGB

Colour is defined as a combination to tristimuli R (red), G (green), and B (blue). From the RGB representation other colour models can be derived. RGB is the most commonly used model for television and for pictures acquired by digital cameras (Cheng et al., 2001).

The brightness value of the scene can be used to represent the three primary colours (R,G,B) through the following equation

$$C = \int_{\lambda} E_c(\lambda) S_c(\lambda) d\lambda \quad (5.1)$$

for $C \in (R, G, B)$ where $E_c(\lambda)$ is the radiance spectrum, $S_c(\lambda)$ is the filter for the colour C and λ is the wave length.

The RGB model can be represented geometrically by a 3- dimensional cube Figure 5.1 where the position of a point is in vector space. White, for instance, can be represented when all the primary colours are at k , where k is the maximum light intensity (when k is set to 1 a unity cube is produced). The derivation of the other colours is shown on the cube.

RGB is suitable for colour display but not good for colour scene segmentation and analysis because of the light correlation among the R, G, and B components (Pietikainen et al., 1996; Littmann and Ritter, 1997). Thus if the intensity changes, all the three components will change accordingly. Also the RGB colour model measurement does not represent colour differences within a uniform scale.

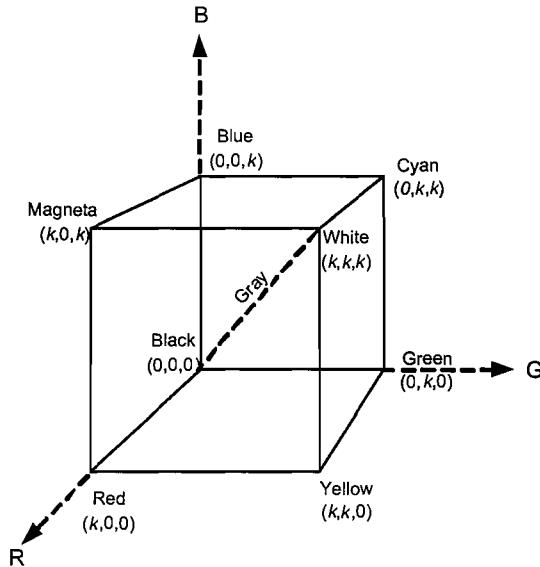


FIGURE 5.1: RGB colour cube

5.2.2 Normalised rgb

The normalised rgb model is an efficient method to get the variations of intensities uniformly across the spectral distribution. The normalised colour space is defined as

$$r = R/(R + G + B) \quad (5.2)$$

$$g = G/(R + G + B) \quad (5.3)$$

$$b = B/(R + G + B) \quad (5.4)$$

The model components have to satisfy the following condition

$$r + g + b = 1 \quad (5.5)$$

Due to condition (5.5), we may only use two of the three colour components since the third can be derived from them (Golland and Bruckstein, 1996).

The normalisation process removes intensity information, thus the rgb values are "pure colours" (Störring, 2004). This property led to one of the advantages of the normalised rgb colour system which is its independence to the brightness of the image (Andreadis

et al., 1990). Normalisation also reduces the sensitivity of the distribution to colour variability (Terrillon et al., 1998). The normalised colours are very noisy when pixel values reflect low intensity (Zheng, 2004; Cheng et al., 2001; Pietikainen et al., 1996).

5.2.3 HSV

The HSV (hue-saturation-value) (Rui et al., 1996; Tsang and Tsang, 1996; Kim et al., 1996; Etemadnia and Alsharif, 2003) model is another commonly used colour space in image processing which is more intuitive to human vision. Each component in this space contributes directly to visual perception (Wan and Kuo, 1998). In HSV each axis can be quantised independently which makes this colour space very useful (Park et al., 1999). Wan and Kuo (Wan and Kuo, 1996) concluded that a colour quantisation scheme based on HSV space performed better than one based on RGB colour space. There are many different variations of HSV model, such as HSB (hue-saturation-brightness) (Tepichin-Rodriguez et al., 1995), HSL (hue-saturation-lightness), and HSI (hue-saturation-intensity) (Carron and Lambert, 1994; Kim and Park, 1996).

Hue (H) is an attribute associated with the dominant wavelength in a mixture of light waves. Thus hue represents dominant colour as perceived by an observer. Saturation (S) refers to relative purity or the amount of light mixed with a hue (Gonzalez and Woods, 1992). Value (V) represents intensity. Hue and saturation taken together are called chromaticity, and therefore may be characterised by its intensity, (V), and chromaticity (HS). This characteristic of the HSV model can be used for shadow suppression. Shadow effects mainly the brightness (intensity) of the background. Therefore, using only the (HS) colour components (jointly or separately) should eliminate the shadow effect on the background resulting in extracting the foreground only without the shadow. In this thesis we implemented all the colour models including the HSV as a three dimensional model (including all the colour components). The two dimensional implementation of the (HS) (or even using each one of them alone) will be left for implementation as future work.

A hue-saturation slice of HSV is derived by projecting the surface of an RGB colour cube onto the $R + G + B = 1$ plane. The saturation and hue of a point on the projection are its polar coordinates r and θ with respect to the centre of the projected surface, while the value (V) of all the points on the projection is simply the length of the diagonal of the colour cube projected (Schwarz et al., 1987). The RGB can be transformed to HSV by a standard procedure (Smith, 1978).

5.2.4 $l_1l_2l_3$

Gevers and Smeulders (Gevers and Smeulders, 1999, 1996) proposed this colour model. The colour model was analysed and evaluated with various colour features by colour-metric histogram matching under varying illumination environment.

The colour model is formulated as follows:

$$l_1 = \frac{(R - G)^2}{(R - G)^2 + (R - B)^2 + (G - B)^2} \quad (5.6)$$

$$l_2 = \frac{(R - B)^2}{(R - G)^2 + (R - B)^2 + (G - B)^2} \quad (5.7)$$

$$l_3 = \frac{(G - B)^2}{(R - G)^2 + (R - B)^2 + (G - B)^2} \quad (5.8)$$

where R, G, B are the colour members of the RGB colour model. The experiments applied by Gevers and Smeulders (Gevers and Smeulders, 1999) showed that this colour model is invariant to viewing direction, surface orientation, highlights, illumination direction, and illumination intensity.

The $l_1l_2l_3$ colour space was applied by Sebe and Lew (Sebe and Lew, 2001) with a maximum likelihood approach in a colour-based retrieval algorithm.

5.2.5 $c_1c_2c_3$

This colour model was proposed by Gevers and Smeulders (Gevers and Smeulders, 1999) as a colour invariant model. The same procedure used in the $l_1l_2l_3$ colour space was used by the authors to verify its competence. This colour space formulation is as follows:

$$c_1 = \arctan\left(\frac{R}{\max\{G, B\}}\right) \quad (5.9)$$

$$c_2 = \arctan\left(\frac{G}{\max\{R, B\}}\right) \quad (5.10)$$

$$c_3 = \arctan\left(\frac{B}{\max\{R, G\}}\right) \quad (5.11)$$

The evaluation by the authors showed that, like the $l_1l_2l_3$ model, this model is invariant to viewing direction, surface orientation, highlights, illumination direction, and illumination intensity.

The model was used with an edge detector in (Salvador et al., 2001) for shadow identification and classification. It was also used to segment only cast shadows for both still images and motion sequences in (Salvador et al., 2004). The result of extraction gave an 80-90% segmentation accuracy most of the time in a sequence of 300 frames.

5.3 Colour Model Evaluation

In order to find a better colour invariant colour space that can eliminate and suppress the shadow from being detected as a motion object, we tested the colour spaces: RGB, normalised rbg, HSV, $l_1l_2l_3$, and $c_1c_2c_3$. The evaluation was done through using each colour model with a motion segmenter while monitoring the effect of the colour model on the three pre-labelled regions: moving object, background and shadow. Due to the difficulty in obtaining ground truth data for the shadow region, the SBD algorithm was used to identify and label the shadow region. Any pixel identified by the SBD as shadow is grouped in the shadow region as long as it is not identified in the silhouette as foreground. This is due to the SBD algorithm outcome which can wrongly label motion pixels as shadows. Also shadow pixels are allowed only to be in the area under the knee level in the sequences used. The background appearing above the knee was at too far a distance from the camera for shadows to affect it, with the lighting arrangement used. This will avoid having pixels that are supposed to appear in the background region being mislabelled as shadows.

The Unary Classifier was used as a motion segmenter to test the colour spaces. The parametrisation of the Unary Classifier is much simpler with less parameters (compared to the MOG algorithm) suggesting the bias associated with parameter choice can be mitigated, thus ensuring focus on shadows alone.

5.3.1 Parameterising Unary Classifier

The soft margin coefficient, ν , set to 0.01 (to include 99% of the background pixels in the background model). Different settings of σ were used: 1, 10, 100 on an indoor motion sequence, 018a060s00L, with a dimension of 240 x 367 pixels. The σ values were chosen initially large to examine the effect of such size on the performance of the UC in accommodating the background pixels variations. Also the same σ values were used on an outdoor motion sequence, 013e037s00L, with the same dimensions used in the data sets. The test was done using all the colour models being tested. The result for the indoor sequence is shown in Table 5.1.

The table shows that increasing the standard deviation from 1 to 100 have minor effect on the result on colour models: $c_1c_2c_3$, $l_1l_2l_3$, Normalised rbg and RGB. HSV colour model is the only colour model affected by changing the kernel σ . The HSV colour

Colour Model	σ	RMSE	PSNR (dB)	FG Error ¹	BG Error ²	Shadow Error ³	Overall Error ⁴	σ_{RMSE}^2
$c_1c_2c_3$	1	0.323	9.827	0.202	10.003	49.989	64.301	4.94E-05
	10	0.322	9.838	0.203	9.969	49.969	64.132	4.86E-05
	100	0.321	9.878	0.202	9.850	49.911	63.539	4.92E-05
HSV	1	0.595	4.513	0.001	39.063	99.983	224.435	5.36E-04
	10	0.402	7.936	2.002	13.487	99.863	103.794	6.83E-04
	100	0.368	8.715	2.405	9.980	99.856	87.703	8.27E-04
$l_1l_2l_3$	1	0.262	11.643	1.536	7.377	16.847	42.436	6.46E-05
	10	0.262	11.652	1.537	7.358	16.851	42.370	6.50E-05
	100	0.255	11.860	1.538	6.938	16.881	40.358	6.57E-05
norm_rgb	1	0.336	9.489	0.131	10.695	55.686	69.786	5.48E-05
	10	0.335	9.497	0.131	10.670	55.657	69.649	5.39E-05
	100	0.341	9.347	0.131	11.187	55.644	72.082	5.21E-05
RGB	1	0.366	8.751	0.053	10.427	99.384	86.236	6.36E-04
	10	0.366	8.763	0.053	10.382	99.374	86.021	6.41E-04
	100	0.364	8.798	0.053	10.242	99.381	85.328	6.37E-04

1 FG Error Foreground Percentage Error

2 BG Error Background Percentage Error

3 BG Shadow Percentage Error

4 The percentage of the overall error compared to the motion pixels only

TABLE 5.1: The effect of changing the standard deviation on an indoor sequence using different colour models

model overall error decreased as σ was increased (inversely proportional). Nevertheless, the fact that the shadow error in HSV is maintained close to 100% even when σ was changed, indicates that continuing to increase σ for this model is pointless since we are targeting a colour model that suppresses shadows. Also the foreground error increased as σ increased.

From the table, choosing any σ value will keep the colour model error order the same. σ will be set to 10 for indoor sequences. A further discussion of the specific amount of error for each region will be discussed in the next section with more sequences used in the evaluation process.

The results for the outdoor motion sequence with different settings to the kernel standard deviation is shown in Table 5.2. From the table, the amount of error decreased as σ increased. However, in most of the colour models the change in the error is minimal, 1-2% (notice: $c_1c_2c_3$, Normalised rgb and RGB). The change is only noticeable when σ changed from 1 to 10 in HSV and $l_1l_2l_3$ and also when σ changed from 10 to 100 in HSV. Continuing to increase the kernel σ for HSV might decrease the percentage error value for HSV. But as in the indoor sequences, the shadow error is large and changes little with increase in σ . Since we are looking for a colour model that suppresses the shadow, continuing to increase the standard deviation is of no benefit to our goal. The error has mostly given a stable value on a σ of 10 for most of the colour models therefore we will use it for a more thorough investigation using more outdoor motion sequences.

Colour Model	σ	RMSE	PSNR (dB)	FG Error ¹	BG Error ²	Shadow Error ³	Overall Error ⁴	σ_{RMSE}^2
$c_1c_2c_3$	1	0.277	11.161	3.976	7.947	18.161	61.643	1.33E-04
	10	0.275	11.216	3.981	7.835	18.147	60.877	1.34E-04
	100	0.270	11.369	4.002	7.533	17.801	58.776	1.32E-04
HSV	1	0.563	4.996	0.614	34.424	99.775	254.569	1.01E-04
	10	0.358	8.935	7.194	11.383	97.611	102.949	1.43E-04
	100	0.326	9.741	8.878	8.636	96.310	85.591	1.47E-04
$l_1l_2l_3$	1	0.341	9.348	27.648	9.267	15.319	93.927	2.28E-04
	10	0.319	9.935	29.754	7.292	14.231	82.265	3.51E-04
	100	0.313	10.093	30.417	6.794	13.690	79.409	4.07E-04
norm_rgb	1	0.287	10.851	3.165	8.598	22.954	66.201	1.03E-04
	10	0.285	10.913	3.186	8.459	22.954	65.277	1.04E-04
	100	0.281	11.034	3.205	8.200	22.808	63.494	1.06E-04
RGB	1	0.333	9.567	2.866	9.947	99.822	89.164	2.12E-04
	10	0.331	9.600	3.008	9.832	99.822	88.534	2.33E-04
	100	0.327	9.706	3.013	9.520	99.837	86.398	2.29E-04

1 FG Error Foreground Percentage Error

2 BG Error Background Percentage Error

3 BG Shadow Percentage Error

4 The percentage of the overall error compared to the motion pixels only

TABLE 5.2: The effect of changing σ on a outdoor sequence using different colour models

5.3.2 Indoor Evaluation

Figure 5.2 shows examples of extracting an indoor image with different colour models. Figure 5.2(a) is the input image. Figure 5.2(b) is the silhouette used with black identifying the background region, white is the foreground region and red is the shadow region.

Figure 5.2(c) gives the extraction using $c_1c_2c_3$ colour model. The moving object is extracted with few holes. Part of the shadow has disappeared. The background is not perfectly extracted and many background pixels mislabelled as motion pixels. Later on after comparing the extraction using this colour model with other colour models, this colour model will show as a better option than other colour models though it is not perfect.

The HSV model extraction in Figure 5.2(d) gave a more noisy background. The shadow region is not suppressed at all as it shows in the image. Holes in the moving object appear in areas where darker regions are in-between light areas (like in the folds of the t-shirt of the moving subject).

For the $l_1l_2l_3$ extracted image in Figure 5.2(e), the background is noisy with motion pixels. The noise motion pixels appearing in the background are larger than in the background extracted using other colour models. The extracted moving object is one

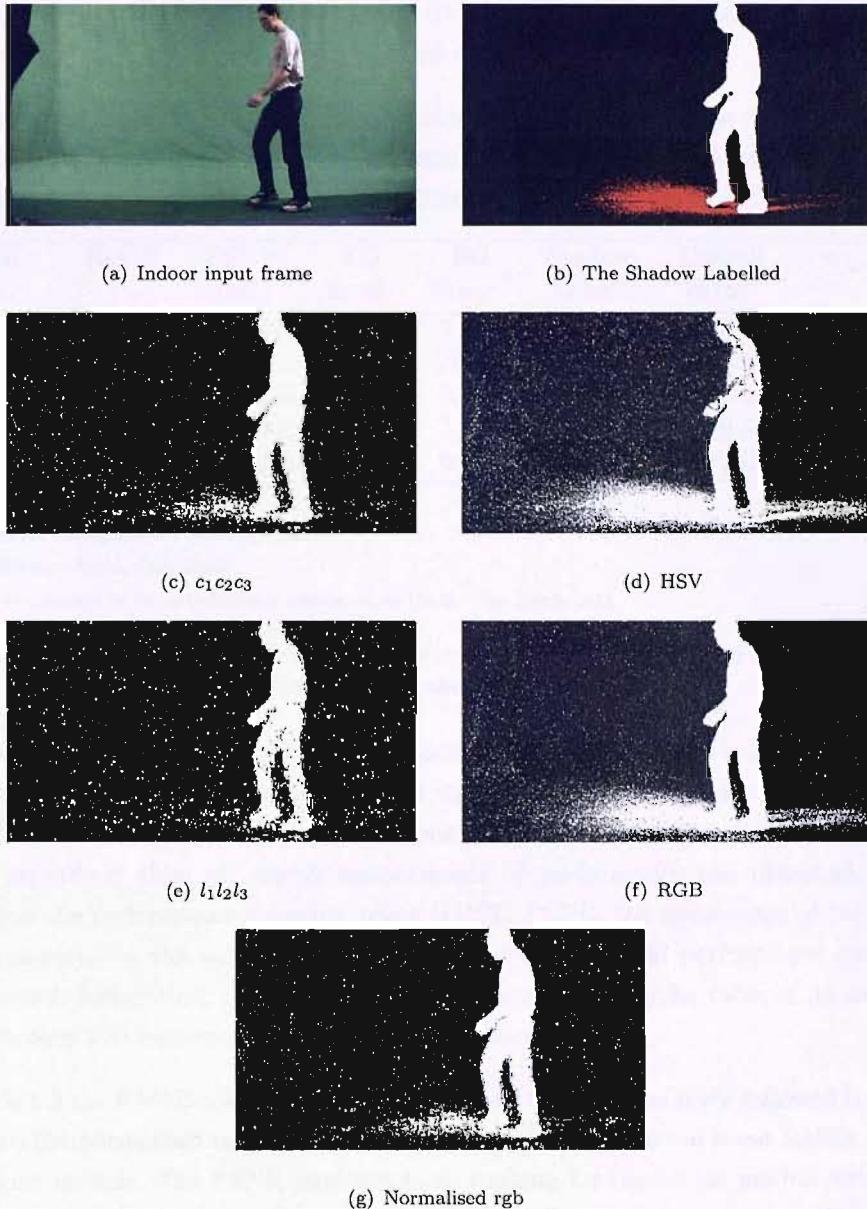


FIGURE 5.2: Indoor images extracted using different colour models.

of the worst compared to the other colour models where holes appear on most of the body (small holes). Also the moving object extracted edges are not as well defined as the other colour models. Even though this model is not a competitor with some of the colour models in extracting the background and the foreground regions, the model gives the best performance in extracting the shadow. From the given sample the algorithm with the $l_1l_2l_3$ succeeded in removing most of the shadow region.

The RGB colour model sample in Figure 5.2(f) reveals that in the given samples this colour model is the best in extracting the moving object with no holes and with well

extracted edges. The background is noisy. The shadow region as a whole is mislabelled where we can see the shadow region labelled as foreground pixels.

The normalised rgb, Figure 5.2(g), gives a similar performance to the $c_1c_2c_3$ where part of the shadow is not suppressed, the background is noisy with motion pixels, and the moving object is extracted without many holes.

Colour Model	RMSE	PSNR (dB)	FG Error ¹	BG Error ²	Shadow Error ³	Overall Error ⁴	σ_{RMSE}^2
$c_1c_2c_3$	0.273	11.288	0.224	7.011	47.434	118.664	3.941E-05
HSV	0.383	8.360	3.897	13.423	99.896	233.140	1.539E-04
$l_1l_2l_3$	0.242	12.358	2.072	5.848	18.165	93.017	3.613E-05
norm_rgb	0.286	10.905	0.173	7.624	53.313	129.567	4.779E-05
RGB	0.330	9.663	0.076	9.516	99.559	173.046	1.385E-04

1 FG Error Foreground Percentage Error

2 BG Error Background Percentage Error

3 BG Shadow Percentage Error

4 The percentage of the overall error compared to the motion pixels only

TABLE 5.3: Averaged overall assessment of different colour models on 10 motion indoor sequences using the UC algorithm

In order to evaluate the tested colour models we are not going to concentrate on the shadow region only since a colour model that performs well on a region might not necessarily perform as well on other regions. Thus the evaluation was done on each region separately then an overall measurement of performance was obtained. Table 5.3 shows the performance measures using RMSE; PSNR; the percentage of the overall error compared to the silhouette's motion pixels; and regional performance measures: background, foreground, and shadow. Each measurement in the table is an averaged measure over 270 frames using 10 different sequences.

In Table 5.3 the RMSE column shows $l_1l_2l_3$ giving the minimum error followed by $c_1c_2c_3$ and then the normalised rgb colour model. RGB and HSV gave the worst RMSE over all the colour models. The PSNR gave the same ranking for the colour models with $l_1l_2l_3$ giving the best result. In the foreground region, RGB scored the minimum foreground error followed by normalised rgb and $c_1c_2c_3$. HSV and $l_1l_2l_3$ gave the worst result in this region (both colour models samples contained erroneous holes in the moving subject in Figure 5.2). In the background region, $l_1l_2l_3$ colour model gave the minimum error followed by $c_1c_2c_3$ and then the normalised rgb. HSV and RGB colour models were the worst performers in this region. For the shadow region, HSV and RGB gave the worst result with most of the shadow region mislabelled. The normalised rgb and the colour model $c_1c_2c_3$ managed to suppress almost half of the shadow region with 53.31% for the first and 47.43% for the second. The $l_1l_2l_3$ model showed the best result in this region with only 18.17% error. The overall percentage error over the silhouette motion pixels gave the minimum error rank to $l_1l_2l_3$ followed by $c_1c_2c_3$ and then the normalised rgb colour model. HSV and RGB colour models performed worst with the

highest percentage error exceeding twice the amount of the motion pixels for the HSV and 173.05% for RGB.

The overall measurement is mainly affected by the dominance of the background region. It is true that $l_1l_2l_3$ is the best performer in the background and the shadow regions but it is one of the worst in the foreground region. Such an extreme result lowers the expectations on this colour model especially if the quality of extracting the foreground region is a main concern. The $c_1c_2c_3$ colour model appears best as a best option with the second best performance in the RMSE, PSNR, background region, shadow region, and the overall percentage error. Though the consistency of the UC was already shown previously in Chap. 4, the UC illustrates a large consistency even when the colour model is changed (the RMSE variance is small for all the different colour models).

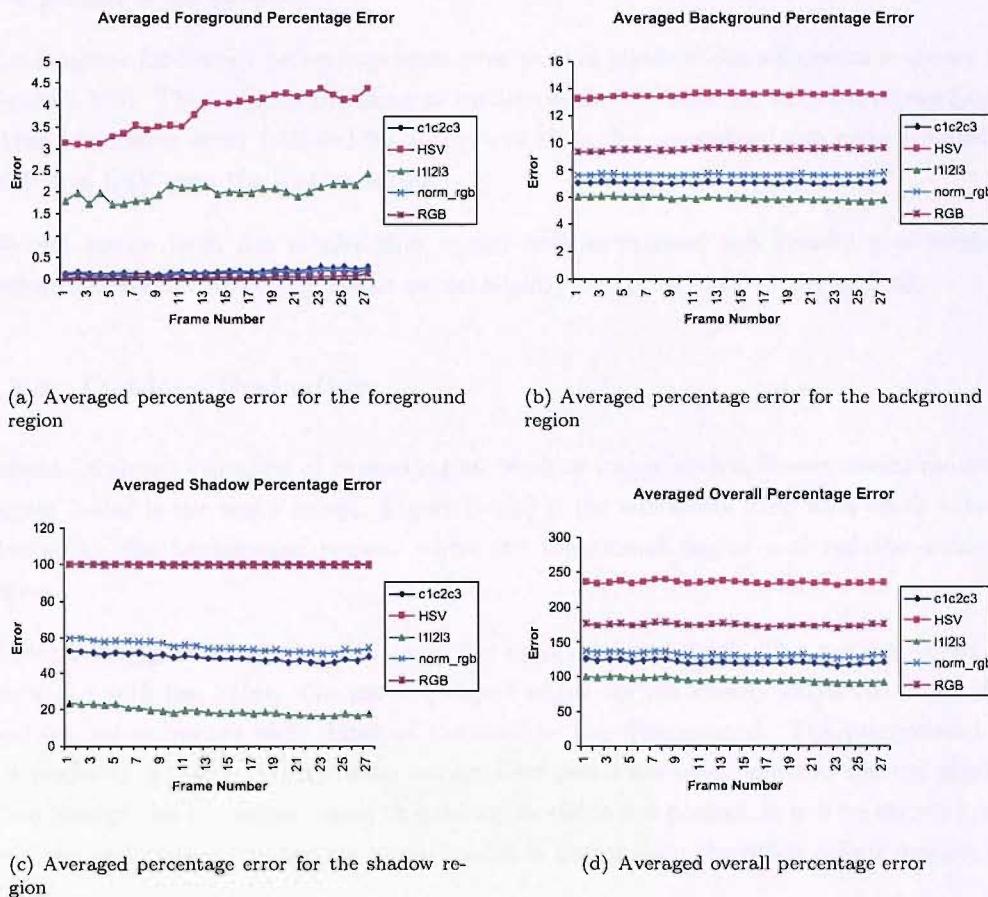


FIGURE 5.3: The averaged error of ten indoor sequences in different regions along with the overall percentage error

Figure 5.3 shows the averaged errors for the different colour models in different regions for 27 frames averaged over 10 different motion sequences.

The foreground percentage error is shown in Figure 5.3(a). While $c_1c_2c_3$, normalised rgb and RGB colour model maintained the extraction with the lowest error, HSV followed by $l_1l_2l_3$ gave the highest percentage error.

Figure 5.3(b) shows the averaged percentage error over the background region. HSV continued to score the highest error as an average over all 10 sequences. RGB scores the second highest percentage error then below that comes the error for the $c_1c_2c_3$ and the normalised rgb, but with small differences. Overall, the $l_1l_2l_3$ model scored the lowest percentage error in this region.

Figure 5.3(c) shows the error in the shadow region. HSV and the RGB colour models gave the highest error in this region. $l_1l_2l_3$ gave the minimum error in the shadow region. $c_1c_2c_3$ and then normalised rgb gave similar results suppressing around 40% to 50% percent of the shadow.

The diagram for overall percentage error over motion pixels in the silhouette is shown in Figure 5.3(d). The result is the same as earlier results in Table 5.3 where it shows $l_1l_2l_3$ giving the lowest error followed by $c_1c_2c_3$ and then the normalised rgb colour models. RGB and HSV gave the highest errors.

We can notice from the results that $c_1c_2c_3$ and normalised rgb usually give similar performance with the $c_1c_2c_3$ colour model slightly better in most of the regions.

5.3.3 Outdoor Evaluation

Figure 5.4 shows examples of extracting an outdoor image with different colour models. Figure 5.4(a) is the input image. Figure 5.4(b) is the silhouette used with black colour identifying the background region, white the foreground region and red the shadow region.

Figure 5.4(c) gives the extraction using the $c_1c_2c_3$ colour model. The moving object is extracted with few holes. The moving object edges are not clearly extracted. Also the feet are not extracted well. Most of the shadow has disappeared. The background is not perfectly extracted where many background pixels are mislabelled as motion pixels. Even though the extraction using this colour model is not perfect, it will be shown later that the performance using this colour model is better than the other colour models.

The background in the HSV extraction in Figure 5.4(d), has more noise than $c_1c_2c_3$. The shadow region is hardly suppressed. More holes appear on the moving object than in the $c_1c_2c_3$ extracted sample.

For $l_1l_2l_3$ sample image in Figure 5.4(e), the extracted moving object is the worst compared to the other colour models where large holes appear on most of the body. The background is noisy with motion pixels. Also the moving object extracted edges are not

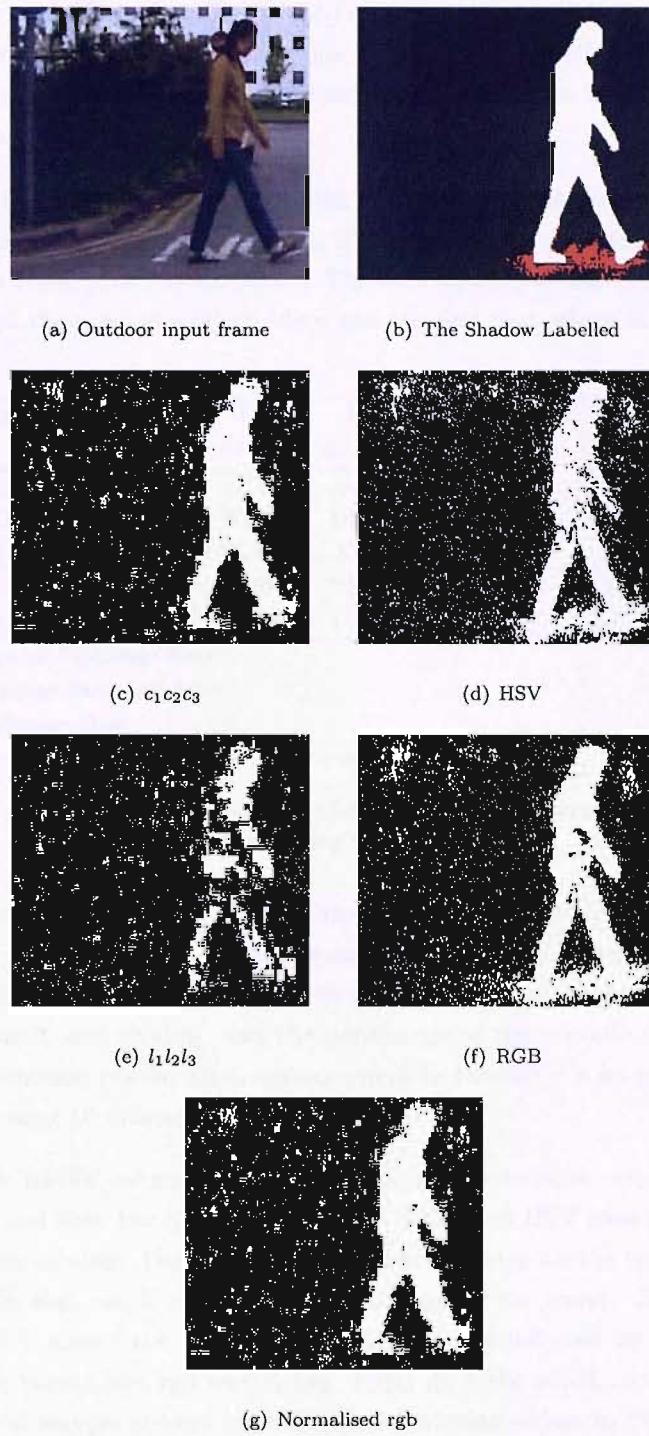


FIGURE 5.4: Outdoor images extracted using different colour models.

as well extracted as the other colour models. This colour model gives the same performance on the shadow region as in indoor sequences where it can extract this region efficiently.

The RGB colour model sample in Figure 5.4(f) shows the moving object with some holes. The background is noisy with motion pixels. This colour model's main weak point is in the shadow region, as it appears in the sample image, where the shadow pixels are resident and labelled as motion pixels.

The same as in the indoor extracted samples, the normalised rgb in Figure 5.4(g) gives a similar performance to the $c_1c_2c_3$ model. The shadow is suppressed well. The background region is noisy with motion pixels. The moving subject extracted well with few holes in it except the moving subject edges and the feet part where it is not extracted well.

Colour Model	RMSE	PSNR (dB)	FG Error ¹	BG Error ³	Shadow Error ²	Overall Error ⁴	σ_{RMSE}^2
$c_1c_2c_3$	0.308	10.278	9.580	9.260	34.275	77.288	1.471E-04
HSV	0.404	7.985	7.528	16.629	99.139	135.106	3.434E-04
$l_1l_2l_3$	0.338	9.464	32.037	8.293	26.661	92.445	1.521E-04
norm_rgb	0.321	9.928	7.967	10.290	40.977	82.768	1.344E-04
RGB	0.365	8.905	4.240	13.673	99.854	109.787	4.036E-04

1 FG Error Foreground Percentage Error

2 BG Error Background Percentage Error

3 BG Shadow Percentage Error

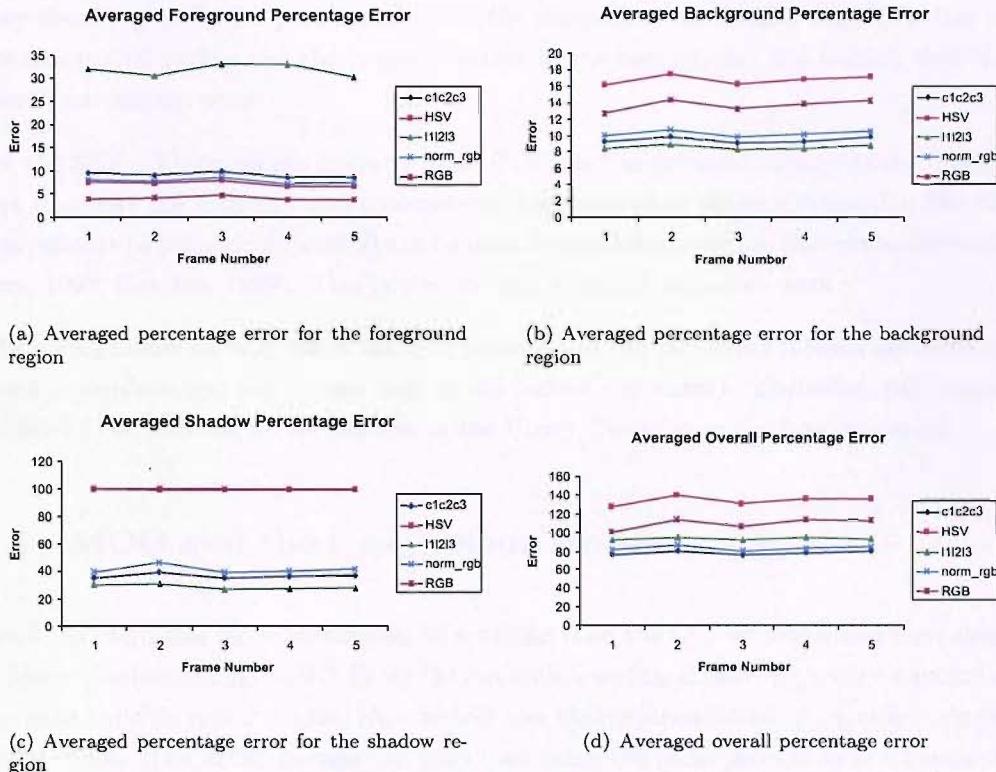
4 The percentage of the overall error compared to the motion pixels only

TABLE 5.4: Averaged overall assessment of different colour models on 10 motion outdoor sequences using the UC algorithm

The same evaluation was done as on the indoor sequences for each region (background, foreground and shadow) and an overall assessment is obtained as well. Table 5.4 shows the performance measures using RMSE; PSNR; regional performance measures: background, foreground, and shadow; and the percentage of the overall error compared to the silhouette's motion pixels. Each measurement in the table is an averaged measure over 50 frames using 10 different sequences.

In Table 5.4, the RMSE column shows $c_1c_2c_3$ giving the minimum error followed by the normalised rgb and then the $l_1l_2l_3$ colour model. RGB and HSV gave the worst RMSE over all the colour models. The PSNR gave the same ranking for the colour models with $c_1c_2c_3$ giving the best result and RGB and HSV giving the worst. In the foreground region error, RGB scored the minimum foreground error followed by the HSV model. Then comes the normalised rgb and $c_1c_2c_3$. $l_1l_2l_3$ gave the worst result in this region (the colour model sample showed big holes in the moving object in Figure 5.4(e)). All the other models gave results with less than 10% error in the foreground region while the $l_1l_2l_3$ colour model value is very large, 32.04%. In the background region, the $l_1l_2l_3$ colour model gave the minimum error followed by $c_1c_2c_3$ and then the normalised rgb. HSV and RGB colour models were the worst performers in this region. For the shadow region, HSV and RGB gave the worst result with most of the shadow region mislabelled. The $l_1l_2l_3$ model showed the best result in this region with only 26.66% of error. $c_1c_2c_3$

followed the top performer in this region with only 34.28% of error. The normalised rgb percentage error then follows with 40.98%. The overall percentage error over the silhouette's motion pixels minimum error is scored by $c_1c_2c_3$ with 77.29% error. The normalised rgb and the $l_1l_2l_3$ followed with an increase of almost 5% for the first and 15% for the second. HSV and RGB colour models both scored more than 100% of error to perform worst of all the colour models. Notice that for all the different colour models with the high overall error recorded, the RMSE variance is small. This means that the UC gives high consistency in motion extraction even if the colour model is changed.



(c) Averaged percentage error for the shadow region

(d) Averaged overall percentage error

FIGURE 5.5: The averaged error of ten outdoor sequences in different regions along with the overall percentage error

Figure 5.5 shows the averaged errors for the different colour models in different regions over 10 different motion sequences.

The foreground percentage error is shown in Figure 5.5(a). While RGB performed the best in this region, $c_1c_2c_3$, normalised rgb and HSV colour models maintained the extraction with the lowest error (less than 10%). $l_1l_2l_3$ gave the highest percentage error.

Figure 5.5(b) shows the averaged percentage error over the background region. HSV maintained to score the highest error. RGB scored the second highest percentage error. Then came the group of normalised rgb, $c_1c_2c_3$ and $l_1l_2l_3$ with the similar and lower error (maintained less than 11%).

For the shadow region, Figure 5.5(c), HSV and the RGB colour models gave the highest error in this region. $l_1l_2l_3$ gave the lowest shadow error. $c_1c_2c_3$ and the normalised rgb gave similar results to $l_1l_2l_3$ colour model with small difference.

The overall percentage error over the motion pixels in the silhouette is shown in Figure 5.5(d). The result is the same as the result in Table 5.4 where it shows $c_1c_2c_3$ giving the lowest error followed by normalised rgb and then the $l_1l_2l_3$ colour models. RGB and HSV gave the highest errors.

The same note can be made on the performance of $c_1c_2c_3$ and the normalised rgb where they usually give similar performance with the normalised rgb usually slightly better in the foreground region and the $c_1c_2c_3$ is better in the background, the shadow and the overall percentage error.

For the HSV, (V) represents intensity while (H,S) are the chromatic components. Therefore to utilise the intensity and chromaticity components in shadow extraction the HS components (separately or jointly) can be used for motion extraction (Gevers and Smeulders, 1999; Bowden, 1999). This procedure will be added as future work.

The $c_1c_2c_3$ model showed one of the best performances in the colour models assessments (best in outdoor and the second best in the indoor sequences). Therefore, this colour model will be selected for further use in the Unary Classifier in the fusion process.

5.4 MOG and the $c_1c_2c_3$ colour model

Since the MOG has more parameters to optimise than the UC, we performed the colour model evaluation using the UC. From the evaluation we found that $c_1c_2c_3$ colour model is the most suitable colour model. Now we will test the performance of $c_1c_2c_3$ colour model on the MOG. The MOG parameters are tuned using the same process as in Chapter 3. The MOG parameters tuning procedure using the $c_1c_2c_3$ colour model with the detailed tests and tables are in Appendix B.

5.4.1 Extraction Results for the MOG using $c_1c_2c_3$ colour model

The final settings for the indoor and the outdoor motion sequences were used to further test more motion sequences.

- **Indoor Motion Sequences**

The 10 indoor sequences were processed using the MOG algorithm with the optimised settings (In each one of them 50-52 frames were used for background modelling). The optimised settings are: learning rate 0.004, the initial weight 0.0005, the initial variance 0.01, 0.4 background threshold and 7 Gaussians per mixture.

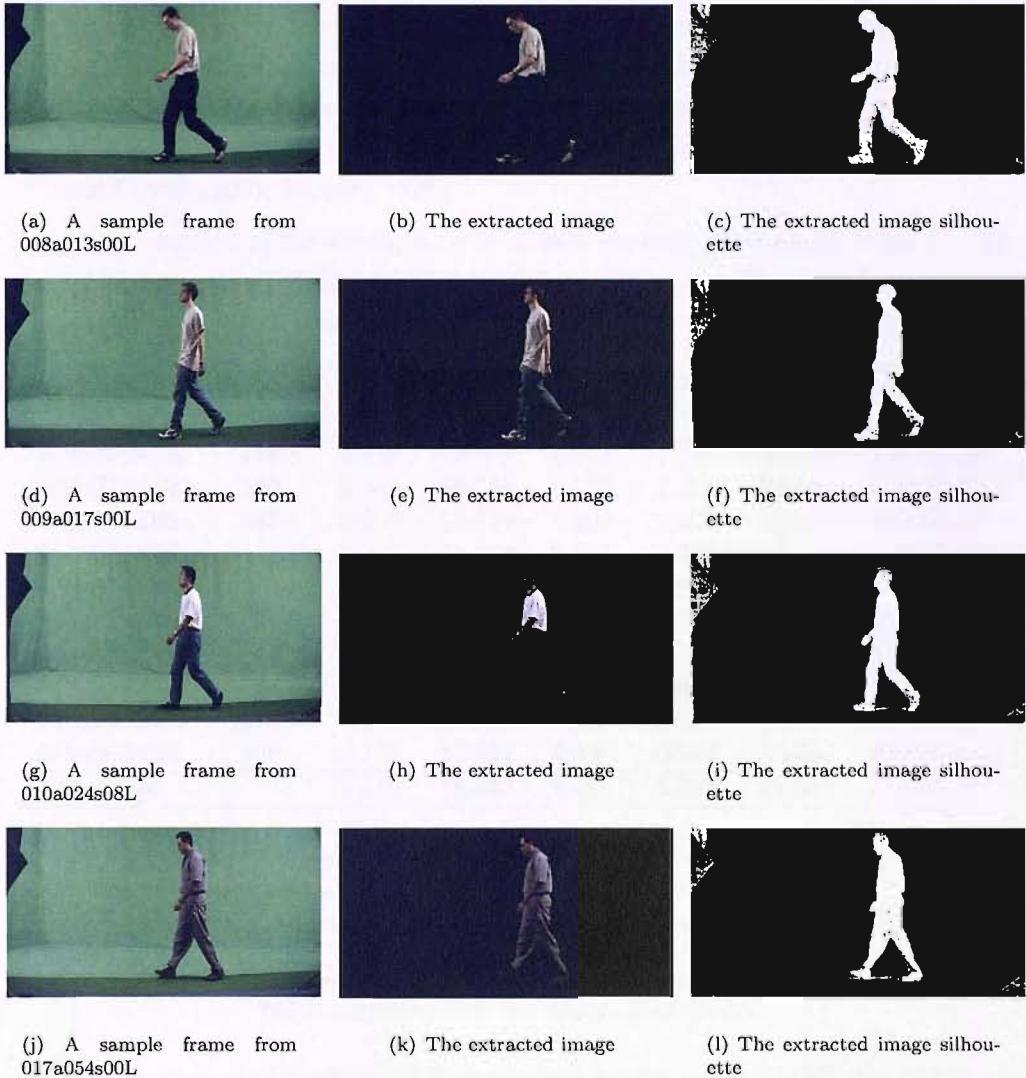


FIGURE 5.6: Examples of indoor images extracted using the MOG algorithm with the $c_1c_2c_3$ colour model.

Samples of the extracted images are shown in Figure 5.6. Looking at the extracted samples in a colour format, in the second column, leads us to conclude that the extraction is optimal. However when turning the result into binary images, the pros and cons of extraction is more obvious. The system performs well in suppressing the shadow in all the shown samples except for small traces in 010a024s08L and 017a054s00L. MOG also performs well in extracting the foreground region with only small holes in the legs of the moving objects. For the background region

the extraction seemed to be noiseless in the coloured view (the second column of images in Figure 5.6). However when looking at the binary view for these images, third column, the extraction process fails to label the top left corner of the frames as part of the background region. This is because the intensity values of this part are very small, close or equal to zero. The same problem is encountered in HSV and normalised colour models. Colours in these colour models becomes unstable near such values where a small perturbation of RGB value might cause a large jump in the transformed values. These value are called singularity values and the essential singularity of normalised coordinates is at black $R = G = B = 0$ (Stokman and Gevers, 2001; Kender, 1976).

The evaluation of the averaged error of each sequence is shown in Table 5.5. The table shows the RMSE; the PSNR; and three different percentage errors, the background, foreground and the overall percentage error compared to the motion pixels.

Sequence Number	No. of Frames	RMSE	PSNR (dB)	FG Error ¹	BG Error ²	Overall Error ³	σ_{RMSE}^2
008a013s00L	178	0.142	16.969	6.598	1.651	27.520	3.981E-05
009a017s00L	169	0.123	18.194	5.772	1.232	23.844	3.940E-05
010a024s08L	187	0.133	17.541	6.204	1.467	28.251	2.813E-05
013a037s00L	114	0.119	18.509	9.282	0.973	26.924	2.496E-05
013a040s00L	184	0.120	18.433	6.816	1.099	24.309	3.458E-05
017a054s00L	188	0.109	19.297	5.750	0.852	17.720	1.924E-05
017a055s00R	162	0.106	19.473	6.142	0.786	17.653	1.637E-05
018a059s00L	188	0.115	18.835	6.983	0.915	20.096	2.617E-05
018a060s00L	179	0.113	18.950	5.308	0.988	19.106	2.827E-05
019a063s00L	186	0.113	18.932	6.030	0.968	20.746	2.442E-05
Average		0.119	18.513	6.488	1.093	22.617	2.814E-05

1 FG Error Foreground Percentage Error

2 BG Error Background Percentage Error

3 The percentage of the overall error compared to the motion pixels only

TABLE 5.5: Overall assessment on a number of motion indoor sequences using the MOG algorithm with the $c_1c_2c_3$ colour model.

Motion sequence 017a055s00R gave the best results in the RMSE, the PSNR, the background error, and the the overall error compared to the silhouette's motion pixels. On the other hand, 010a024s08L gave the worst results in most of the measurement except for the foreground and the background error. The averaged error for the foreground error is 6.49%. The averaged background error is smaller, 1.09%. The overall percentage error compared to the silhouette's motion pixels is 22.62%. The RMSE variance is small in all the tested motion sequences with a maximum value of 3.981E-05 in sequence 008a013s00L. This means that in indoor motion sequences, the $c_1c_2c_3$ MOG performs in a large consistency.

When comparing these results with the results of the indoor MOG using the RGB colour, Table 3.11, the $c_1c_2c_3$ MOG results are smaller in the background and the overall errors. The results of Table 3.11 are 0.84%, 2.73%, and 41.05% for the foreground error, the background error and the overall error respectively. While in Table 5.5, the results are 6.85%, 1.09% and 22.62% for the same regions (foreground, background, and overall error respectively). The RMSE decreased almost 25% from 0.16 to 0.12. The PSNR also improved scoring here 18.51dB while in the RGB colour scored 16.16dB.

- **Outdoor Motion Sequences**

In the process of evaluating the MOG with the $c_1c_2c_3$ colour model, 10 outdoor sequences were tested using this algorithm. 48-55 background frames were used for background adaption. The MOG algorithm with the optimised outdoor parameters were used. The parameters are: learning rate of 0.04, an initial weight of 0.0005, an initial variance of 0.007, 0.4 for the background threshold and 7 Gaussians per mixture.

Figure 5.7 shows samples of outdoor extracted images. From the shown samples $c_1c_2c_3$ MOG with the used settings failed to extract the motion properly. Most of the foreground region was mislabelled especially in motion sequence 009e017s01L. The shadow region seems to be extracted well except for small part of it. The background region has excessive noise.

The evaluation of the averaged error of each outdoor sequence is shown in Table 5.6. The table shows the RMSE, the PSNR, the background/foreground percentage error and the overall percentage error compared to the silhouette's motion pixels.

Sequence Number	No. of Frames	RMSE	PSNR (dB)	FG Error ¹	BG Error ²	Overall Error ³	σ_{RMSE}^2
008e013s00L	100	0.331	9.723	69.099	2.050	82.254	2.863E-03
009e017s01L	96	0.353	9.054	76.855	3.316	100.031	3.811E-04
010e024s00L	94	0.353	9.051	61.725	4.283	88.035	1.702E-04
013e037s00L	158	0.325	9.781	61.111	3.364	84.894	2.337E-04
013e040s00L	151	0.341	9.364	76.594	2.949	98.693	1.762E-04
017e054s00L	112	0.379	8.440	82.023	3.338	102.445	4.085E-04
017e055s00R	88	0.337	9.464	68.843	3.014	89.496	3.019E-04
018e059s01L	104	0.351	9.105	75.844	3.088	97.111	1.593E-04
018e060s00L	88	0.341	9.340	65.465	3.150	85.301	1.904E-05
019e063s05L	112	0.283	10.967	71.267	1.577	86.675	1.089E-04
Average		0.339	9.429	70.882	3.013	91.493	4.822E-04

¹ FG Error Foreground Percentage Error

² BG Error Background Percentage Error

³ The percentage of the overall error compared to the motion pixels only

TABLE 5.6: Overall assessment on a number of motion outdoor sequences using the MOG algorithm with the $c_1c_2c_3$ colour model.

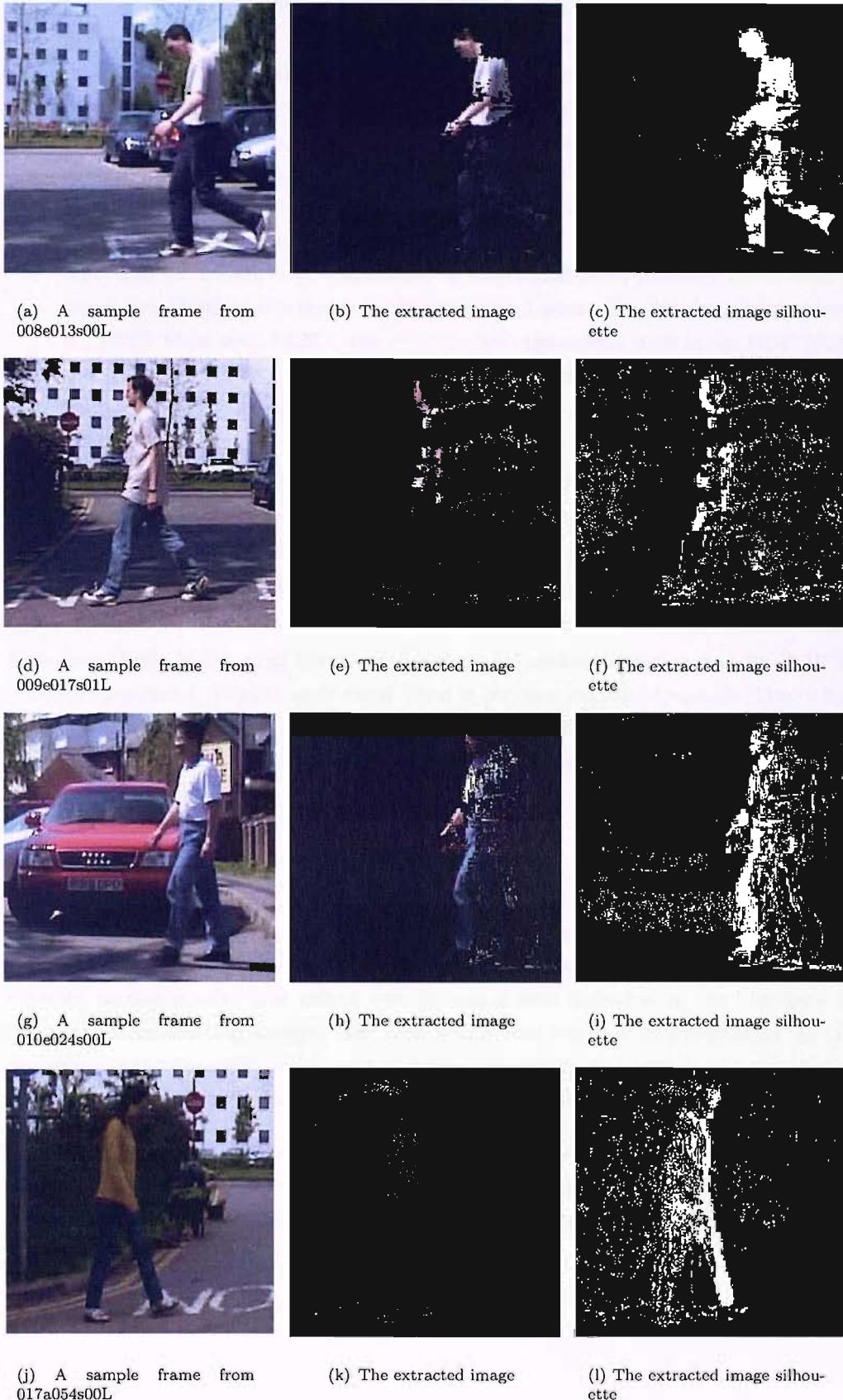


FIGURE 5.7: Examples of outdoor images extracted using the MOG algorithm with the $c_1c_2c_3$ colour model.

The performance of the $c_1c_2c_3$ MOG with the settings used is not satisfying. The averaged RMSE error is 0.34 while the PSNR gave an average of 9.43dB. The motion pixels are mostly misclassified where less than 30% in average are classified correctly. The background region performance was better in numbers where the percentage of error on average was 3.01%. The overall error is large where the average error is close to cover the full amount of the motion pixels, 91.49%. When comparing these results with the results of the outdoor MOG using the RGB colour model, Table 3.11, the amount of RMSE increased by almost 57% (from 0.22 to 0.34) and the PSNR value detriorated by 29% (13.37dB to 9.43dB). The results of the $c_1c_2c_3$ MOG is also larger in the foreground error. The foreground error here is 70.88% while with RGB it was 10.01%. Also the overall error in the RGB MOG was 37.50% compared to 91.49% in $c_1c_2c_3$ MOG. However the background error here is smaller than in the RGB MOG, 3.01% compared to 3.98% in the RGB MOG.

Motion sequence 008e013s00L gives the highest RMSE variance, 2.863E-03. The RMSE variances recorded are still small which means that though $c_1c_2c_3$ MOG gives high overall errors in outdoor sequences, its performance variation is contained in a small range in all the motion sequences, 1.904E-05 to 2.863E-03.

As an overall the MOG using the $c_1c_2c_3$ colour model performed better than RGB MOG in indoor sequences. While it performed worst in outdoor motion sequences. Thus when further combining the MOG with other classifiers, we will use $c_1c_2c_3$ colour model for indoor sequences only. For the outdoor motion sequences we will use RGB colour model.

5.5 Conclusions

In this chapter we presented multiple colour invariant models that can be used with a motion segmenter in order to add the criteria of suppressing shadow pixels from the detected motion pixels. The colour models tested were indicated in the literature as invariant to illumination changes. The Unary Classifier was used as a segmenter. In the assessment procedure different overall measures were used. In addition, the assessment was performed on each region (shadow, background and foreground) individually.

Due to the satisfying performance of the $c_1c_2c_3$ model on outdoor and its reasonable performance in indoor (the second best performer with better foreground performance than the best performer), this colour model will be selected for further use in the Unary Classifier.

An assessment was made on using $c_1c_2c_3$ colour model with the mixture of Gaussians. The MOG performance was enhanced with the $c_1c_2c_3$ in indoor sequences only. Therefore for the mixture of Gaussians $c_1c_2c_3$ colour model will be used only for indoor sequences but for outdoor motion sequences RGB colour model will be used.

Chapter 6

Combining Classifiers

6.1 Introduction

The potency of individual classification is challenged by pattern recognition systems based on combining measures (Valev and Asaithambi, 2001; Leandro Rodrguez-Liales et al., 2003), which can show better classification (Ruta and Gabrys, 2005; Duin, 2002; Chen et al., 1997; Ho et al., 1994; Huang and Suen, 1995). Classifiers that differ in their classification decision can offer complementary information about the patterns to be classified, which can be harnessed to improve performance of the selected classifier (Kittler et al., 1998).

The objective of combination is not to rely on a single decision scheme. Instead, the decisions of single classifiers are combined to derive a consensus decision, where the combination method should enable us to use the benefits and avoid the weaknesses of each classifier in order to achieve the optimal possible performance. Combining identical classifiers will confer no performance benefits at the expenses of increased complexity. On the other hand, different but low performing classifiers are unlikely to bring any benefits in combined performance (Ruta and Gabrys, 2005). It is believed that optimal combined classifiers should be diverse (i.e. with minimum number of coincident failures) and have at the same time good individual performance (Sharkey and Sharkey, 1997). In the previous chapters we have presented three different classifiers. After improving the classifier's performance either by optimising the classifier's parameters, as in the MOG, or by improving the overall design of the technique, as in the SBD and the UC, we succeeded in reaching a reasonable performance for each classifier. In addition, it was highlighted that the performance of each classifier is different in each region and the erroneous noise points were mostly not overlapping (diversity). The techniques are quite different and this should promote diversity.

Another important factor that has to be considered is the combination method itself. Among all the combination methods, simple methods like the Sum, the Product, the

Maximum, and the Median have received much attention (Czyz et al., 2004; Alexandre et al., 2001; Kuncheva et al., 2001; Kuncheva, 2002; Kittler et al., 1998).

Many different studies differ as to which rule is better in performance. Kittler et al. (1998) used different combination schemes namely the Sum, Product, Maximum, Minimum, Median and majority vote rules. They reported that the Sum rule produces the most reliable decisions. They also proved that the Sum is more resilient to error than the Product rule and called the Sum “remarkably robust”. In (Alkoot and Kittler, 1999) a comparison was made between the four simple combination rules (Sum, Product, Maximum, Median) and a single expert decision. The study reached a general result stating that, the results prove the combiners to be better than the single expert, especially the Sum and the Median. Tax et al. (1997) compared the Product and the Mean rules and concluded that the Product rule leads to a better performance when all the classifiers produce small errors. If at least one classifier produces large errors then the Mean rule gives better result (It is well known that a robust estimate of the Mean is the Median (Kittler et al., 1998)). Shakhnarovich and Darrell (2002) experimentally assessed the performance of the Maximum, Minimum, Sum and Product rules for combining face and gait cues. From the experiments the authors reached a conclusion that while the combination almost always improved the classification accuracy of the system, the best performance was produced by using the Product rule. The Minimum rule performed poorly in the experiments and on occasions resulted in lower performance to give an overall performance less than that of the best individual classifier.

Classifier combination was applied to many different applications such as biometrics (gait, face, ear recognition, etc.) (Czyz et al., 2004; Jing and Zhang, 2003; Shakhnarovich et al., 2001; Kale et al., 2004; Chang et al., 2003; Bazin et al., 2005), handwriting recognition (Rahman and Fairhurst, 1997; Xu et al., 1992), speech recognition (Tur et al., 2005), and information retrieval (Lee, 1998; Nottelmann and Straccia, 2005).

In the following sections we will present different new combinations of the SBD, the MOG and the UC classifiers in a novel probabilistic form. When combining only two classifiers the Maximum rule is used for the economy of analysis. For the combination of the three classifiers more elaborate combinations are used. We will use the Sum, the Product, the Maximum and the Median combination rules. We will start by presenting the posterior probabilities derived. Then we will describe the combination methods used to obtain the final decision. Finally we will present the combination results. Notice that the improved SBD and the improved UC are the classifiers used for combination and comparison and we refer to them in this chapter simply as SBD and UC, respectively.

6.2 Bayesian Classification

A natural way to combine probabilistic classifiers is through Bayes theorem. The fact that both algorithms operate using pixel wise operations facilitated the process of combination. The combination of the classifiers using Bayes theorem is given as

$$P(C|\mathbf{x}) = \frac{p(\mathbf{x}|C)P(C)}{p(\mathbf{x})} = \frac{p(\mathbf{x}|C)P(C)}{\sum p(\mathbf{x}|C)P(C)}. \quad (6.1)$$

where C is the class and \mathbf{x} is the tested pixel.

For classifiers that do not have a natural probabilistic output, one can approximate a probability by fitting a logistic function to the output (Bishop, 1996; Platt, 1999; Wahba, 1992). The posterior probability can be expressed using the logistic sigmoid function

$$P(C|\mathbf{x}) = \frac{1}{1 + \exp(-a)}. \quad (6.2)$$

An estimation of a will be given in the following section.

6.3 Probability Estimation

The posterior probabilities for the classifiers used are determined using the logistic sigmoid function formulated as

$$P(C|\mathbf{x}) = \frac{1}{1 + \exp(-(W_C - D_{\mathbf{x}}))} \quad (6.3)$$

where $-(W_C - D_{\mathbf{x}})$ is the decision function used for each algorithm. Expanding the classifiers under consideration (SBD, MOG, and UC) can all be placed in framework 6.3 by “substituting” the appropriate expression for $-(W_C - D_{\mathbf{x}})$.

- **The MOG classifier**

For the MOG algorithm, the background is modelled as Gaussians. Each Background Gaussian is considered as a class. $W_C = 2.5\sigma$ and $D_{\mathbf{x}}$ is calculated as the distance between the the tested pixel and the closest background Gaussian mean. The foreground probability is calculated as follows

$$P(C_{FG}|\mathbf{x}) = 1 - P(C_{BG}|\mathbf{x}) \quad (6.4)$$

- **The SBD classifier**

For the SBD classifier, the values of W_C for the background and the shadow

classes are calculated using the same procedure. The class mean is calculated as the arithmetic mean of the class thresholds (the thresholds are the upper and the lower bound for the class). W_C is then calculated as the distance from one of the class boundaries (thresholds) to the estimated mean.

$D_{\mathbf{x}}$ is calculated as the distance between the brightness distortion parameter, $\hat{\beta}_i$, for a tested pixel and the class mean, refer to Chapter 2 for the details of this parameter. The foreground probability is calculated as in the MOG classifier.

- **The UC classifier**

To calculate $(W_C - D_{\mathbf{x}})$ for the UC we used the decision function formulated in Chapter 4

$$(W_C - D_{\mathbf{x}}) = \left((R_{UC} + NR_{UC})^2 - \sum_{i,j}^{\ell} \alpha_i \alpha_j k(\mathbf{x}_i, \mathbf{x}_j) + 2 \sum_i^{\ell} \alpha_i k(\mathbf{x}_i, \mathbf{x}) - k(\mathbf{x}, \mathbf{x}) \right) \quad (6.5)$$

6.4 Combination Rules

In this section, we present the different combination rules that use posterior probabilities. We use the Bayes decision rule as the principle for combining the classifiers. The Bayes rule assigns the pixel to the class with the maximum posterior probability.

$$f(x) = \begin{cases} C_a & \text{if } P(C_a|\mathbf{x}) > P(C_b|\mathbf{x}) \\ C_b & \text{otherwise.} \end{cases}$$

When combining different classifiers with independent conditional probabilities the product rule can be used

$$f(x)_{Product} = \begin{cases} C_a & \text{if } \prod_i P(C_a|\mathbf{x}_i) > \prod_i P(C_b|\mathbf{x}_i) \\ C_b & \text{otherwise.} \end{cases}$$

where i represents the used classifier.

When the conditional probabilities are not conditionally independent, this rule is violated. For sufficiently accurate classifiers, scores are likely to be positively correlated because the classifiers will agree on the majority of classifications and classify them correctly. In addition to the product rule, several other combination rules can be used and may be more appropriate when score independence is not satisfied (Czyz et al., 2004)

$$f(x)_{Sum} = \begin{cases} C_a & \text{if } \sum_i P(C_a|\mathbf{x}_i) > \sum_i P(C_b|\mathbf{x}_i) \\ C_b & \text{otherwise} \end{cases}$$

$$f(x)_{Maximum} = \begin{cases} C_a & \text{if } \max_i P(C_a|\mathbf{x}_i) > \max_i P(C_b|\mathbf{x}_i) \\ C_b & \text{otherwise} \end{cases}$$

$$f(x)_{Median} = \begin{cases} C_a & \text{if } \text{med}_i P(C_a|\mathbf{x}_i) > \text{med}_i P(C_b|\mathbf{x}_i) \\ C_b & \text{otherwise.} \end{cases}$$

In experiments, the Maximum rule will only be used when combining two classifiers for economy of analysis. For combining three classifiers, we will investigate all four combination rules.

6.5 Weight averaging the classifiers decisions

The combination of the classifiers can be optimised by adding a weighting factor to the overall probability of classifiers (Kittler et al., 1997; Kittler and Hojjatoleslami, 1998). The weight functions as a confidence factor for each classifier. The weight is conditioned to $\sum_i \omega_i = 1$. The combination rules become

$$f(x)_{Product} = \begin{cases} C_a & \text{if } \prod_i [\omega_i P(C_a|\mathbf{x}_i)] > \prod_i [\omega_i P(C_b|\mathbf{x}_i)] \\ C_b & \text{otherwise} \end{cases}$$

$$f(x)_{Sum} = \begin{cases} C_a & \text{if } \sum_i [\omega_i P(C_a|\mathbf{x}_i)] > \sum_i [\omega_i P(C_b|\mathbf{x}_i)] \\ C_b & \text{otherwise} \end{cases}$$

$$f(x)_{Maximum} = \begin{cases} C_a & \text{if } \max_i [\omega_i P(C_a|\mathbf{x}_i)] > \max_i [\omega_i P(C_b|\mathbf{x}_i)] \\ C_b & \text{otherwise} \end{cases}$$

$$f(x)_{Median} = \begin{cases} C_a & \text{if } \text{med}_i [\omega_i P(C_a|\mathbf{x}_i)] > \text{med}_i [\omega_i P(C_b|\mathbf{x}_i)] \\ C_b & \text{otherwise.} \end{cases}$$

We will also use this method to combine classifiers. The classifier weight is derived from the overall performance on the foreground/background extraction expressed as a percentage in comparison with the other techniques.

6.6 Experimental results

The experimental results are implemented by combining two and three classifiers.

6.6.1 Two Classifiers Combination

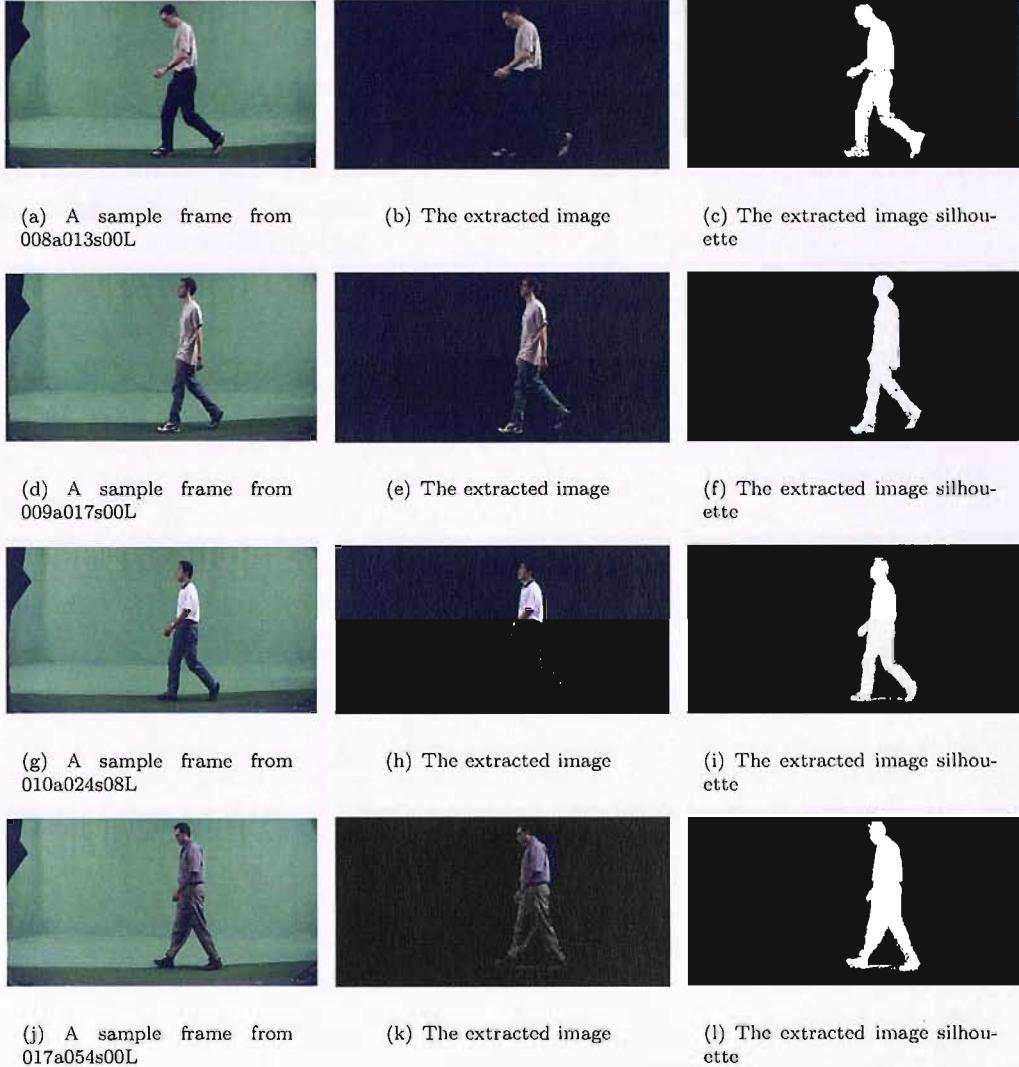


FIGURE 6.1: Indoor images extracted using the combined UC and SBD classifiers

Using the weighted Maximum rule, the UC, the SBD and the MOG are combined in pairs. To choose the values of the weights, N Motion samples (non-background) are used for each motion sequence. Then a search is performed using gradient decent approach for the weights until optimal values are found that produces the minimum overall error. 10 samples are used for indoor sequences and 5 for outdoor sequences.

Figures 6.1 shows the results of combining the UC and the SBD on indoor motion sequences. The samples show a clean background with no noticeable noise. Most of the shadow has disappeared. 010a024s08L motion sequence gave small traces of shadow. The sample of 017a054s00L shows a larger shadow but still as an overall result most of the shadow had disappeared even in this sample. For the moving subject extraction, most of the extracted samples are extracted well with few small negligible holes.

When comparing the images shown with the samples of the original classifiers, Figures 4.5 and 2.10, the combined results outperformed the original SBD extraction especially in the foreground region where in the SBD large holes appeared mostly in the legs of the moving subject. In the combined UC/SBD samples, few holes can be observed and with much smaller size. The result of the UC and the combined UC/SBD is comparable in all the regions.

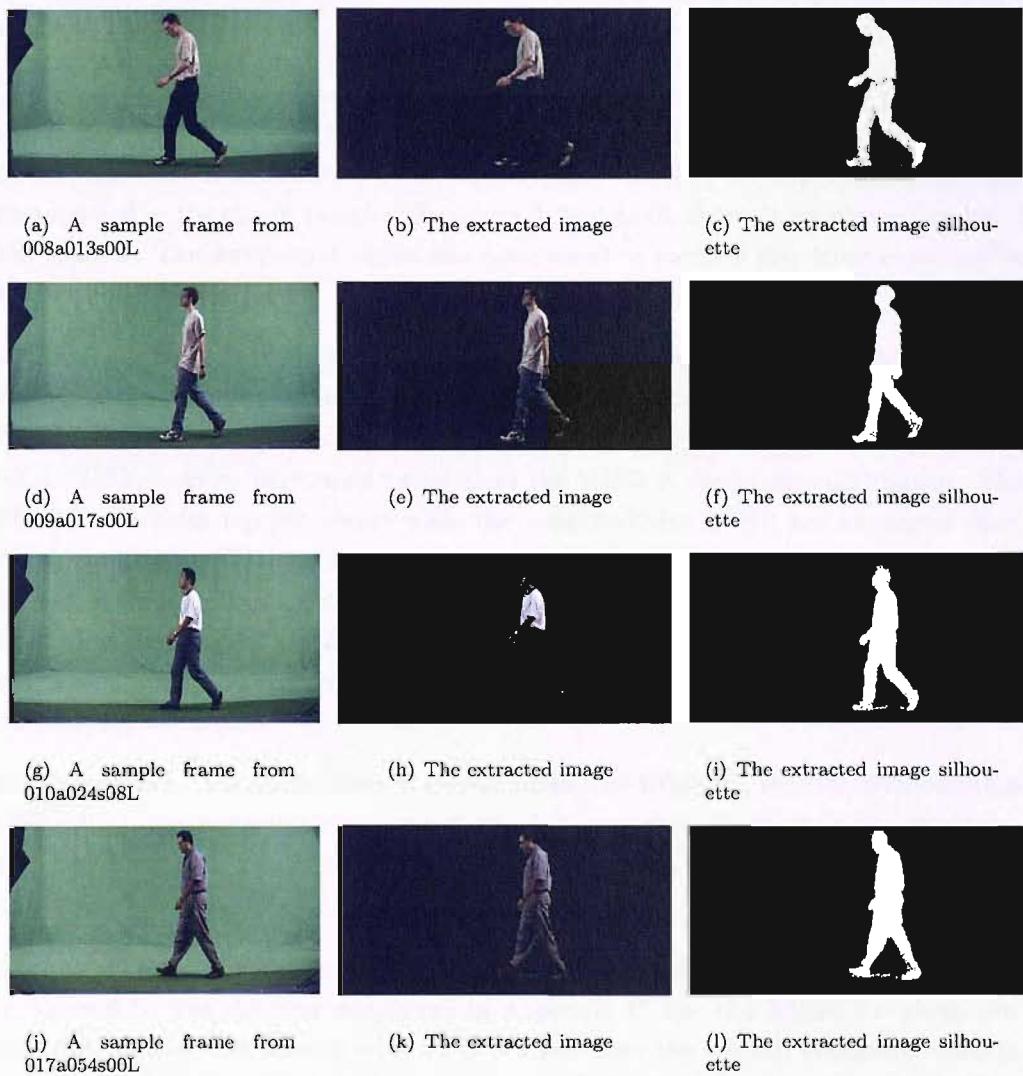


FIGURE 6.2: Indoor images extracted using the combined UC and MOG classifiers

Figure 6.2 represents samples of extraction of indoor motion sequences using the combined UC/MOG classifiers. The combined UC/MOG managed to label the background and the shadow regions correctly with traces of shadows showing in the samples of 010a024s08L and 017a054s00L motion sequences. The foreground extraction is fairly well extracted with some small holes in the samples.

The combined results managed to overcome the problems of the MOG extraction where it misclassifies areas of low intensities. In the MOG the top left (the black curtain) is a noisy area but the combined UC/MOG classified it correctly. The foreground extraction of the MOG suffered from a large number of holes especially in the legs area. The combined UC/MOG extraction of the foreground region is better than in the MOG with less holes. The extraction of the combined UC/MOG is similar to the UC classifier extraction in all the region except for the shadow region where the combined UC/MOG performed better in this region. Clearly, the combined classifier enjoys the advantages of the different classifiers.

The extraction of the combined MOG/SBD classifiers for indoor motion sequences is shown in Figure 6.3. In the background region the new classifier managed to label most of the region pixels correctly. For the shadow region, most, if not all, of the shadow has disappeared in the shown samples. Sequence 017a054s00L shows more obvious traces of the shadow. The foreground region has some small to medium size holes especially in the area of the legs.

Comparing the combined results of the MOG/SBD classifier of Figure 6.3 with the results of the original techniques, the MOG in Figure 3.5 and the SBD in Figure 2.10, the combined classifier's overall performance is better than the other two. The combined MOG/SBD classifier performed better than the MOG in the background region. The MOG had a noisy top left corner while the combined MOG/SBD has an overall clean background. The SBD also behaved well in extracting the background region. For the foreground region, the combined MOG/SBD has some small holes, especially in the area of the legs, but the MOG has more small holes and the SBD suffered even from larger holes. For the shadow region their behaviour is similar where they all behave well in suppressing this region.

When comparing the three different combinations, the UC/SBD, the UC/MOG and the SBD/MOG, the extraction results are fairly similar particularly in the background and the shadow regions. In the foreground region the combined MOG/SBD has more smaller holes than the other two.

The overall result for combining two classifiers using indoor motion sequences is shown in Table 6.1. The detailed results are in Appendix C. For the RMSE the three new classifiers gave better results, with a result lower than the original classifiers. Also in the PSNR, the results of the new classifiers are higher than for the original classifiers. The combined classifiers all scored a PSNR value above 23.2. In the foreground region,

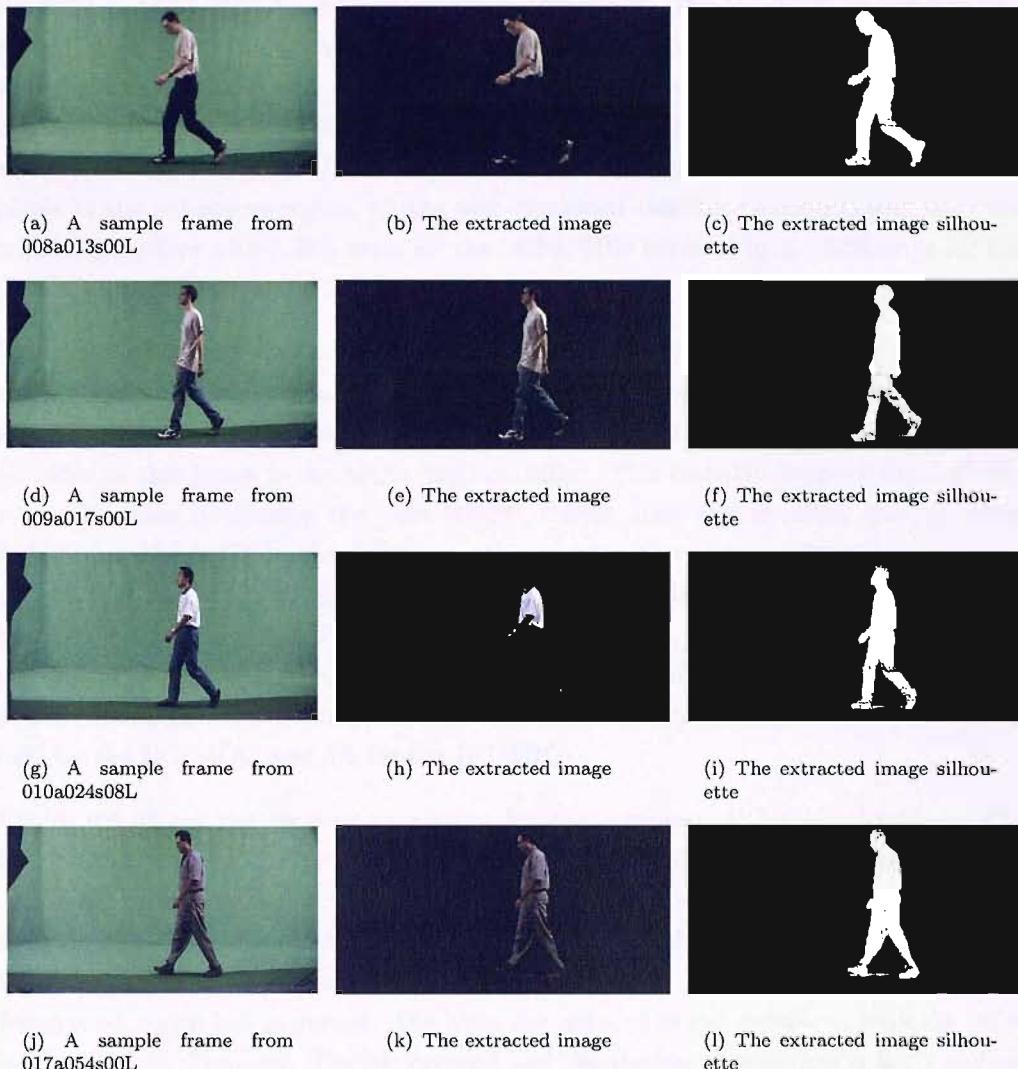


FIGURE 6.3: Indoor images extracted using the combined MOG and SBD classifiers

Combined Classifiers	RMSE	PSNR (dB)	FG Error ¹	BG Error ²	Overall Error ³	σ_{RMSE}^2
UC and SBD	0.070	23.201	2.805	0.331	7.661	3.188E-05
UC and MOG	0.069	23.267	3.296	0.287	7.521	3.214E-05
MOG and SBD	0.068	23.385	4.787	0.174	7.330	4.109E-05
SBD	0.078	22.342	2.659	0.491	9.775	4.047E-05
MOG	0.119	18.513	6.488	1.093	22.617	2.814E-05
UC	0.072	22.872	2.801	0.373	8.265	3.394E-05

¹ FG Error Percentage Error² BG Error Percentage Error³ The percentage of the overall error compared to the motion pixels only

TABLE 6.1: The overall assessment of combining three sets of two combined classifiers using indoor motion sequences

the SBD and the UC gave the best results in this region. The UC/SBD classifier scored a result close to the UC results. The worst result was scored by the MOG with a 6.49% error. For the background region, all the three combinations scored better than the original classifiers with 0.17% error for the MOG/SBD, 0.29% error for the UC/MOG and 0.33% error for the UC/SBD. Finally, in the overall error, compared to the motion pixels in the silhouette region, all the new combined classifiers scored better than the original classifiers with 7.33% error for the MOG/SBD followed by a 7.52% error for the UC/MOG and 7.66% error for the UC/SBD.

The averaged RMSE variance is small for all the provided classifiers which means that each classifier by itself provides high consistency in its performance. Utilising this accuracy in comparing the classifiers using the RMSE, we can say that the most successful classifier in this group is the MOG/SBD classifier. This classifier outperformed all the other classifiers by scoring the least RMSE, 0.068. Also this classifier gave a better fusing model by utilising the differences of its originator and outperforming them with a significant amount (compared to the others). The MOG/SBD managed to give better results than the SBD by 13% and better than the MOG by 43%. The UC/MOG and the UC/SBD gave also reasonable RMSE value, 0.069 and 0.70 respectively, but those two classifiers gave a small improvement difference from one of their originator, the UC (4% for the UC/MOG and 3% for the UC/SBD).

Figure 6.4 shows the outdoor extraction for the combined UC/SBD classifier. The background region is well extracted with few noise pixels. The foreground region is extracted with some holes (of varying size). The shadow is resident in most of the sequences except sample 013e037s00L. When comparing the result of this classifier with the UC, Figure 4.6, and the SBD classifiers, Figure 2.11, the quality of extracting the foreground region has improved. The holes are reduced in size compared with the holes by the original classifiers. The background and the shadow suppression is fairly similar in the UC/SBD combined classifier and its originating classifiers.

Figure 6.5 presents the outdoor extraction of the combined UC/MOG classifier. The foreground is extracted with varying quality of extraction. The extraction of 013e037s00L resulted in extracting most of the foreground region pixels with few holes. On the other hand, the foreground extraction of sequence 008e013s00L sample resulted in some large holes.

Compared with the original classifiers result, the UC in Figure 4.6 and the MOG in Figure 3.6, the background suppression for the combined UC/MOG classifier is better than the MOG and similar to the UC in most of the samples (sample 009e017s01L of UC/MOG showed a noisy background). For the shadow region, the MOG classifier samples all showed shadows (the MOG here uses the RGB colour model which is not a colour invariant model). The combined classifier managed to suppress the shadow similar to the UC. For the foreground region, the MOG was the best performer in one

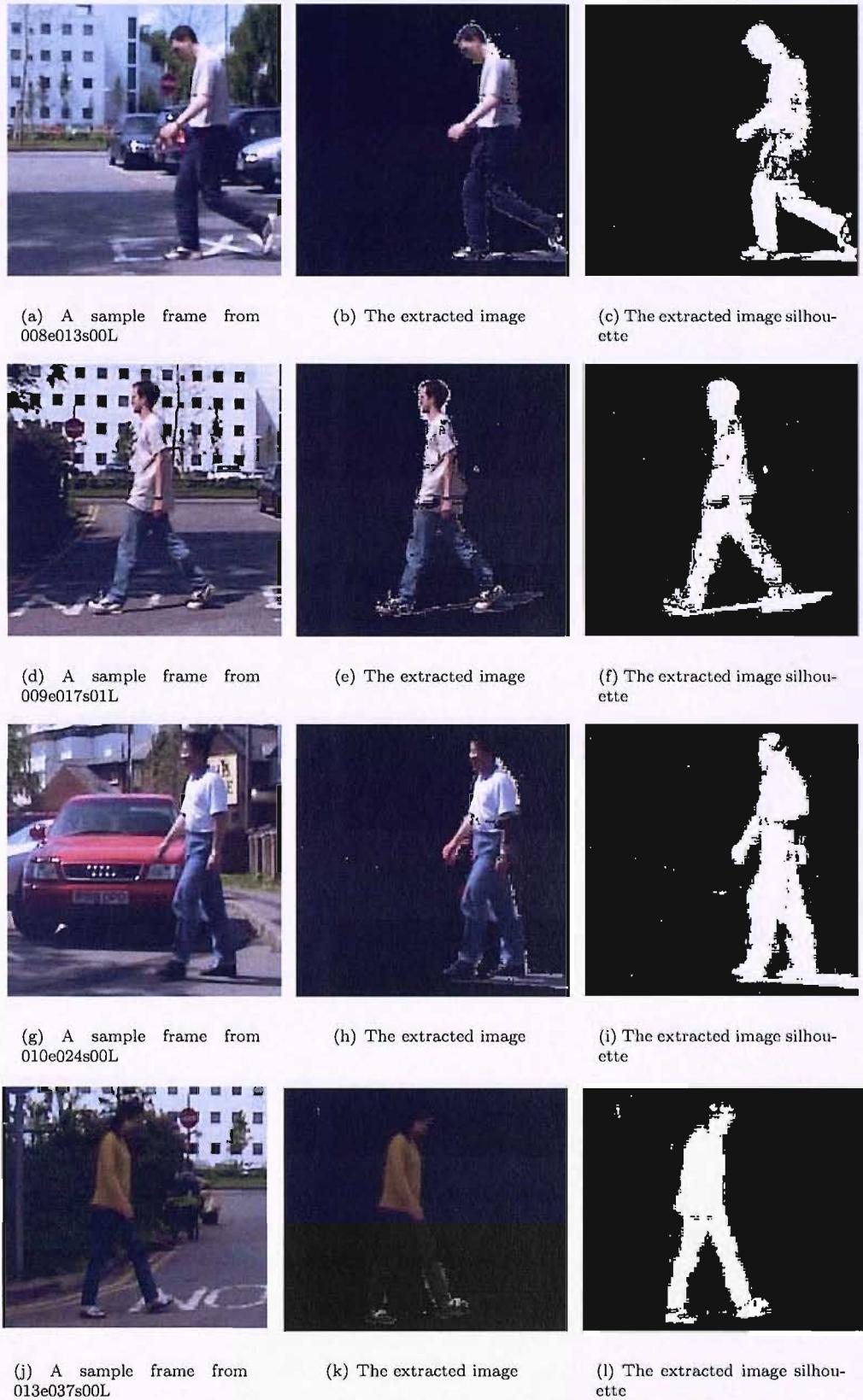


FIGURE 6.4: Outdoor images extracted using the combined UC and SBD classifiers

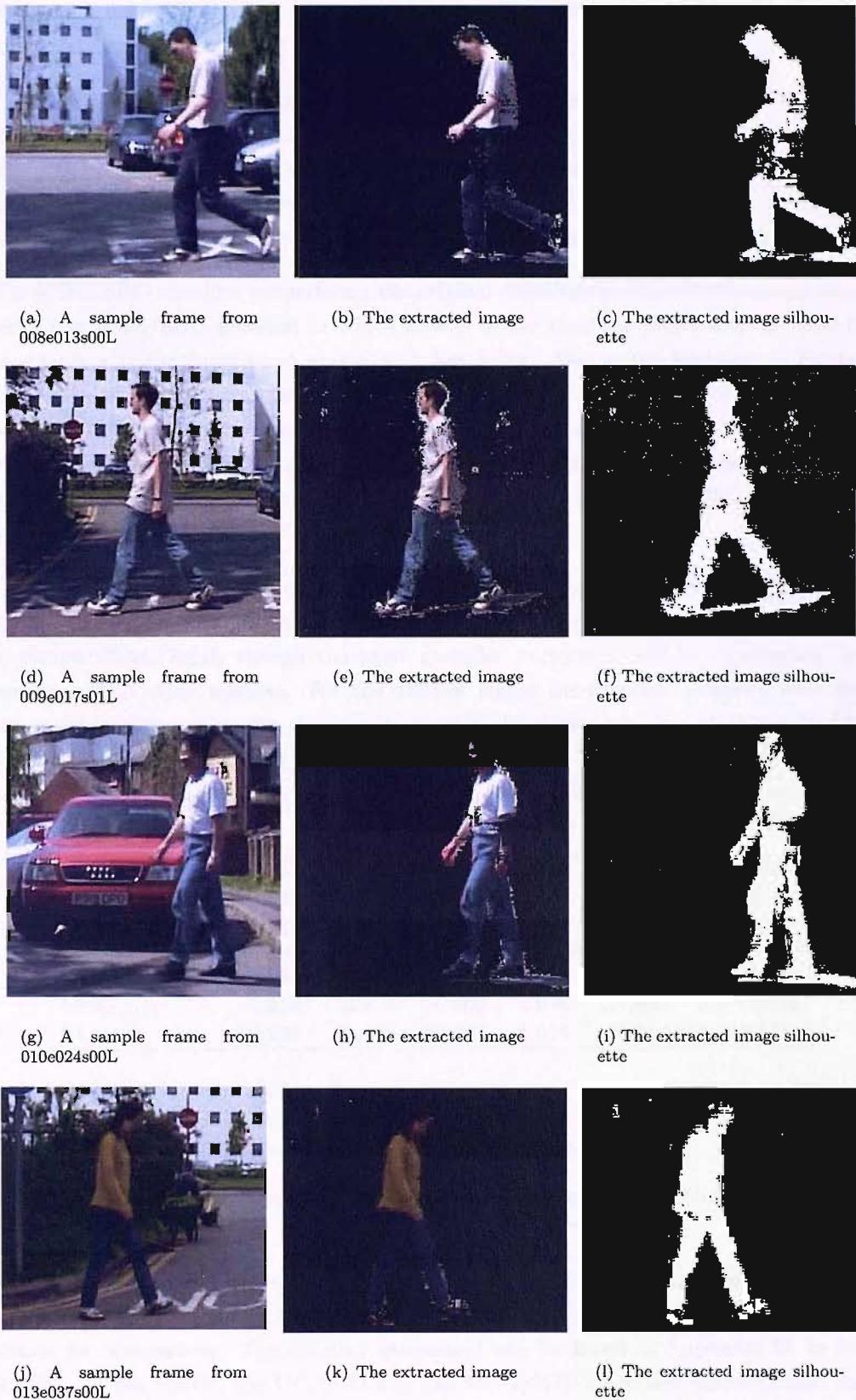


FIGURE 6.5: Outdoor images extracted using the combined UC and MOG classifiers

of the samples, 008e013s00L, but was not in another sample, in 013e037s00L the the combined classifier was better.

Figure 6.6 shows the outdoor motion sequences extraction using the combined MOG/SBD classifier. The classifier performs well in the background region with only few noise pixels. The algorithm succeeds in eliminating the shadow of one sample out of the four presented samples. In the foreground region, the combined classifier manages to classify correctly most of the region but there still exist some holes in the moving subject.

The MOG/SBD classifier outperforms the original classifiers in some of the shown samples. The MOG/SBD classifier performs mostly better than the MOG and the SBD in producing a better foreground region with less holes. Also in the background region, the new classifier performs as well as the SBD classifier in producing a high quality background labelling while the MOG does not produce such quality in the background region (with a noisy background in MOG). For the shadow region, the worst performer is the MOG while the combined classifier's performance is similar to the SBD in this region.

From the shown samples, the UC/SBD gave the best performance in the foreground region. In the background region the UC/MOG was the worst performer especially in sample 009e017s01L though the same classifier performed well in suppressing the background in other samples. For the shadow region the original classifiers were not successful in suppressing the shadows in most of the shown samples which resulted in the combined classifier giving a similar result.

Combined Classifiers	RMSE	PSNR (dB)	FG Error ¹	BG Error ²	Overall Error ³	σ_{RMSE}^2
UC and SBD	0.182	14.961	13.158	2.010	26.801	1.936E-04
UC and MOG	0.193	14.380	15.623	2.104	30.113	1.400E-04
MOG and SBD	0.184	14.862	15.911	1.685	27.337	2.222E-04
SBD	0.189	14.618	14.453	2.153	29.077	2.126E-04
MOG	0.216	13.374	10.009	3.976	37.499	2.370E-04
UC	0.207	13.856	22.027	1.919	35.068	1.297E-04

¹ FG Error Foreground Percentage Error

² BG Error Background Percentage Error

³ The percentage of the overall error compared to the motion pixels only

TABLE 6.2: The overall assessment of combining three sets of two combined classifiers using outdoor motion sequences

Table 6.2 presents the overall assessment of outdoor motion sequences using the new combinations of the UC/SBD, the UC/MOG and the MOG/SBD, with original classifiers results for comparison. The detailed assessment can be found in Appendix C. In the RMSE and the PSNR, the UC/SBD and the MOG/SBD classifiers outperformed the original classifiers. The UC/MOG gave better result than their originators, the UC

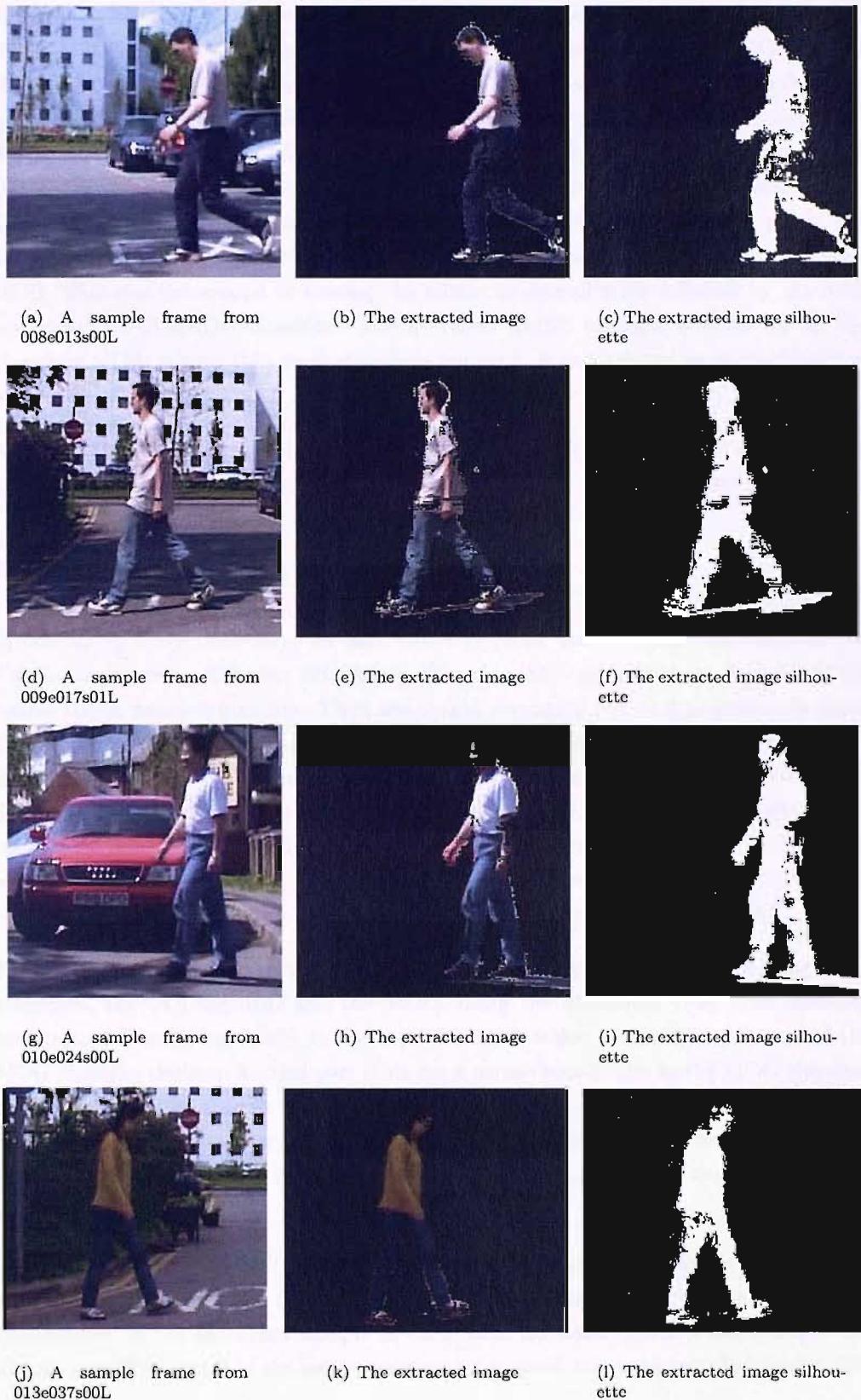


FIGURE 6.6: Outdoor images extracted using the combined MOG and SBD classifiers

and the MOG, but fell below the results of the SBD. For the foreground percentage, the UC/SBD scored the second best result in this region. The UC/MOG and the MOG/SBD took the fourth and fifth position with a relatively small increase in the error value compared to the SBD which was placed in the third position. For the background region, the MOG/SBD classifier scored the best result in this region followed by the UC, the UC/SBD and the UC/MOG classifiers with minor difference between them. Using the RMSE and the overall error compared to the silhouette's motion pixels, the new UC/SBD classifier scored the best result in extraction outdoor motion sequences. MOG/SBD was the second in scoring the minimum overall error followed by the SBD and then the UC/MOG classifier. The averaged RMSE variance is small for all the classifiers. This means that each classifiers variation is constrained in a small limited range of displacement.

It is clear that classifier combination achieves better performance. All the combinations managed to produce an overall error less than their originating classifiers.

6.6.2 Three Classifiers Combination

In combining three classifiers, we used the Maximum, the Median, and the Sum and the Product rules using different principles. The classifiers are combined first using the simple Bayes decision function. Then the weight averaging probability method is used. Each classifier weight is derived from the individual classifier performance. For indoor sequences, the SBD performance was 90%, the MOG was 77%, and the UC was 92%. So after a normalisation process, the SBD weight will be 0.348, the MOG weight will be 0.297, and the UC weight will be 0.355. For outdoor sequences, the SBD performance was 71%, for the MOG 63%, and for the UC 65%. Therefore after normalisation the weights will be: 0.357 for the SBD, 0.317 for the MOG, and 0.326 for the UC.

Figure 6.7 shows the result of extracting indoor motion sequences by combining three classifiers, the UC, the SBD and the MOG, using the Maximum rule. The resulting background is noisy especially in the top left corner which indicates dominance of the MOG classifier decision for this part (this error occurs specifically in the MOG classifier samples only). The shadow has disappeared from two samples but a small part is still resident in the last two samples, 010a024s08L and 017a054s00L. The foreground region is extracted well with few holes (holes are more noticeable in the sample of motion sequence 008a013s00L).

Figure 6.8 shows the result of extracting indoor motion sequences using a new classifier that combines the UC, the SBD, and the MOG classifiers using the Median rule. The background of the extracted sample is clear with no noise pixels. The shadow has disappeared from most of the sequences except for small traces in sample 010a024s08L

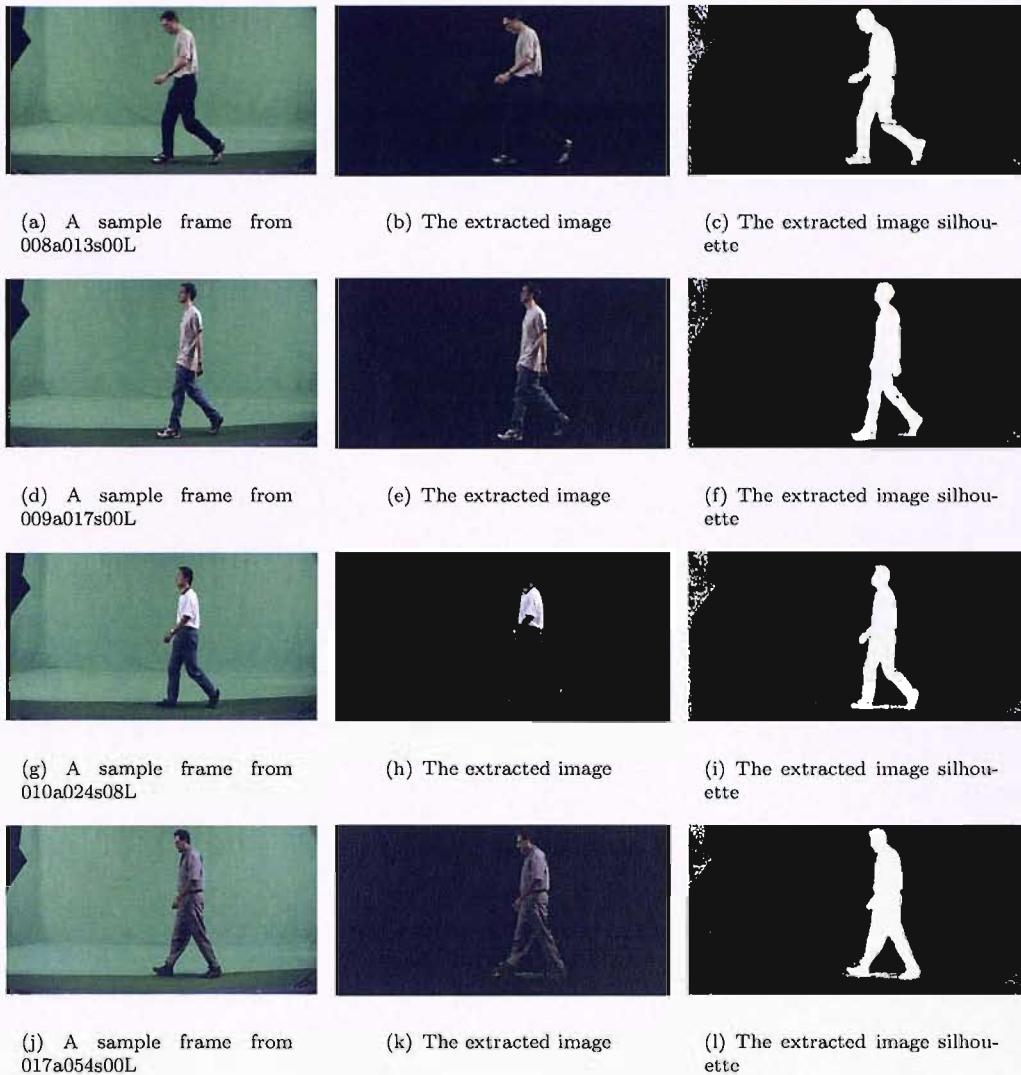


FIGURE 6.7: Indoor images extracted by combining the UC, the SBD, and the MOG classifiers using the Maximum rule

and 017a054s00L. The foreground region is extracted well with some mislabelled pixels, which are more obvious in 010a024s08L.

Figure 6.9 presents the indoor classification using the Product of the UC, the SBD and the MOG classifiers. The background is mostly clean except for the left top corner and another small corner in the bottom right corner of the scene. The shadow had mostly disappeared in most of the samples. The foreground region is extracted well but with more small holes than in the previous combinations of three classifiers.

Figure 6.10 shows an indoor sequences extraction by the Sum of the UC, the SBD and the MOG classifiers. We notice an improvement in extracting the moving subject compared with the Product combiner, though some pixels are misclassified. The left top corner is

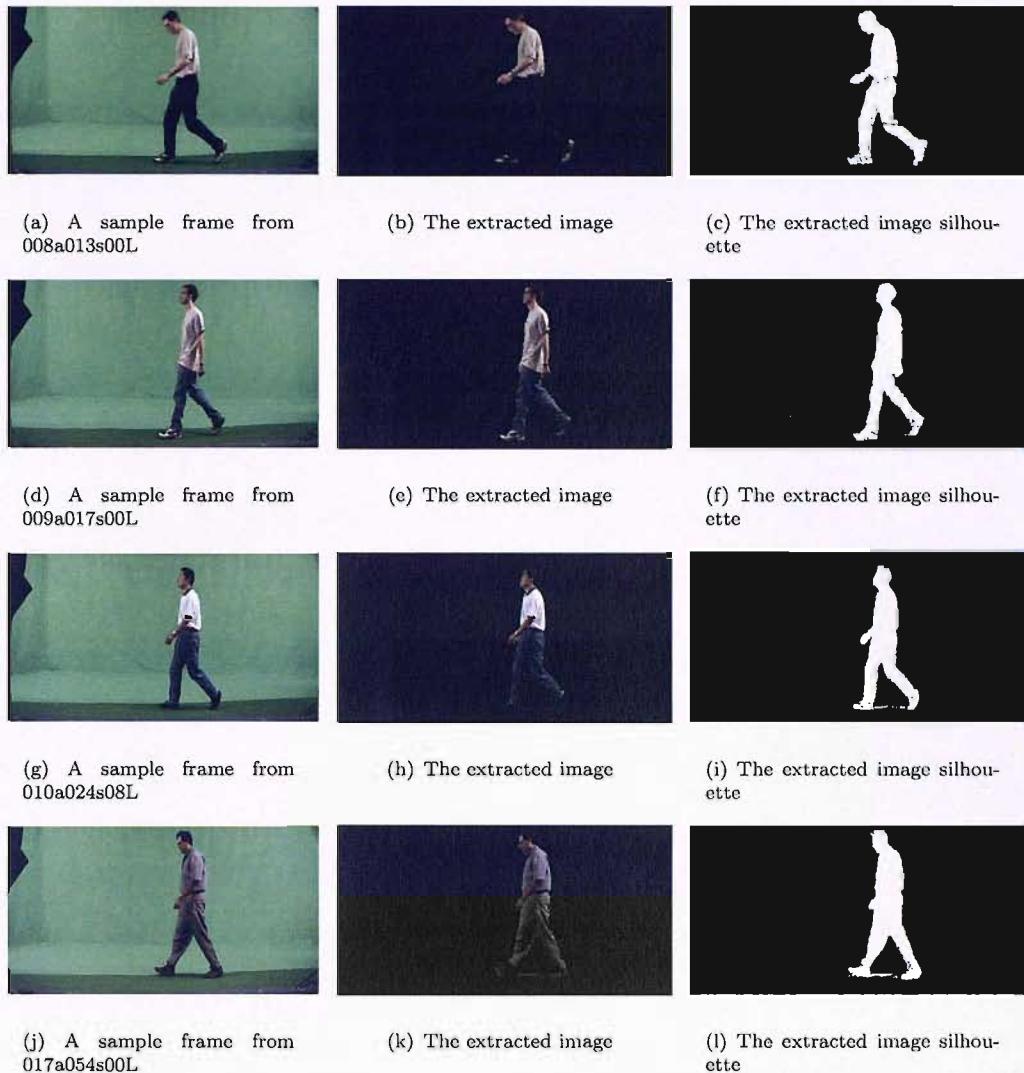


FIGURE 6.8: Indoor images extracted by combining the UC, the SBD, and the MOG classifiers using the Median rule

also noisy in this classifier as well as the bottom right corner. Most of the shadow has disappeared from all the samples except some small parts showing in motion sequences 010a024s08L and 017a054s00L.

Comparing the results of Figure 6.7 - 6.10 with their originating classifiers, Figures 2.10, 3.5, and 4.5, the Median combination behaves as well as the UC and the SBD in classifying the background region. The Maximum, the Sum, and the Product classifiers performed well in most of the scene parts except the top left corner which was noisy, similar to the result of the MOG.

In the foreground region all the new classifiers performed well except for the Product rule combination which suffered from more small holes in the moving subject. The

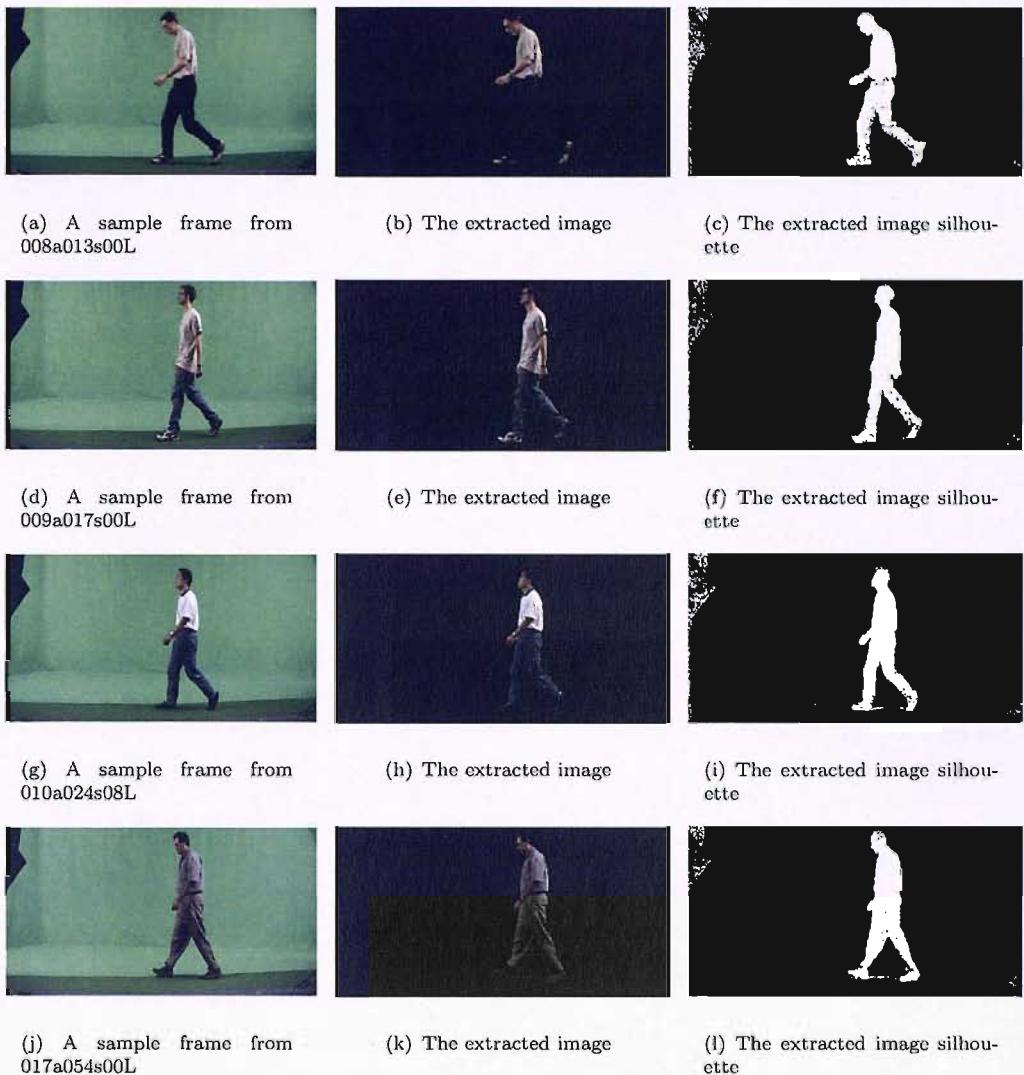


FIGURE 6.9: Indoor images extracted by combining the UC, the SBD, and the MOG classifiers using the Product rule

good performers gave a similar result to the UC while the Product classifier gave results similar to the MOG.

In the shadow region the Median and the Product combinations gave the best result with a minimum traces of shadows showing in two samples only. Also the Max and Sum rule behaved well in this region but with a little more shadow.

Figures 6.11 - 6.14 shows the samples of extracting indoor sequences using a the average weighted combination (see Section 6.5) of the UC, the SBD, and MOG classifiers using the Maximum, the Median, the Product, and the Sum combination rules, respectively. When extracting the background region the median combination rule gave the best result with no noticeable noise in the background except for the last sample, 017a054s00L,

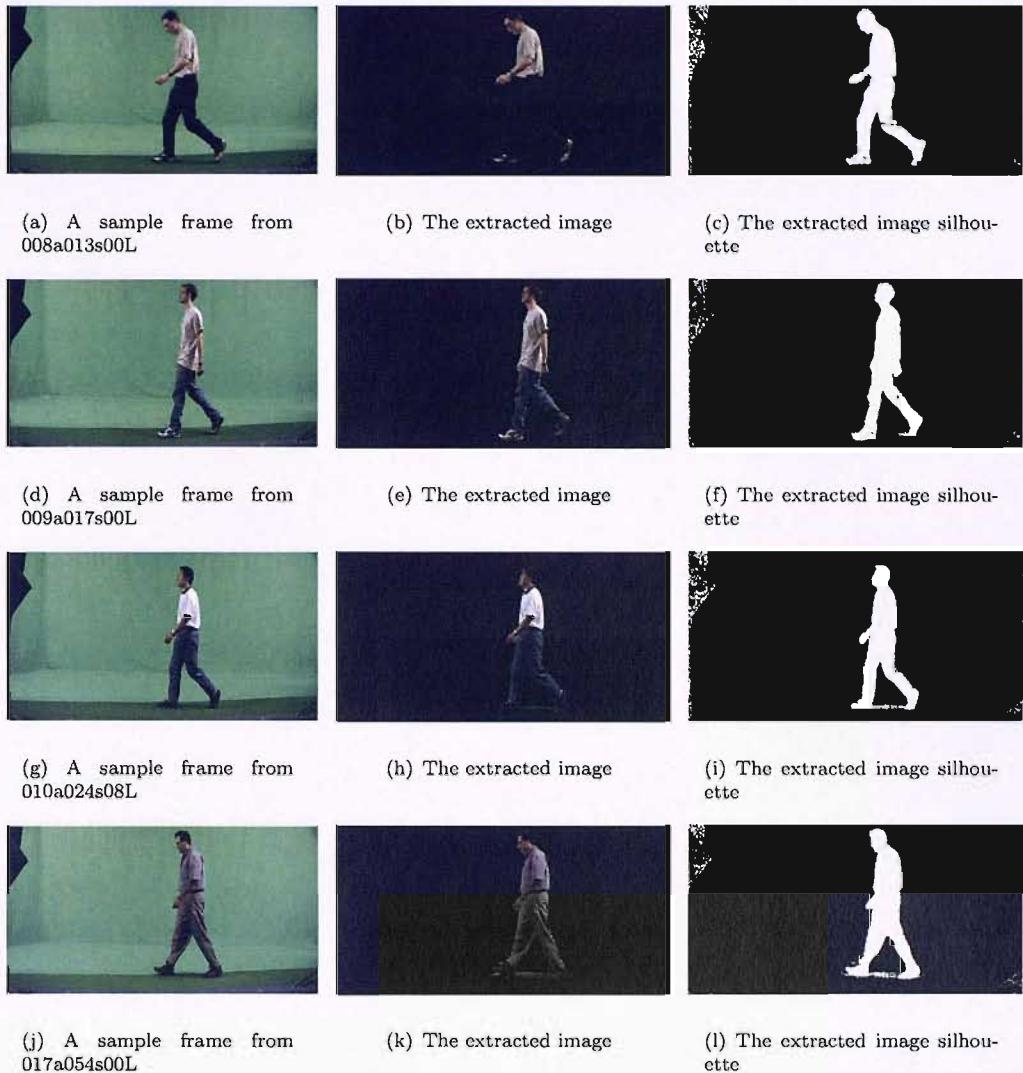


FIGURE 6.10: Indoor images extracted by combining the UC, the SBD, and the MOG classifiers using the Sum rule

where some isolated noise exists. The Maximum rule gives the second best performance in the background region with some noisy pixels showing noticeably in the top left corner. Again, 017a054s00L gave more background noise in this classifiers than the other samples. The Sum rule background noise is more than the previous two classifiers. The top left corner is the noisiest part of the background region. The Product classifier gives the highest background noise compared to the other three combined weighted classifiers.

For the shadow region similar results were apparent in all the four weighted combination classifiers. In all the samples most of the shadow region is removed except some small parts showing in samples of 010a024s08L and 017a054s00L.

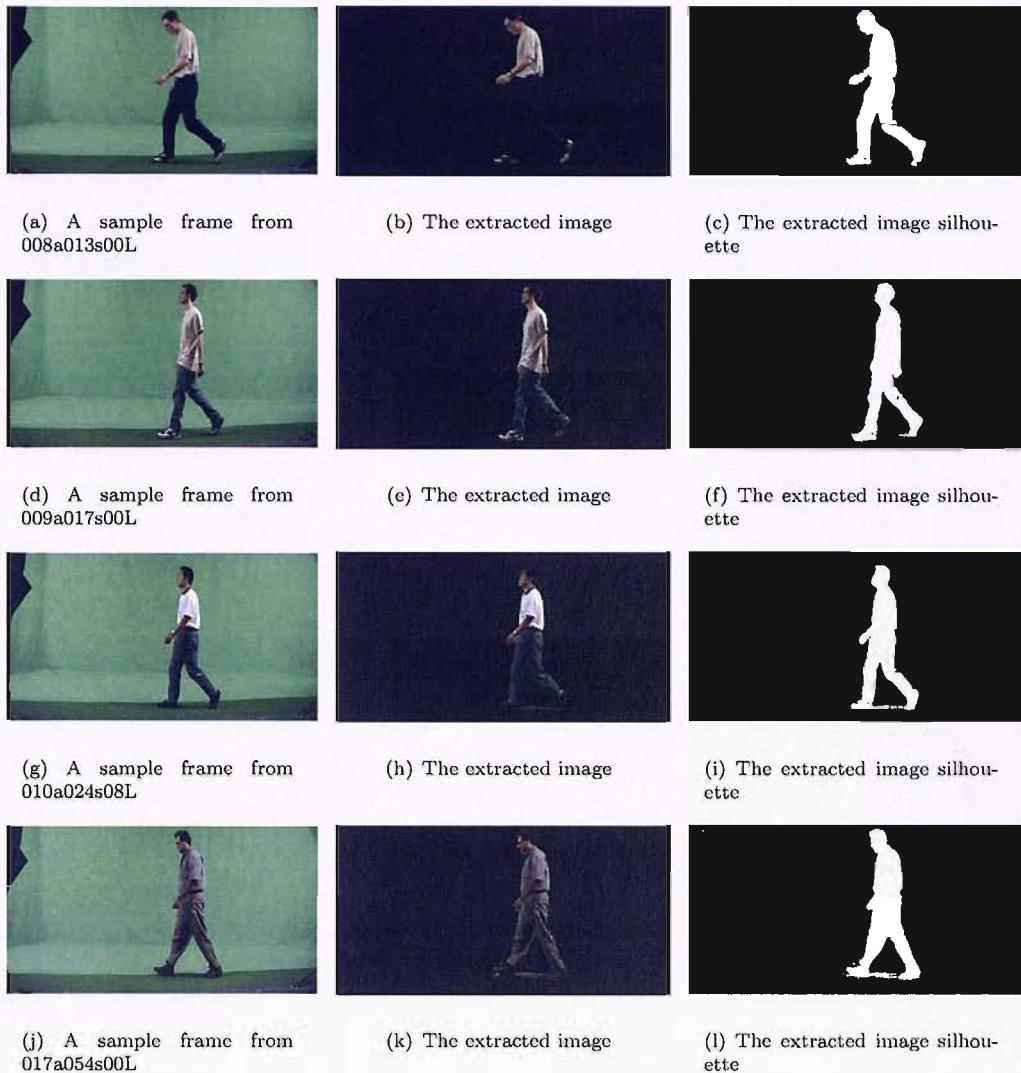


FIGURE 6.11: Indoor images extracted by combining the UC, the SBD, and the MOG classifiers using the weighted Maximum rule

In the foreground region all the four classifiers gave a quality extraction of this region. Small holes are noticed in 008a013s00L and 009a017s00L.

When comparing the results of the four weighted combinations with their originating classifiers (Figures 2.10, 4.5 and 3.5), the new classifiers performed better than the MOG and the SBD in labelling the foreground pixels. The UC gave similar performance to the combined classifiers in this region. For the shadow region, the combined classifiers behaved similarly to the SBD but the MOG and the UC performance seems to be slightly better. In the background region, the new combination using the weighted Median rule gave the best performance in this region. The UC and the SBD gave similar results. The new combinations using the weighted Maximum, Product and Sum gave similar results to the MOG, with a noisy background corner.

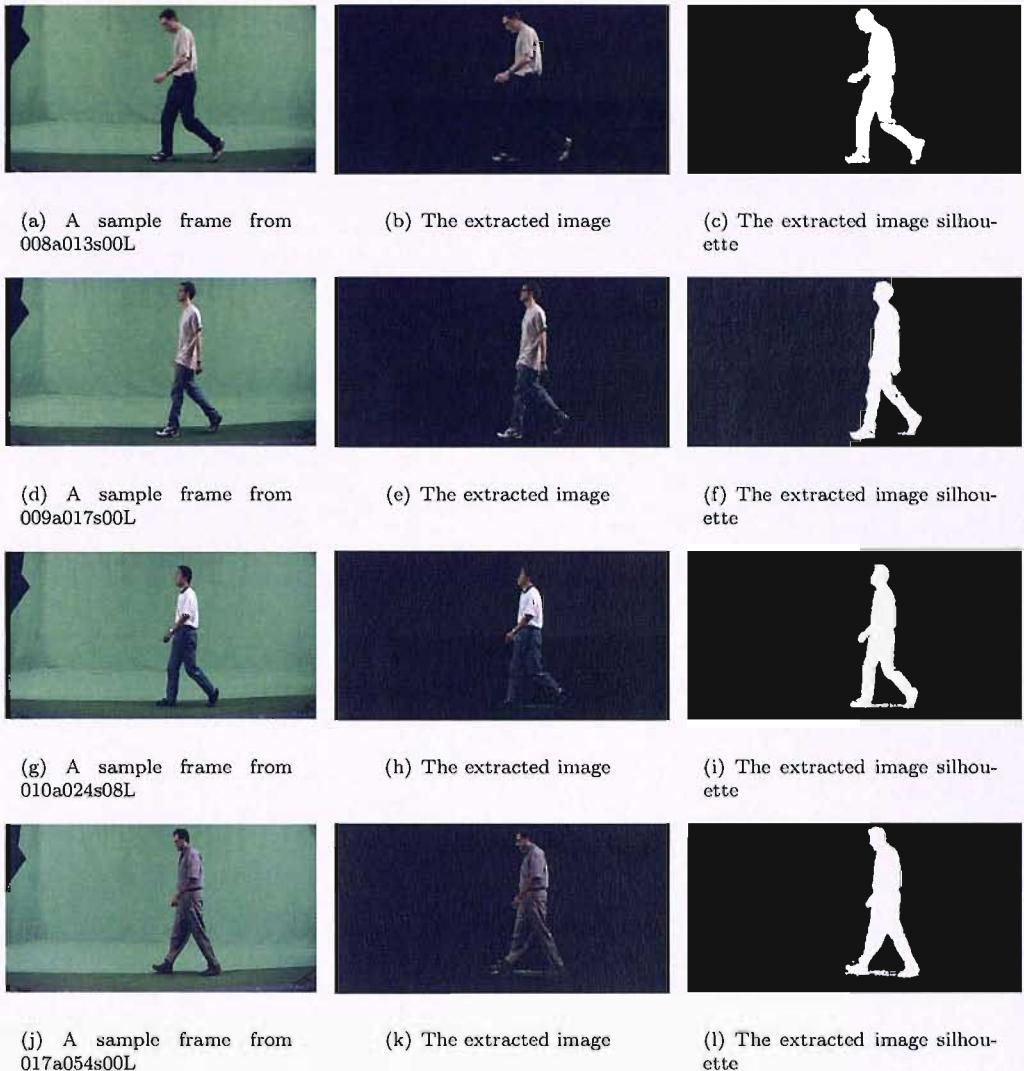


FIGURE 6.12: Indoor images extracted by combining the UC, the SBD, and the MOG classifiers using the weighted Median rule

Table 6.3 presents the overall results for indoor motion sequences. For the detailed assessment for each combination please refer to Appendix C. The table is divided to three parts. The first part, Table 6.3(a), presents the Max, the Median, the Product, and the Sum combination rules. The second part, Table 6.3(b), present the weighted combination results using the same combination rules in part 6.3(a). The last part shows the results of the original classifiers for comparison.

For the RMSE and the PSNR in the unweighted combinations, the best value is scored by the Median with the least RMSE, 0.07, and the highest PSNR, 23.55dB. When using the weights, we notice an improvement in the RMSE and the PSNR of the Max and the Product combinations. The RMSE and the PSNR values of the Median and the

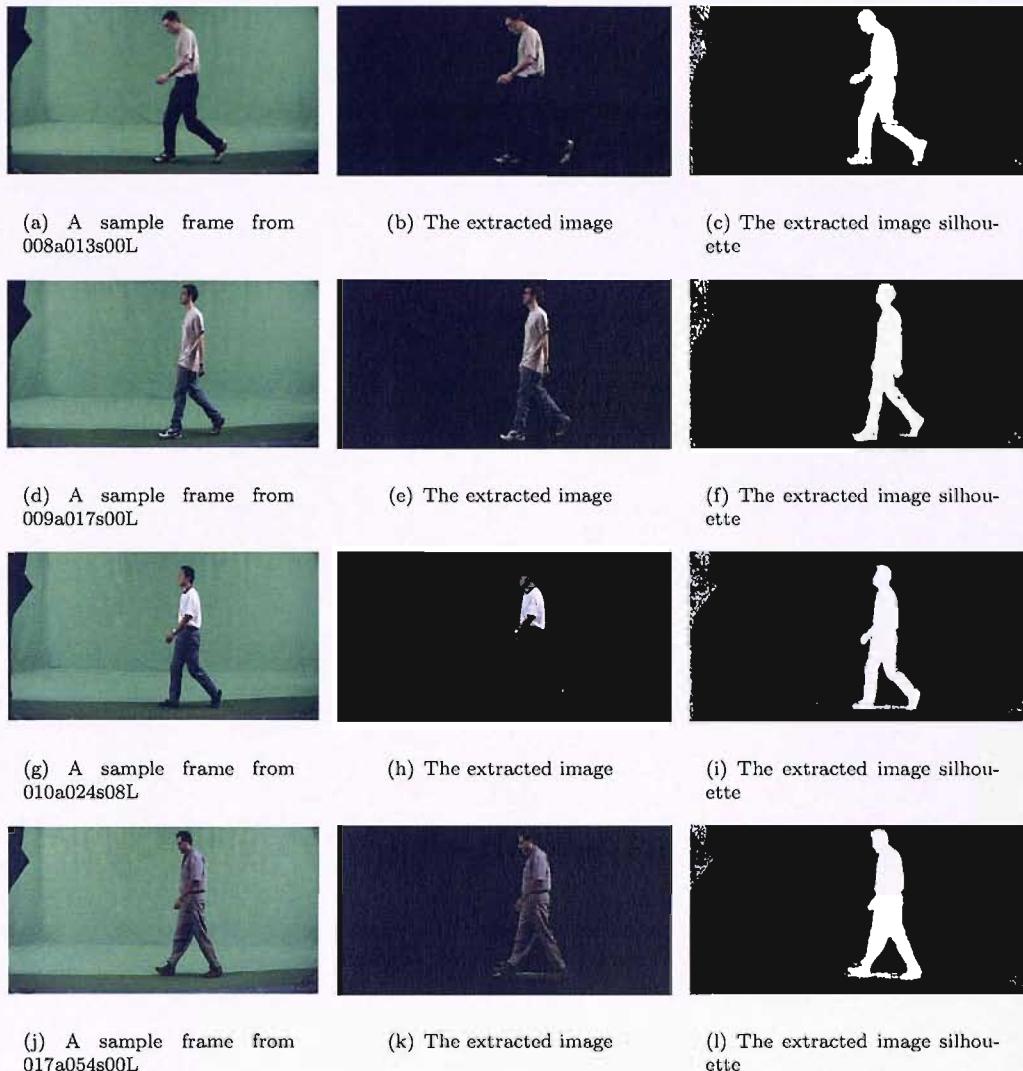


FIGURE 6.13: Indoor images extracted by combining the UC, the SBD, and the MOG classifiers using the weighted Product rule

Product either remained unchanged, for RMSE in the product, or deteriorated slightly. The Median in this part also scored the best result in the RMSE and the PSNR.

For the foreground percentage error, the Max and the Sum scored the best results in Table 6.3(a) followed by the Median and then the Product rule which scored the worst result in this region. In the background percentage error the result is inverted where the best result is scored by the Product and the worst by the Max and the Median. For the results of the foreground region in the weighted combination, Table 6.3(b), the foreground error is similar with a small difference in all the different combination rules. The Max rule scored the best result in this region followed by the Median and then the Sum and the Product. The relation seen in the unweighted combination does not hold here, where the Product rule scored the worst foreground and background error as

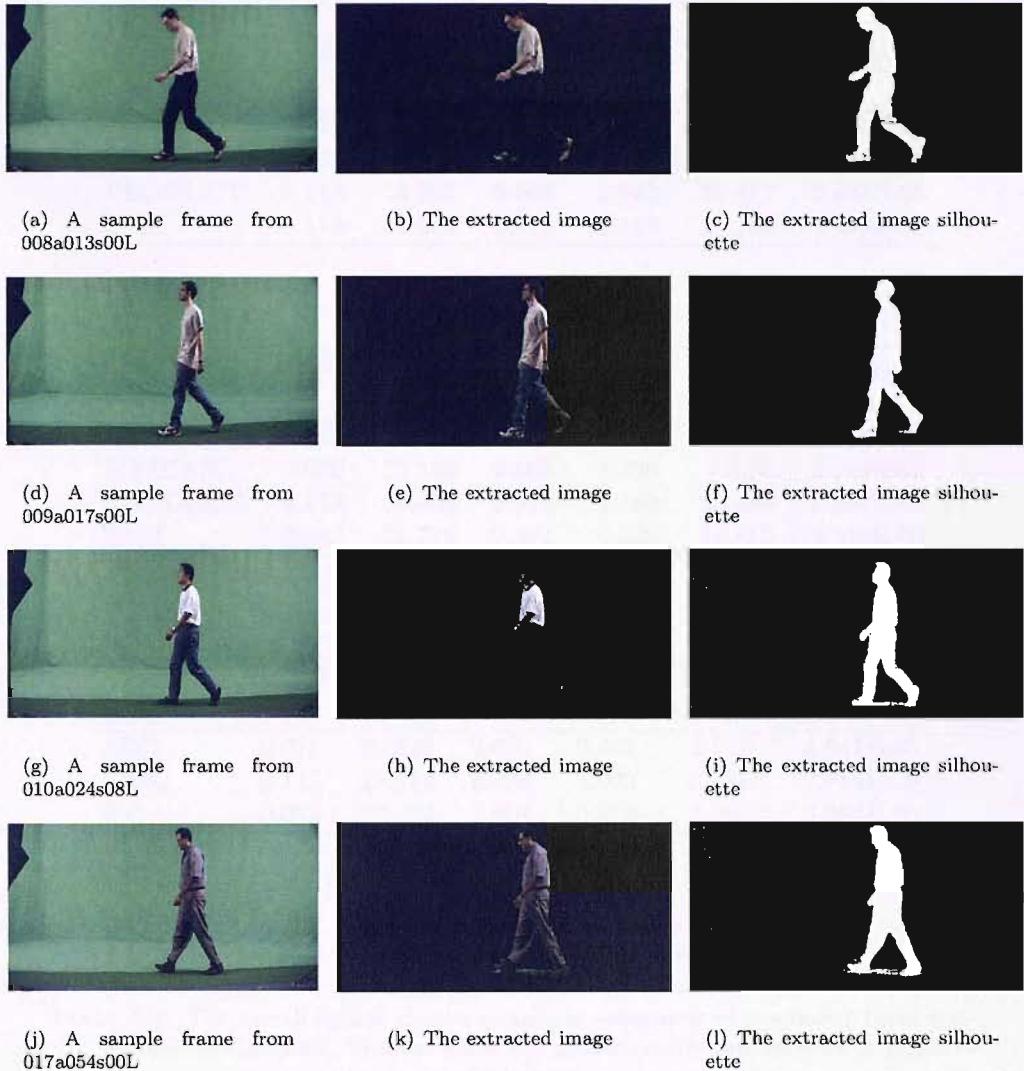


FIGURE 6.14: Indoor images extracted by combining the UC, the SBD, and the MOG classifiers using the weighted Sum rule

well. Also the Sum rule scored the second worst foreground and background error. The minimum background error is scored by the Median with 0.34% followed by the Max rule with a 0.39%.

The overall error compared to the silhouette's motion pixels gave the first and the second position to the new Median classifiers of parts 6.3(a) and 6.3(b) with an error of 7.07% and 7.85% respectively. The weighted Maximum and Sum scored reasonably competitive overall results but the result of the remaining four combined classifiers (the unweighted Maximum, Product, and Sum and the weighted Product) were not satisfactory. We notice that the Product combiner does not perform well whether a weight is used or not. This is because the Product is more vulnerable of making one classifier's erroneous result more dominant (when compared to the other combination rules) especially if the

(a) Combining by using only the classifiers probabilities						
Combined Classifiers	RMSE	PSNR (dB)	FG Error ¹	BG Error ²	Overall Error ³	σ_{RMSE}^2
MAX	0.113	19.004	2.972	1.169	20.169	3.351E-05
MEDIAN	0.067	23.546	3.922	0.216	7.073	3.472E-05
PRODUCT	0.113	18.952	6.543	0.940	20.417	3.104E-05
SUM	0.113	19.004	2.972	1.169	20.168	3.309E-05

(b) Combining by using weighted probabilities						
Combined Classifiers	RMSE	PSNR (dB)	FG Error ¹	BG Error ²	Overall Error ³	σ_{RMSE}^2
MAX	0.074	22.737	2.954	0.387	8.598	4.084E-05
MEDIAN	0.070	23.149	2.963	0.336	7.849	3.464E-05
PRODUCT	0.113	19.004	2.972	1.169	20.168	3.351E-05
SUM	0.082	21.779	2.971	0.525	10.655	3.448E-05

(c) Original classifiers result						
Original Classifiers	RMSE	PSNR (dB)	FG Error ¹	BG Error ²	Overall Error ³	σ_{RMSE}^2
SBD	0.078	22.342	2.659	0.491	9.775	4.047E-05
MOG	0.119	18.513	6.488	1.093	22.617	2.814E-05
UC	0.072	22.872	2.801	0.373	8.265	3.394E-05

1 FG Error Foreground Percentage Error

2 BG Error Background Percentage Error

3 The percentage of the overall error compared to the motion pixels only

TABLE 6.3: The overall indoor motion sequences assessment of combining three classifiers using the Maximum, Median, Sum, and product using two different principle of combination

classifier's probability is too small or equal to zero. For the unweighted Maximum and Sum, when compared to the averaged weighted result of the same combination rule, it is obvious that using weights for each classifier can rescale each classifier probability in comparison with the other classifiers in order to get a better performance.

Comparing the results of the new combined classifiers with the their original classifiers in the RMSE, the PSNR and the overall error, the Median rule in parts 6.3(a) and 6.3(b) scored better results than all the original classifiers. The Maximum weighted combination rule, Table 6.3(b), scored competitive results but fell behind the best originating classifier, the UC. Similar to the UC and the SBD, the weighted Maximum rule managed to score better results than the modified SBD and the MOG. For the foreground error, the weighted combination managed to maintain an error less than 3.0% in all the combinations used. The unweighted combination performed similarly in the Maximum and the Sum rules but the error exceeded this percentage in the Median and the Product

rules. Finally in the background error, the SBD and the UC both gave results less than 0.5%. Most of the weighted combination scored similar or at least close results except for the Product rule which scored the worst results in this region. In the unweighted combinations, only one classifier managed to keep up to this level, the Median. The best performers in the background are the unweighted Median followed by the weighted Median classifier.

The averaged RMSE variance gives small results for all the shown classifiers with a maximum RMSE variance in the weighted Maximum with a variance of 4.084E-05. These small value means that each classifier performance result is highly consistent.

As an overall result, combining classifiers for indoor motion sequences can lead to better performance results. The Median combiner is a powerful tool with a potential to give competitive results. The combination without weights can give the best results in one (or more) of the combination rules but using weights can ensure better results in many combination rules by optimising the contribution of the inputs.

Figures 6.15 - 6.18 presents samples of the extraction of outdoor motion sequences using the combination rules, the Maximum, the Median, the Product and the Sum. For the background labelling the Median provides the best performance in this region with only few scattered noisy pixels. All the other three classifiers suffers from noise in the background area. For the shadow region in most of the provided samples, shadows are still unsuppressed except for the fourth sample, 013e037s00L, where the majority of the shadow disappeared from this sample. For the foreground region, though the overall labelling is acceptable, the moving subjects suffered from holes of different sizes. In comparing the results of the different combinations no clear difference is noticed in samples of 008e013s00L and 009e017s01L. For the other two samples, 010e024s00L and 013e037s00L, though the Median region gave the best results in sample 013e037s00L but in sample 010e024s00L gave the highest foreground error. The Product combination gave the worst foreground result for sample 013e037s00L.

When comparing samples result of the outdoor extraction of the combined classifiers with their original classifiers, the UC in Figure 4.6, the MOG in Figure 3.6, and the SBD in Figure 2.11, the Median rule again managed to outperform all the classifiers. The best performers for the background region are the Median, the SBD, and the UC (with only very few noise pixels). For the foreground region, though the combined results suffered from some holes in the foreground region but the holes are smaller in size than some of the holes in the originating classifiers. For the shadow region, the UC seems to be the best performer in this region while the MOG is the worst. The new combined classifiers gave better shadow suppression than the MOG with a similar result to the SBD but not as good as the UC in this region.

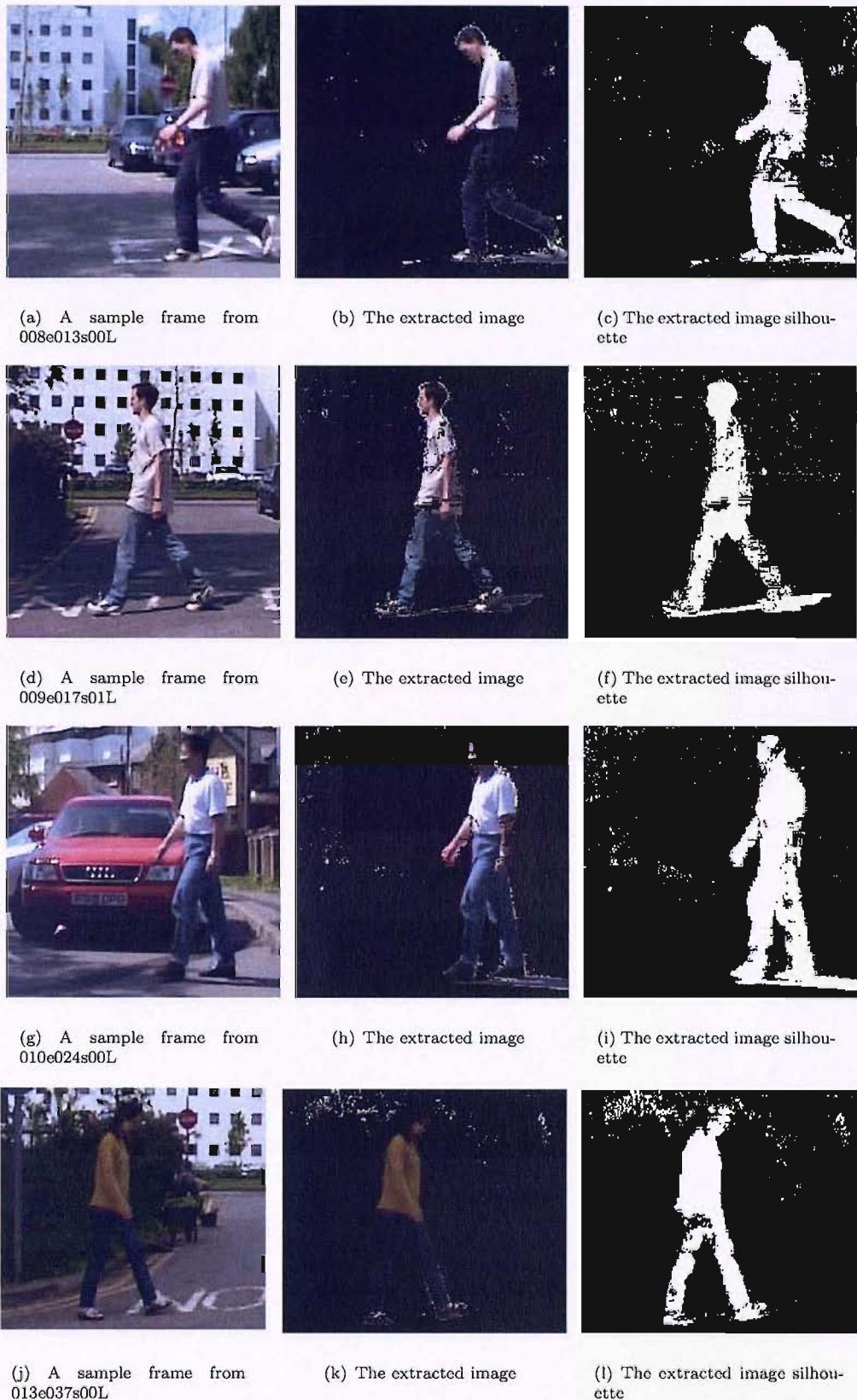


FIGURE 6.15: Outdoor images extracted by combining the UC, the SBD, and the MOG classifiers using the Maximum rule

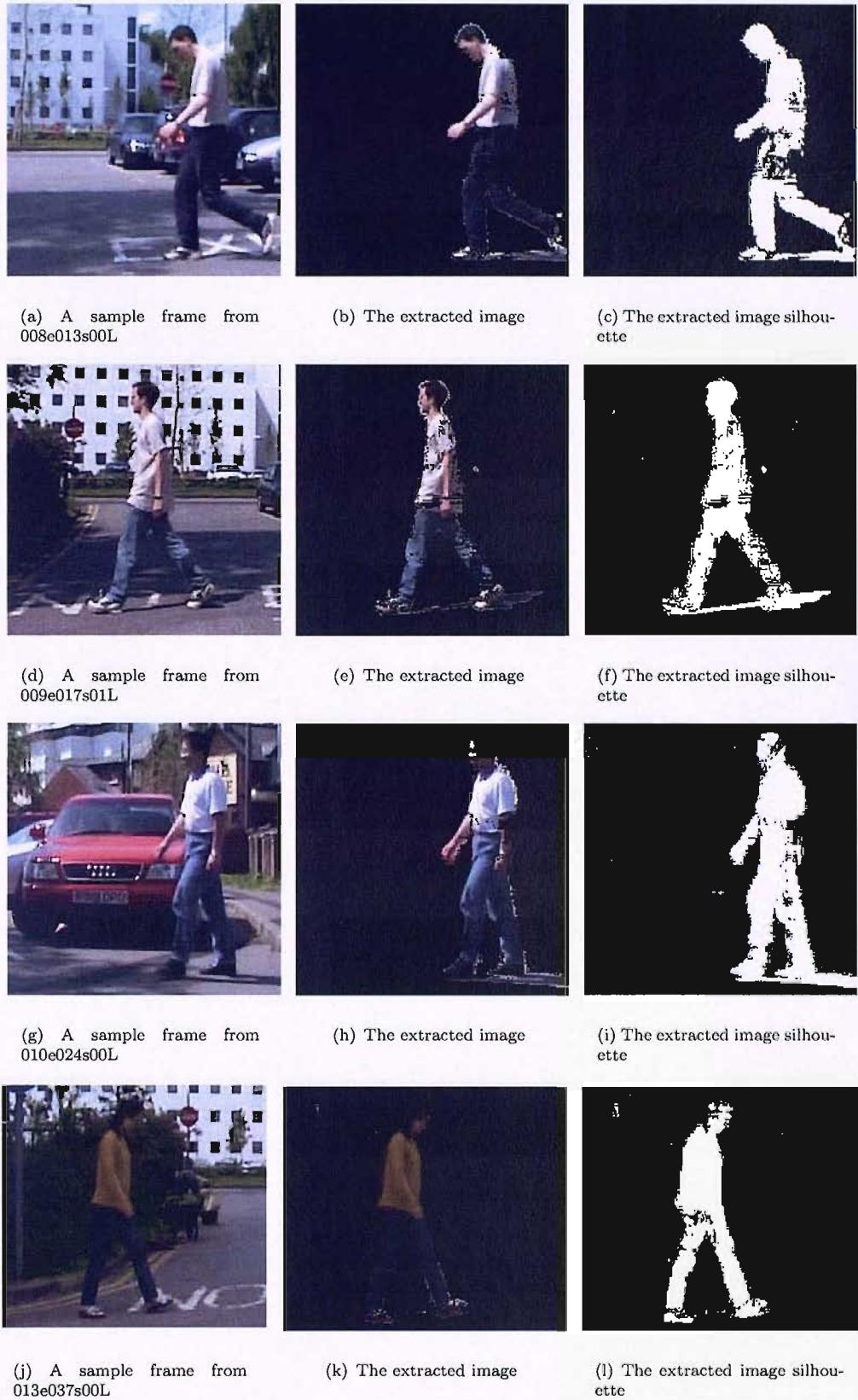


FIGURE 6.16: Outdoor images extracted by combining the UC, the SBD, and the MOG classifiers using the Median rule

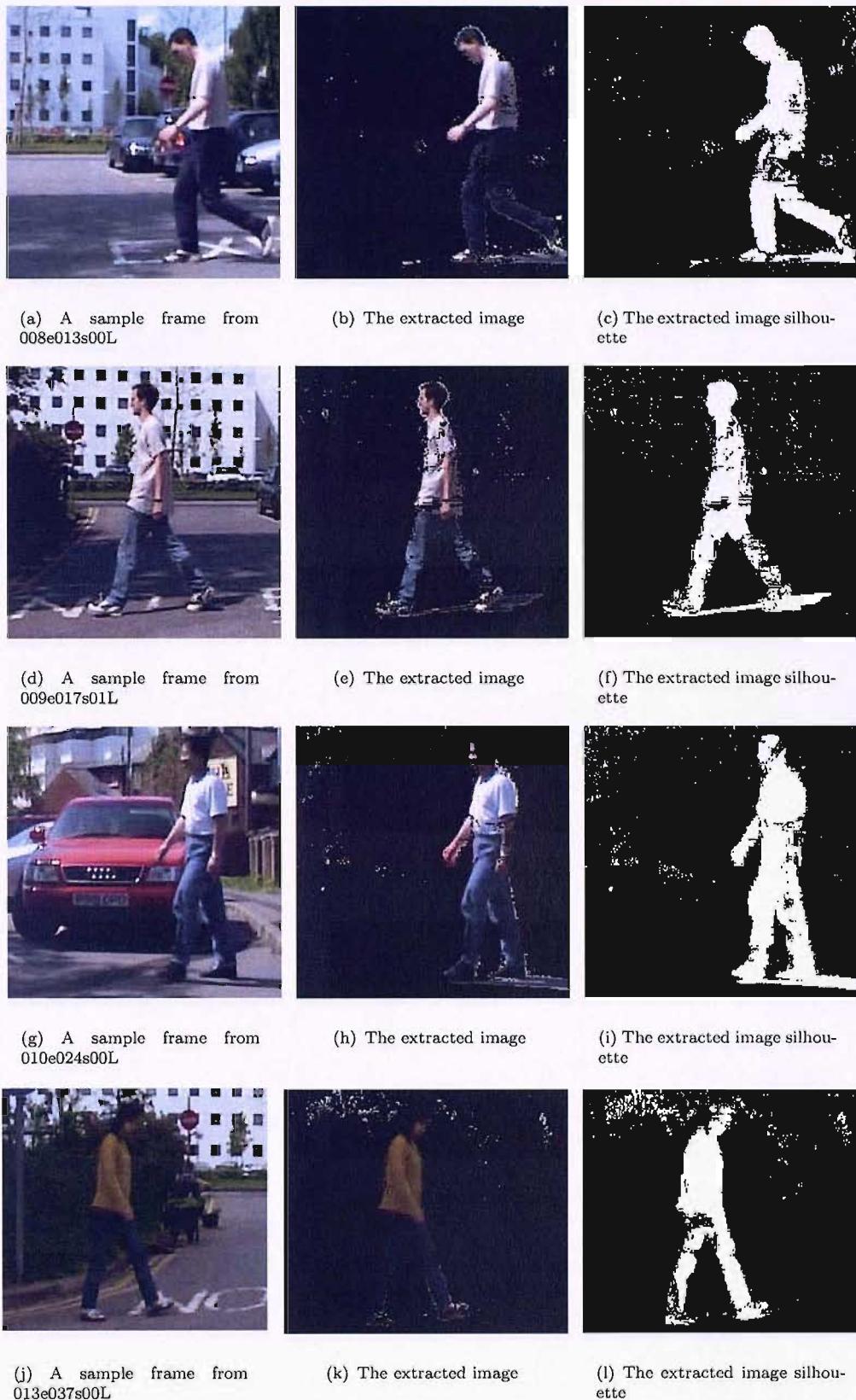


FIGURE 6.17: Outdoor images extracted by combining the UC, the SBD, and the MOG classifiers using the Product rule

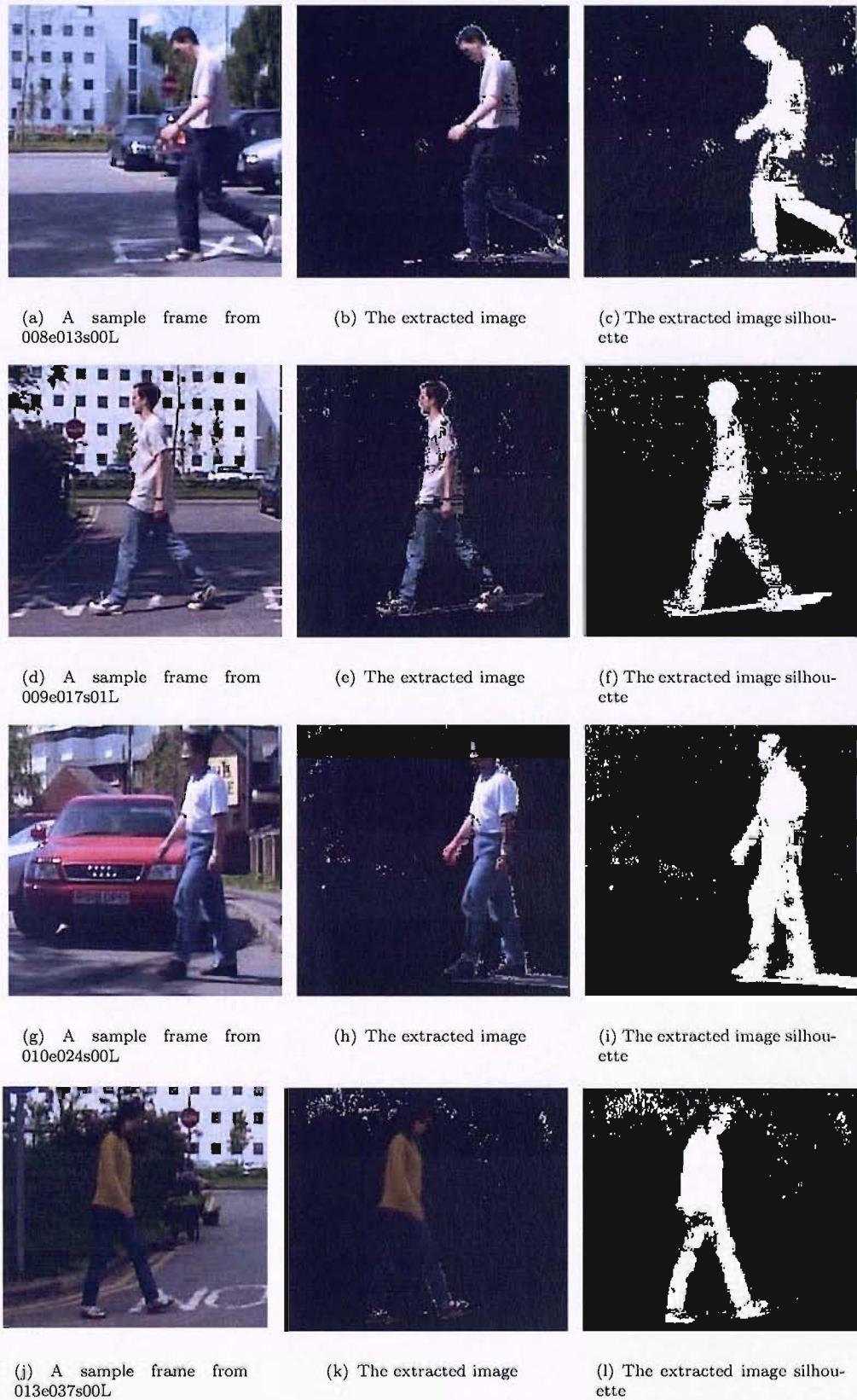


FIGURE 6.18: Outdoor images extracted by combining the UC, the SBD, and the MOG classifiers using the Sum rule

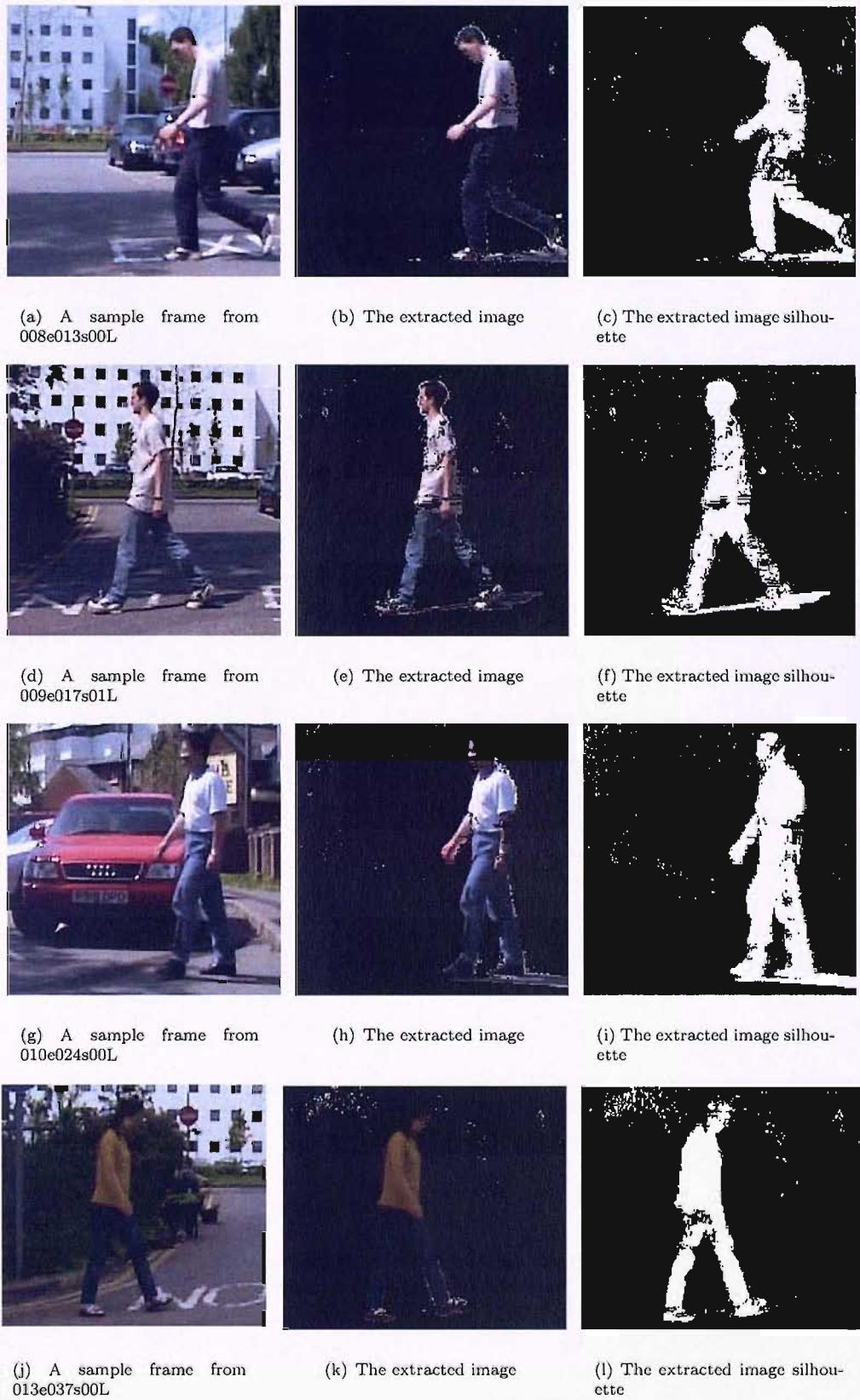


FIGURE 6.19: Outdoor images extracted by combining the UC, the SBD, and the MOG classifiers using the weighted Maximum rule

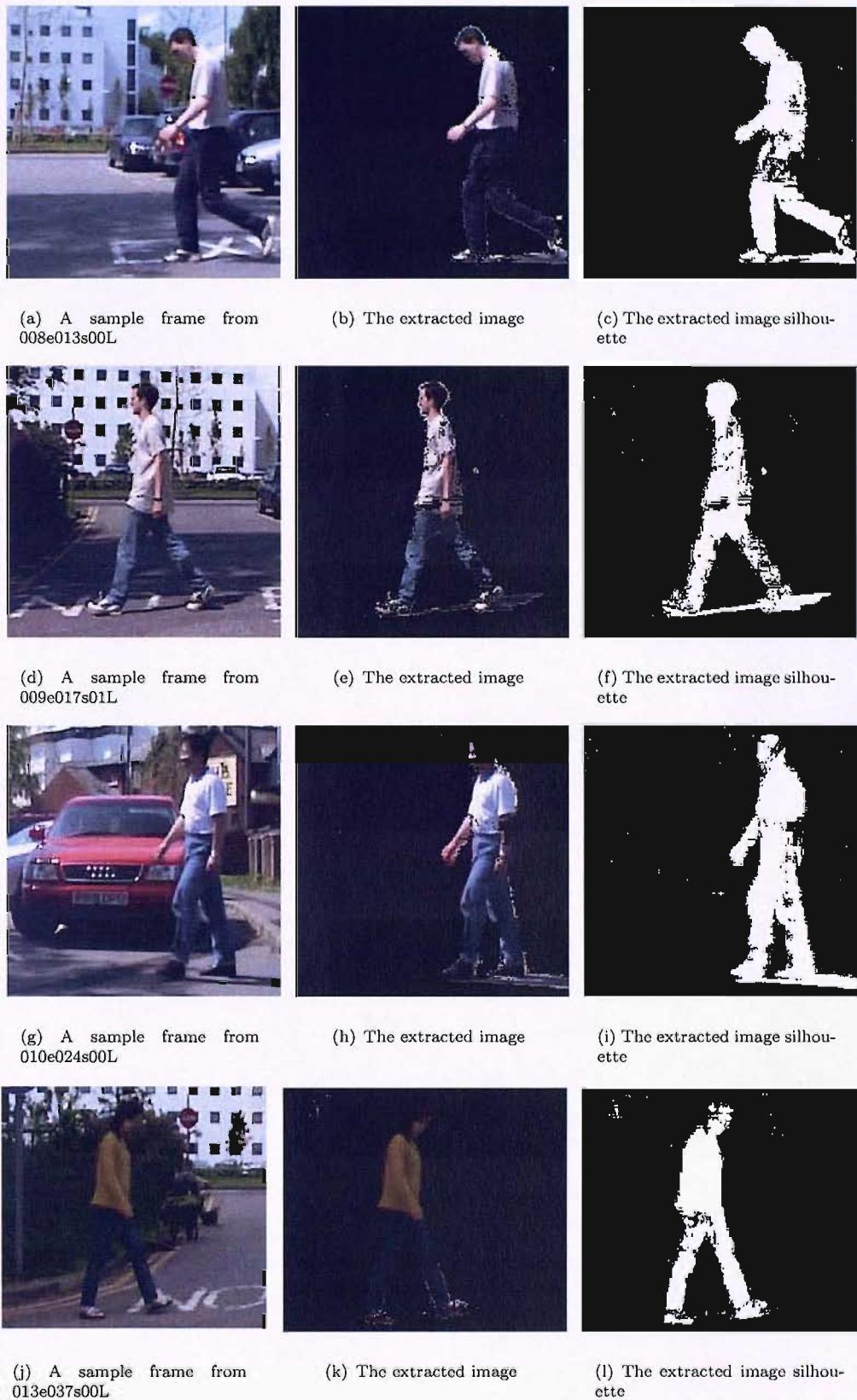


FIGURE 6.20: Outdoor images extracted by combining the UC, the SBD, and the MOG classifiers using the weighted Median rule

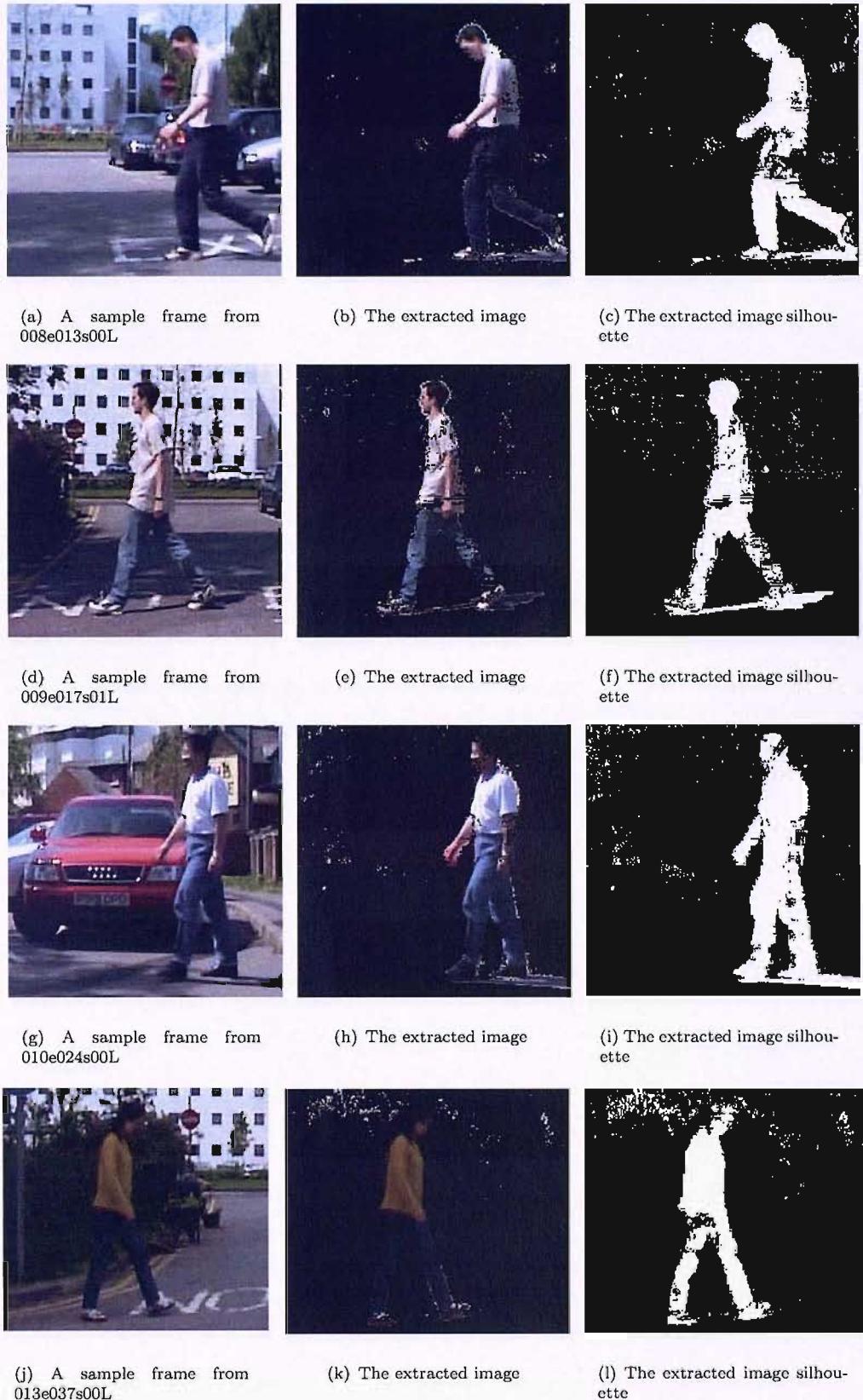


FIGURE 6.21: Outdoor images extracted by combining the UC, the SBD, and the MOG classifiers using the weighted Product rule

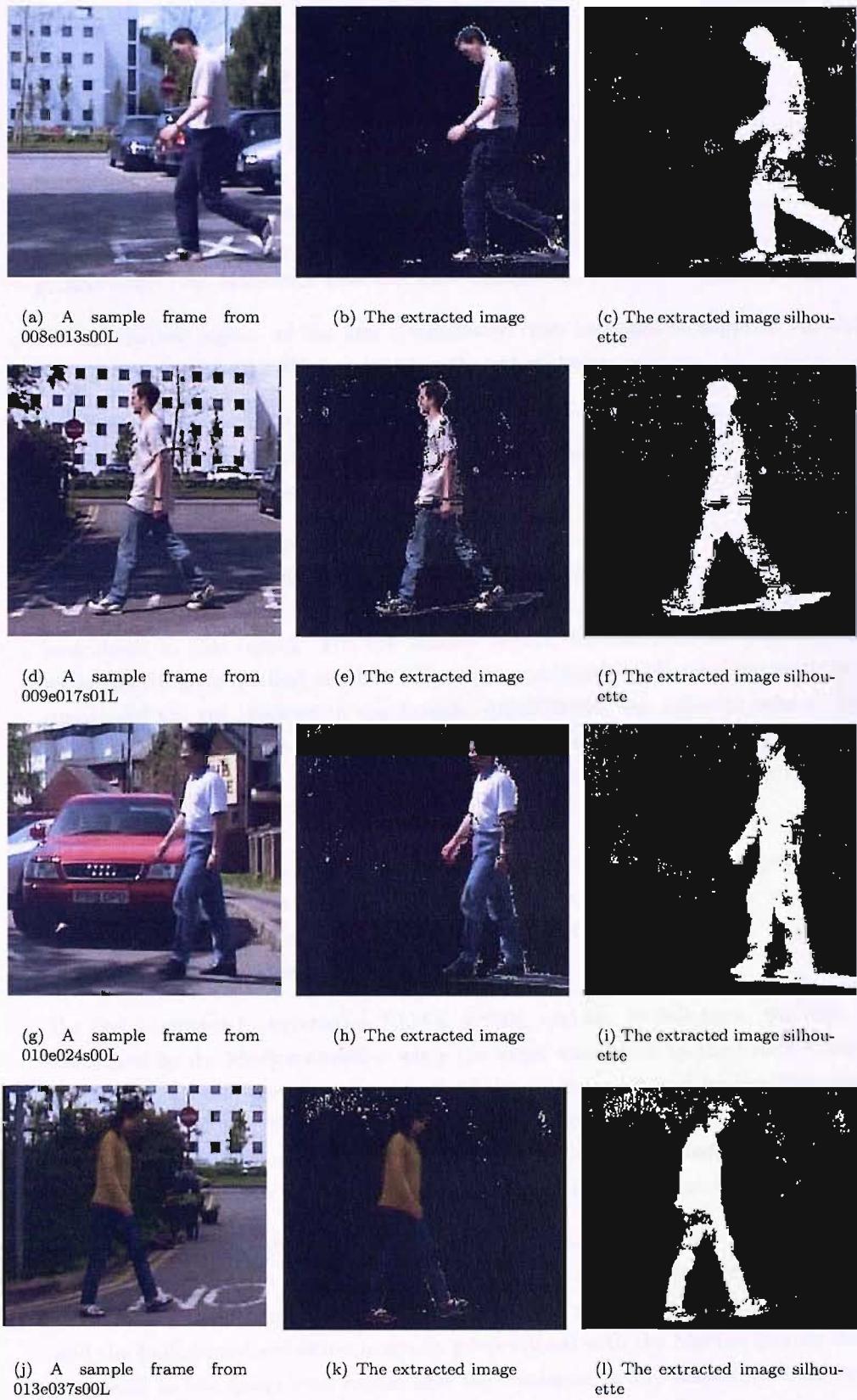


FIGURE 6.22: Outdoor images extracted by combining the UC, the SBD, and the MOG classifiers using the weighted Sum rule

Figures 6.19 - 6.22 shows outdoor motion sequences extracted using the weighted combination of the Maximum, the Median, the Product, and the Sum rules. For the background region the Median combination rule, Figures 6.20, managed to correctly label most of the background pixels. The other three combination rules suffered more background noise (the Maximum gave the least background error among the those three).

For the shadow region, all the four combination rules managed to suppress the shadow only in sample 013e037s00L but failed in the other three.

In the foreground region, the Product rule gave the best performance in sample 010e024s00L while the Median rule gave the best performance in sample 013e037s00L. The other samples did not give any clear preference for a combination rule.

The results of the combined classifiers gave in some cases better results than the original classifiers (Figure 2.11 for the SBD, Figure 3.6 for the MOG, and Figure 4.6 for the UC). In the background region, the median combiner, the SBD and the UC gave the best result in this region. For the shadow region, the MOG is the worst performer while the UC gave the best results. The weight combined classifiers along with the SBD suppressed the the shadows in one sample, 013e037s00L, but failed in others. For the foreground region, the weight combined classifiers suffered from the mislabelling of some of the foreground pixels but the size of the mislabelled groups is usually smaller than the original classifiers.

Table 6.4 shows the assessment result for the the combination classifiers using the Maximum, the Median, the Product, and the Sum rule using a principle of unweighted combination, Table 6.4(a), and a weighted combination, Table 6.4(b). Also the table provides the original classifiers results, Table 6.4(c).

For the unweighted combination RMSE, PSNR, and the overall error, the best values are scored by the Median classifier while the worst was scored by the Product classifier. For the foreground percentage error, the best result was scored by the Sum classifier with an error of 15.17%. The worst error was scored by the Product classifier. In the background error column, the minimum error is scored by the Median classifier, 1.58%, while the highest error was scored by the Sum and the Maximum classifiers.

For the weighted combination rules, the Median classifier gave the best result for the RMSE, the PSNR and the overall error compared to the silhouette's motion pixels. The worst result in these measurement is scored by the product classifier. The foreground and the background errors are inversely proportional with the Median scoring the highest result in the foreground region and the minimum in the background region. The Product scored the minimum error in the foreground region and the highest error in the background region.

(a) Combining by using only the classifiers probabilities						
Combined Classifiers	RMSE	PSNR (dB)	FG Error ¹	BG Error ²	Overall Error ³	σ_{RMSE}^2
MAX	0.203	13.982	15.197	2.668	33.569	2.088E-04
MEDIAN	0.181	14.966	15.921	1.580	26.628	2.135E-04
PRODUCT	0.206	13.868	16.395	2.601	34.317	2.305E-04
SUM	0.203	13.985	15.170	2.668	33.545	2.074E-04

(b) Combining by using weighted probabilities						
Combined Classifiers	RMSE	PSNR (dB)	FG Error ¹	BG Error ²	Overall Error ³	σ_{RMSE}^2
MAX	0.192	14.468	15.628	2.074	29.835	1.896E-04
MEDIAN	0.182	14.928	15.905	1.613	26.848	2.117E-04
PRODUCT	0.203	13.982	15.197	2.668	33.569	2.088E-04
SUM	0.198	14.196	15.241	2.424	31.896	1.988E-04

(c) Original classifiers result						
Original Classifiers	RMSE	PSNR (dB)	FG Error ¹	BG Error ²	Overall Error ³	σ_{RMSE}^2
SBD	0.189	14.618	14.453	2.153	29.077	2.126E-04
MOG	0.216	13.374	10.009	3.976	37.499	2.370E-04
UC	0.207	13.856	22.027	1.919	35.068	1.297E-04

1 FG Error Foreground Percentage Error

2 BG Error Background Percentage Error

3 The percentage of the overall error compared to the motion pixels only

TABLE 6.4: The overall outdoor motion sequences assessment of combining three classifiers using the Maximum, Median, Sum, and product using two different principle of combination

The averaged RMSE variance ranged from 1.988E-04 to 2.305E-04 in the combined classifiers. These small values illustrates that the combined classifiers are highly consistent in their performance.

Now we will rank the classifiers according to their performance in each region in Table 6.4. For the background region, the best performer is the unweighted Median classifier followed by the weighted Median classifier and the UC. The worst result is scored by the MOG. In the foreground region, the best performer is the MOG followed by the SBD classifiers and then the combined classifiers. For the whole frame according to the RMSE, the PSNR, and the overall error compared to the silhouette's motion pixels, the best result is scored by the unweighted Median, followed by the weighted Median, and then the SBD. The worst result is again scored by the MOG.

6.7 Conclusions

In this chapter we presented ways of producing better classifiers. We used the Bayes rule to combine classifiers. We combined two classifiers using an average weighted Maximum rule classifier. The resulting new classifiers all managed to score better results than their original classifiers in indoor and outdoor motion sequences. We used two different principles to combine three classifiers: weighted averaged classifiers and unweighted classifiers. Also we used different rules of combining classifiers: the Maximum, the Median, the Product and the Sum. The new classifiers, especially the Median classifiers, gave better results than some of the state of the art motion extractors, the MOG and the SBD. The Median gave more robust results compared to the other combination rules. Intuitively, the median might be better for working with disparate measures.

The weighted classifiers in general gave better results than the unweighted classifiers when the same combination rules are used except for the Median. The unweighted Median classifier gave the best result when compared to all the other classifiers.

A comparison between the performance of the two and three classifiers can not be applied with the given tests, since the combination conditions of each is different. Different weights selection procedure were used for the two and the three classifiers (exploring the two different weight selection procedures on two and three classifiers will be included in the future work). Also in the three classifiers we used more combination rules with a weighted and an unweighted principles, while in two classifiers we used only the weighted Maximum rule. Investigating these options will be left as a future work.

We can reach to an overall conclusion that optimised combination of classifiers enjoys performance advantages of separate classifiers.

Chapter 7

Conclusions and Future Work

7.1 Conclusions

In this work we provided a systematic procedure to optimise the performance of state of the art classifiers (mixture of Gaussian and Statistical Background Disturbance). Also, the same procedure was applied to the Unary Classifier.

In the assessment procedure we used five different measurement tools. Three measurements assessed each frame by using the whole scene (RMSE, PSNR, Overall error). The RMSE and the PSNR are commonly used measurement tools in pattern recognition. The overall error measurement gave a relative scale of the overall scene error to the silhouette's motion pixels. Two more measurements were used which are concerned with regional assessment to provide more information on the performance of a classifier on a certain region (background, foreground). These were analysed on images in controlled laboratory scenario, and outdoors where many parameters varied.

The Statistical Background Disturbance algorithm (Horprasert et al., 1999, 2000) is one of the state of the art motion extractors, Chapter 2. An advantage of this algorithm is its ability to classify pixels in one of four categories: foreground, background, shadow and highlights. The algorithm gives an acceptable level of shadow recognition in indoor sequences but does not perform as well in outdoor motion sequences. Also in order to accomplish a quality extraction of shadows, the quality of foreground detection is sometimes sacrificed which results in the appearance of holes on moving objects. In order to solve this problem, an improved version of the SBD was presented which mainly contributes to the improvement of shadow selection criteria by further investigating the validity of the shadow pixels through a distance measure to the background. The condition resulted in a small improvement in the shadow detection which improved the quality of foreground extraction as well.

The Mixture of Gaussians (Stauffer and Grimson, 2000, 1999) is another state of the art algorithm, Chapter 3 . The classifier performs well in extracting the foreground of indoor sequences. Also, the classifier performs well on the background but with more isolated mislabelled pixels. For the outdoor foreground and background the result is not as precise as for the indoor sequences where large holes might result in the moving object and large connected mislabelled background pixels can be seen. Also one of the main problems of this technique is the shadow region. Shadows are usually labelled as motion pixels. In order to overcome this problem, we tested different colour invariant models and found a colour model that performs well in shadow suppression though the use of this colour model resulted in the appearance of some noise in different parts of the scene.

We then established a new motion classifier which we called the modified Unary Classifier, Chapter 4. We used the original Unary Classifier for the first time as a valid classifier for the motion extraction area. The Classifier gave a high quality extraction for the foreground area but suffered noise in the background both for indoor and for outdoor motion sequences. So we modified the original classifier by improving its decision function to make the classifier competent enough to accommodate the background model. Also the classifier does not have the criteria of shadow detection or suppression. We used an invariant colour model to solve this problem. The resulting novel classifier challenged state of the art classifiers in this field and presented better results in the indoor environment and reasonably competent results in outdoor motion sequences.

Two of the classifiers, the MOG and the modified UC, had no original capability to suppress shadows. To solve the problem, we investigated colour invariant models, Chapter 5, and we monitored the final outcome of using each colour model. We noticed that some of these colour models behaved in indoor sequences more effectively than in outdoor sequences. The final outcome was an improvement in the overall performance of the classifier as a result of shadow suppression.

After the assessment of the behavior of the MOG, the improved SBD and the improved UC we noticed that each one of them performs differently in each region and may produce non-overlapping misclassifications. This diversity along with the satisfactory performance of those classifiers led us to the direction of combining those classifiers in order to produce better motion extractors, Chapter 6.

In Chapter 6, we combined the three classifiers (the MOG, the improved SBD and the improved UC) aiming to produce better motion extractors. In order to do that we used the logistic sigmoid function to estimate the probabilities for the SBD and the MOG. For the UC we used their decision function to calculate the probability.

The different combinations of two classifiers produced better results than their originating classifiers. The new MOG/SBD classifier gave the best results in indoor motion

sequences while the new UC/SBD classifier gave the best performance in outdoor sequences.

We combined three classifiers together using the Bayes decision function along with the Maximum, the Median, the Product or the Sum rules. The result of this combination produced the best motion extractor with the least overall error in indoor and outdoor sequences which is the Median classifier. Also this classifier gave the minimum error in labelling the background region.

In addition, we combined three classifiers together using weight averaged Bayes decision function using the Maximum, the Median, the Product or the Sum rules. The weighting process improved the performance of some of the classifiers (Max, Product, and Sum). The Median classifier though gave the best results among the four weighted classifiers.

7.2 Future Work

There are still many interesting questions to be answered in this research area, aiming to improve performance.

- To use different kernels with the UC. Possible kernels which await investigation include: linear, polynomial, exponential radial and sigmoid. The UC is based on the SVM theory and the performance of SVM largely depends on the used kernel (Amari and Wu, 1999). Using a different kernel may affect the overall performance of the UC classifier which might lead to a better motion extracting classifier.
- To modify the nonadaptive classifiers to be adaptive. One of the major problem with the SBD and the UC is that they are nonadaptive i.e. if the background model changes (by global lighting change or by adding an object to the background) the system might fail to identify and suppress the background which results in a noisy or a distorted output.
- The shadow extraction modification in the SBD algorithm can be further modified by adding a lower bound to the shadow region. We explained, in Chapter 2, that broad shadow thresholds can reduce foreground performance. The result was improved by conditioning the choice of the shadow region. This conditioning could be explored further.
- The HSV colour model separates the intensity (V) from the chromatic components (H,S). Thus to utilise this property for shadow extraction, we can use the HS components (one or both of them) only for the classification while excluding the (V) component. This provides an added advantage that with the value components removed much of the lighting/shading difference are absent (Bowden, 1999; Gevers and Smeulders, 1999).

- When merging three classifiers we used different combination rules (the Maximum, the Median, the Sum, and the Product). We noticed that the overall performance of the combined classifiers changed when different combination rules were used. For two classifiers, we used only the Maximum rule. Therefore, combining two classifiers can be further explored by using different combination rules which may result in a better performing classifier.
- In merging two classifiers, in all the merged classifiers we managed to obtain a low overall error. This is due to the procedure used to find a better performing classifier. In this procedure we searched for the optimal weights setting that minimises the overall error. The same procedure can be applied when combining three classifiers.
- More simple and sophisticated combination rules could also be used, like Majority Voting, Bagging and Boosting.
- Performance could be further investigated by analysing the objective performance on a larger database. In gait recognition, this could concern the variance of the feature vectors on the indoor data compared with that derived for the outdoor data.

These suggestions might improve these new systems which has already been demonstrated to meet their aims, namely improved foreground background segmentation in image sequences. Naturally we seek to analyse the advantages these new systems confer in practical applications

Appendix A

Unary Classifier sphere radius varying on motion sequences

In Chapter 4 a modified UC was presented. The modification is based on enlarging the UC size until a proper size is reached that reduces the overall error. The radius size was changed in indoor and outdoor motion sequences. The following sections of this appendix will present the results of changes the radius sizes for each motion sequence.

A.1 Indoor motion sequences

N	ROC ¹	RMSE	PSNR (dB)	FG Error ²	BG Error ³	Overall Error ⁴	σ_{RMSE}^2
0	0.910	0.286	10.879	0.212	8.816	112.038	7.982E-05
2	0.974	0.133	17.537	0.775	1.858	24.258	7.812E-05
4	0.976	0.106	19.492	1.321	1.122	15.452	6.265E-05
6	0.974	0.094	20.549	1.789	0.820	12.102	5.466E-05
8	0.971	0.087	21.240	2.233	0.644	10.319	5.073E-05
10	0.968	0.082	21.726	2.639	0.525	9.225	4.677E-05
12	0.965	0.079	22.066	3.031	0.438	8.528	4.420E-05
14	0.962	0.077	22.305	3.427	0.371	8.073	4.327E-05
16	0.958	0.076	22.441	3.838	0.318	7.826	4.306E-05

¹ ROC Optimal Cutoff Measure

² FG Error Foreground Percentage Error

³ BG Error Background Percentage Error

⁴ The percentage of the overall error compared to the motion pixels only

TABLE A.1: Assessment on the indoor motion sequences 008a013s00L using the improved UC algorithm

N	ROC ¹	RMSE	PSNR (dB)	FG Error ²	BG Error ³	Overall Error ⁴	σ_{RMSE}^2
0	0.924	0.265	11.547	0.117	7.475	109.847	7.279E-06
2	0.980	0.119	18.471	0.495	1.490	22.336	2.619E-05
4	0.982	0.098	20.191	0.881	0.968	15.041	2.910E-05
6	0.980	0.088	21.104	1.243	0.749	12.189	2.816E-05
8	0.978	0.082	21.756	1.557	0.612	10.490	2.724E-05
10	0.976	0.077	22.250	1.880	0.513	9.364	2.579E-05
12	0.974	0.074	22.625	2.193	0.438	8.590	2.442E-05
14	0.971	0.072	22.910	2.532	0.378	8.046	2.424E-05
16	0.968	0.070	23.135	2.863	0.328	7.644	2.483E-05

1 ROC Optimal Cutoff Measure

2 FG Error Foreground Percentage Error

3 BG Error Background Percentage Error

4 The percentage of the overall error compared to the motion pixels only

TABLE A.2: Assessment on the indoor motion sequences 009a017s00L using the improved UC algorithm

N	ROC ¹	RMSE	PSNR (dB)	FG Error ²	BG Error ³	Overall Error ⁴	σ_{RMSE}^2
0	0.924	0.262	11.626	0.326	7.329	110.363	6.062E-05
2	0.977	0.113	18.973	0.974	1.295	20.343	4.157E-05
4	0.977	0.093	20.648	1.467	0.825	13.807	2.446E-05
6	0.975	0.084	21.498	1.909	0.631	11.350	2.057E-05
8	0.972	0.079	22.074	2.330	0.509	9.941	1.997E-05
10	0.968	0.075	22.500	2.736	0.420	9.012	2.032E-05
12	0.965	0.073	22.809	3.126	0.352	8.393	2.098E-05
14	0.962	0.071	23.012	3.526	0.300	8.011	2.232E-05
16	0.958	0.070	23.154	3.937	0.255	7.759	2.512E-05

1 ROC Optimal Cutoff Measure

2 FG Error Foreground Percentage Error

3 BG Error Background Percentage Error

4 The percentage of the overall error compared to the motion pixels only

TABLE A.3: Assessment on the indoor motion sequences 010a024s08L using the improved UC algorithm

N	ROC ¹	RMSE	PSNR (dB)	FG Error ²	BG Error ³	Overall Error ⁴	σ_{RMSE}^2
0	0.922	0.263	11.597	0.498	7.291	132.105	4.966E-05
2	0.973	0.107	19.414	1.544	1.135	21.991	6.805E-05
4	0.970	0.085	21.489	2.346	0.628	13.640	5.580E-05
6	0.965	0.076	22.410	3.070	0.443	11.025	5.070E-05
8	0.959	0.072	22.890	3.750	0.341	9.866	4.950E-05
10	0.953	0.070	23.144	4.409	0.274	9.310	5.052E-05
12	0.947	0.069	23.250	5.054	0.225	9.092	5.328E-05
14	0.941	0.069	23.260	5.676	0.190	9.078	5.587E-05
16	0.936	0.070	23.211	6.288	0.162	9.186	5.860E-05

1 ROC Optimal Cutoff Measure

2 FG Error Foreground Percentage Error

3 BG Error Background Percentage Error

4 The percentage of the overall error compared to the motion pixels only

TABLE A.4: Assessment on the indoor motion sequences 013a037s00L using the improved UC algorithm

N	ROC ¹	RMSE	PSNR (dB)	FG Error ²	BG Error ³	Overall Error ⁴	σ_{RMSE}^2
0	0.924	0.262	11.655	0.326	7.258	115.937	7.540E-05
2	0.976	0.117	18.701	1.039	1.393	23.166	1.253E-04
4	0.975	0.097	20.285	1.598	0.915	16.101	1.070E-04
6	0.972	0.089	21.062	2.085	0.718	13.451	9.248E-05
8	0.969	0.084	21.600	2.530	0.591	11.878	8.452E-05
10	0.966	0.080	21.992	2.943	0.499	10.839	7.646E-05
12	0.962	0.077	22.292	3.346	0.427	10.102	6.996E-05
14	0.959	0.075	22.499	3.773	0.369	9.615	6.328E-05
16	0.955	0.074	22.665	4.193	0.319	9.245	5.980E-05

1 ROC Optimal Cutoff Measure

2 FG Error Foreground Percentage Error

3 BG Error Background Percentage Error

4 The percentage of the overall error compared to the motion pixels only

TABLE A.5: Assessment on the indoor motion sequences 013a040s00L using the improved UC algorithm

N	ROC ¹	RMSE	PSNR (dB)	FG Error ²	BG Error ³	Overall Error ⁴	σ_{RMSE}^2
0	0.882	0.331	9.607	0.094	11.724	164.872	8.615E-06
2	0.976	0.137	17.270	0.409	1.988	28.302	3.692E-05
4	0.981	0.108	19.378	0.696	1.196	17.450	4.554E-05
6	0.981	0.097	20.248	0.944	0.954	14.283	4.425E-05
8	0.980	0.091	20.807	1.173	0.814	12.556	4.072E-05
10	0.979	0.087	21.251	1.388	0.712	11.335	3.905E-05
12	0.978	0.083	21.612	1.608	0.632	10.428	3.697E-05
14	0.976	0.080	21.918	1.819	0.566	9.719	3.537E-05
16	0.975	0.078	22.189	2.028	0.509	9.127	3.284E-05

¹ ROC Optimal Cutoff Measure² FG Error Foreground Percentage Error³ BG Error Background Percentage Error⁴ The percentage of the overall error compared to the motion pixels only

TABLE A.6: Assessment on the indoor motion sequences 017a054s00L using the improved UC algorithm

N	ROC ¹	RMSE	PSNR (dB)	FG Error ²	BG Error ³	Overall Error ⁴	σ_{RMSE}^2
0	0.920	0.270	11.373	0.181	7.782	114.147	2.264E-05
2	0.980	0.118	18.554	0.521	1.465	21.929	4.774E-05
4	0.983	0.094	20.582	0.855	0.885	13.776	4.491E-05
6	0.981	0.083	21.644	1.202	0.657	10.788	3.799E-05
8	0.979	0.077	22.352	1.546	0.522	9.159	3.314E-05
10	0.977	0.072	22.863	1.905	0.428	8.140	3.155E-05
12	0.973	0.069	23.217	2.300	0.357	7.505	3.132E-05
14	0.970	0.067	23.460	2.686	0.302	7.097	3.157E-05
16	0.967	0.066	23.621	3.062	0.259	6.839	3.160E-05

¹ ROC Optimal Cutoff Measure² FG Error Foreground Percentage Error³ BG Error Background Percentage Error⁴ The percentage of the overall error compared to the motion pixels only

TABLE A.7: Assessment on the indoor motion sequences 017a055s00R using the improved UC algorithm

N	ROC ¹	RMSE	PSNR (dB)	FG Error ²	BG Error ³	Overall Error ⁴	σ_{RMSE}^2
0	0.919	0.273	11.290	0.172	7.966	111.480	6.465E-05
2	0.977	0.126	18.014	0.627	1.662	23.816	6.785E-05
4	0.979	0.103	19.775	1.066	1.062	15.847	4.760E-05
6	0.978	0.092	20.718	1.439	0.813	12.739	3.812E-05
8	0.976	0.085	21.385	1.788	0.657	10.912	3.152E-05
10	0.973	0.081	21.866	2.105	0.551	9.761	2.734E-05
12	0.971	0.078	22.222	2.432	0.472	8.988	2.477E-05
14	0.969	0.075	22.522	2.741	0.407	8.388	2.351E-05
16	0.966	0.073	22.735	3.055	0.355	7.985	2.278E-05

¹ ROC Optimal Cutoff Measure² FG Error Foreground Percentage Error³ BG Error Background Percentage Error⁴ The percentage of the overall error compared to the motion pixels only

TABLE A.8: Assessment on the indoor motion sequences 018a060s00L using the improved UC algorithm

N	ROC ¹	RMSE	PSNR (dB)	FG Error ²	BG Error ³	Overall Error ⁴	σ_{RMSE}^2
0	0.926	0.262	11.651	0.104	7.284	110.984	1.188E-05
2	0.985	0.101	19.930	0.451	1.059	16.499	2.360E-05
4	0.986	0.081	21.807	0.796	0.656	10.721	2.573E-05
6	0.984	0.073	22.726	1.135	0.498	8.675	2.245E-05
8	0.981	0.068	23.342	1.484	0.400	7.528	2.059E-05
10	0.978	0.065	23.784	1.865	0.326	6.801	2.092E-05
12	0.975	0.063	24.067	2.270	0.272	6.378	2.239E-05
14	0.971	0.062	24.235	2.693	0.228	6.136	2.295E-05
16	0.967	0.061	24.306	3.128	0.193	6.042	2.483E-05

¹ ROC Optimal Cutoff Measure² FG Error Foreground Percentage Error³ BG Error Background Percentage Error⁴ The percentage of the overall error compared to the motion pixels only

TABLE A.9: Assessment on the indoor motion sequences 019a063s00L using the improved UC algorithm

A.2 Outdoor motion sequences

N	ROC ¹	RMSE	PSNR (dB)	FG Error ²	BG Error ³	Overall Error ⁴	σ_{RMSE}^2
0	0.820	0.330	9.625	6.418	11.618	81.768	1.683E-04
2	0.833	0.206	13.753	14.012	2.720	31.691	2.365E-04
4	0.791	0.202	13.915	19.138	1.740	30.414	1.774E-04
6	0.756	0.206	13.723	23.015	1.345	31.711	1.564E-04
8	0.722	0.214	13.425	26.653	1.128	33.939	1.617E-04
10	0.693	0.220	13.159	29.719	0.982	36.053	1.561E-04
12	0.667	0.226	12.925	32.406	0.874	38.043	1.558E-04
14	0.642	0.232	12.691	34.985	0.799	40.137	1.480E-04
16	0.619	0.238	12.487	37.382	0.727	42.064	1.498E-04

1 ROC Optimal Cutoff Measure

2 FG Error Foreground Percentage Error

3 BG Error Background Percentage Error

4 The percentage of the overall error compared to the motion pixels only

TABLE A.10: Assessment on the outdoor motion sequences 008e013s00L using the improved UC algorithm

N	ROC ¹	RMSE	PSNR (dB)	FG Error ²	BG Error ³	Overall Error ⁴	σ_{RMSE}^2
0	0.744	0.314	10.086	16.709	8.882	79.158	3.189E-04
2	0.638	0.248	12.102	33.998	2.218	49.529	9.127E-05
4	0.551	0.259	11.726	43.353	1.519	53.987	9.738E-05
6	0.488	0.270	11.371	50.005	1.226	58.586	1.108E-04
8	0.439	0.279	11.100	55.043	1.044	62.347	1.109E-04
10	0.404	0.285	10.906	58.722	0.926	65.198	1.165E-04
12	0.373	0.291	10.734	61.836	0.857	67.831	1.069E-04
14	0.348	0.295	10.602	64.390	0.791	69.921	1.009E-04
16	0.328	0.299	10.499	66.453	0.737	71.603	1.033E-04

1 ROC Optimal Cutoff Measure

2 FG Error Foreground Percentage Error

3 BG Error Background Percentage Error

4 The percentage of the overall error compared to the motion pixels only

TABLE A.11: Assessment on the outdoor motion sequences 009e017s01L using the improved UC algorithm

N	ROC ¹	RMSE	PSNR (dB)	FG Error ²	BG Error ³	Overall Error ⁴	σ_{RMSE}^2
0	0.816	0.343	9.298	5.610	12.800	83.350	3.243E-05
2	0.820	0.236	12.567	13.808	4.228	39.512	9.564E-05
4	0.761	0.236	12.566	20.857	3.094	39.722	1.954E-04
6	0.703	0.245	12.234	27.081	2.600	42.995	2.494E-04
8	0.658	0.253	11.967	31.897	2.260	45.770	2.916E-04
10	0.626	0.258	11.772	35.436	2.018	47.839	2.825E-04
12	0.600	0.263	11.618	38.224	1.830	49.479	2.545E-04
14	0.576	0.267	11.466	40.749	1.686	51.128	2.170E-04
16	0.556	0.271	11.352	42.811	1.552	52.367	1.796E-04

¹ ROC Optimal Cutoff Measure² FG Error Foreground Percentage Error³ BG Error Background Percentage Error⁴ The percentage of the overall error compared to the motion pixels only

TABLE A.12: Assessment on the outdoor motion sequences 010e024s00L using the improved UC algorithm

N	ROC ¹	RMSE	PSNR (dB)	FG Error ²	BG Error ³	Overall Error ⁴	σ_{RMSE}^2
0	0.879	0.275	11.216	3.981	8.096	60.877	1.339E-04
2	0.897	0.158	16.021	8.617	1.640	20.140	7.644E-05
4	0.868	0.153	16.307	12.249	0.933	18.815	6.633E-05
6	0.838	0.159	16.007	15.488	0.670	20.202	1.289E-04
8	0.811	0.166	15.632	18.433	0.523	22.110	2.254E-04
10	0.784	0.174	15.250	21.207	0.435	24.272	3.622E-04
12	0.758	0.181	14.915	23.837	0.360	26.379	5.329E-04
14	0.732	0.188	14.590	26.458	0.306	28.620	7.442E-04
16	0.710	0.194	14.325	28.707	0.270	30.610	9.448E-04

¹ ROC Optimal Cutoff Measure² FG Error Foreground Percentage Error³ BG Error Background Percentage Error⁴ The percentage of the overall error compared to the motion pixels only

TABLE A.13: Assessment on the outdoor motion sequences 013e037s00L using the improved UC algorithm

N	ROC ¹	RMSE	PSNR (dB)	FG Error ²	BG Error ³	Overall Error ⁴	σ_{RMSE}^2
0	0.663	0.376	8.506	20.397	13.335	120.390	3.501E-04
2	0.613	0.260	11.708	35.773	2.889	57.453	2.494E-05
4	0.543	0.259	11.731	43.961	1.763	57.197	6.092E-05
6	0.492	0.265	11.555	49.467	1.349	59.595	9.699E-05
8	0.451	0.270	11.378	53.779	1.107	62.091	1.080E-04
10	0.418	0.275	11.222	57.248	0.945	64.346	1.089E-04
12	0.388	0.280	11.074	60.331	0.830	66.567	1.039E-04
14	0.364	0.283	10.957	62.857	0.735	68.383	9.904E-05
16	0.344	0.287	10.858	64.973	0.662	69.947	9.132E-05

1 ROC Optimal Cutoff Measure

2 FG Error Foreground Percentage Error

3 BG Error Background Percentage Error

4 The percentage of the overall error compared to the motion pixels only

TABLE A.14: Assessment on the outdoor motion sequences 013e040s00L using the improved UC algorithm

N	ROC ¹	RMSE	PSNR (dB)	FG Error ²	BG Error ³	Overall Error ⁴	σ_{RMSE}^2
0	0.738	0.312	10.118	17.726	8.454	69.696	1.621E-04
2	0.633	0.257	11.820	34.688	2.041	47.166	3.969E-04
4	0.548	0.269	11.434	43.971	1.229	51.492	3.524E-04
6	0.490	0.279	11.105	50.147	0.877	55.516	3.583E-04
8	0.444	0.288	10.840	54.930	0.661	58.982	3.461E-04
10	0.407	0.295	10.621	58.793	0.525	62.015	3.549E-04
12	0.379	0.300	10.463	61.646	0.436	64.317	3.739E-04
14	0.355	0.305	10.322	64.172	0.370	66.432	3.821E-04
16	0.334	0.309	10.209	66.243	0.317	68.176	3.834E-04

1 ROC Optimal Cutoff Measure

2 FG Error Foreground Percentage Error

3 BG Error Background Percentage Error

4 The percentage of the overall error compared to the motion pixels only

TABLE A.15: Assessment on the outdoor motion sequences 017e054s00L using the improved UC algorithm

N	ROC ¹	RMSE	PSNR (dB)	FG Error ²	BG Error ³	Overall Error ⁴	σ_{RMSE}^2
0	0.803	0.329	9.653	8.528	11.199	85.951	1.552E-04
2	0.789	0.216	13.312	18.372	2.684	36.851	5.556E-05
4	0.742	0.213	13.433	24.040	1.720	35.875	1.188E-04
6	0.706	0.217	13.269	28.058	1.343	37.277	1.421E-04
8	0.678	0.221	13.108	31.107	1.107	38.694	1.441E-04
10	0.656	0.224	12.999	33.449	0.912	39.707	1.549E-04
12	0.635	0.228	12.857	35.788	0.765	41.028	1.703E-04
14	0.614	0.232	12.707	37.986	0.657	42.488	1.751E-04
16	0.595	0.236	12.572	39.934	0.570	43.842	1.903E-04

1 ROC Optimal Cutoff Measure

2 FG Error Foreground Percentage Error

3 BG Error Background Percentage Error

4 The percentage of the overall error compared to the motion pixels only

TABLE A.16: Assessment on the outdoor motion sequences 017e055s00R using the improved UC algorithm

N	ROC ¹	RMSE	PSNR (dB)	FG Error ²	BG Error ³	Overall Error ⁴	σ_{RMSE}^2
0	0.858	0.294	10.623	4.960	9.256	63.491	3.101E-05
2	0.867	0.191	14.369	10.720	2.546	26.821	4.980E-05
4	0.832	0.188	14.545	15.117	1.677	25.728	3.786E-05
6	0.794	0.194	14.257	19.277	1.303	27.521	8.233E-05
8	0.759	0.201	13.929	23.007	1.061	29.722	1.423E-04
10	0.728	0.209	13.618	26.290	0.903	31.999	2.401E-04
12	0.695	0.218	13.273	29.687	0.795	34.710	3.268E-04
14	0.667	0.225	12.986	32.591	0.715	37.113	3.842E-04
16	0.641	0.232	12.734	35.290	0.642	39.354	4.251E-04

1 ROC Optimal Cutoff Measure

2 FG Error Foreground Percentage Error

3 BG Error Background Percentage Error

4 The percentage of the overall error compared to the motion pixels only

TABLE A.17: Assessment on the outdoor motion sequences 018e060s00L using the improved UC algorithm

N	ROC ¹	RMSE	PSNR (dB)	FG Error ²	BG Error ³	Overall Error ⁴	σ_{RMSE}^2
0	0.878	0.266	11.512	4.914	7.284	76.572	2.061E-05
2	0.872	0.160	15.929	11.136	1.697	27.726	1.064E-04
4	0.837	0.156	16.134	15.103	1.165	26.511	1.322E-04
6	0.812	0.158	16.039	17.876	0.944	27.131	1.419E-04
8	0.793	0.160	15.925	19.920	0.811	27.875	1.565E-04
10	0.776	0.163	15.797	21.725	0.714	28.732	1.682E-04
12	0.762	0.165	15.690	23.177	0.639	29.453	1.701E-04
14	0.748	0.167	15.581	24.618	0.566	30.176	1.540E-04
16	0.739	0.168	15.525	25.557	0.510	30.569	1.487E-04

1 ROC Optimal Cutoff Measure

2 FG Error Foreground Percentage Error

3 BG Error Background Percentage Error

4 The percentage of the overall error compared to the motion pixels only

TABLE A.18: Assessment on the outdoor motion sequences 019e063s05L using the improved UC algorithm

Appendix B

MOG parameters setting procedure using the $c_1c_2c_3$ colour model

The tests are performed with the following basic parameter settings: a learning rate of 0.005 for indoor sequences and 0.01 for outdoor sequences, an initial weight of 0.05, a background threshold of 0.6, an initial variance set to 0.01 for indoor sequences and 0.002 for outdoor sequences, a background threshold of 0.6, and 5 Gaussians per mixture. In each test one of the parameters will be varied while having the other parameters fixed until a suitable value is reached.

- **The learning rate (α)**

The initial learning rate can be set between [0-1]. The effect of changing the learning rate will be shown first through testing variable learning rates on an indoor motion sequence. The MOG background adaption time will be tested using different learning rates. Also we well be measuring the effect of this change on the performance of the algorithm in motion detection.

The following learning rates were tested: 0.1, 0.01, and 0.001. 50 background frames were used to test how fast the algorithm will adapt to the background. Figure B.1(a) shows the effect of using different learning rates on background adaptation. Using a very slow learning rate of 0.001 resulted on the system finishing all the background frames without adapting to the background. On the other extreme using a fast adaption rate of 0.1 made the system adapt in 3 frames. Such fast adaption is not suitable since a foreground object having a wide single colour surface can be considered as background in such fast adaption. A learning rate of 0.01 behaved reasonably.

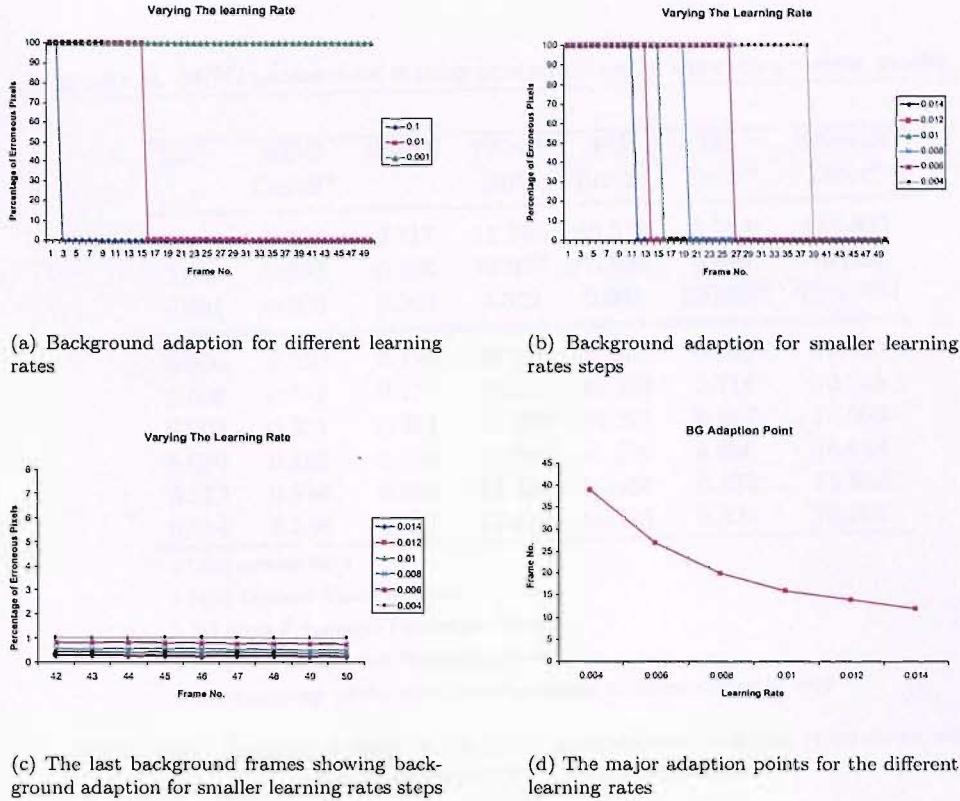


FIGURE B.1: The effect of varying the learning rate on background adaption for MOG with a $c_1c_2c_3$ colour model using an indoor motion sequence

In Figure B.1(b) small steps of learning rates around 0.01 are further explored. Figure B.1(c) gives a very close look at the last 9 frames error for the small steps used. All the learning rates adapted within the 50 background frames with the fastest learning rate, 0.014, adapting within 12 frames and the slowest, 0.004, adapting in 39 frames. Figure B.1(d) gives the adaption frame for the learning rate used.

To choose a specific value of adaptation among the values shown in Figure B.1(b), the algorithm was tested with an indoor motion sequence after the first 50 background frames. Table B.1 shows the result.

The motion sequence with the different learning rates was analysed with different assessment tests ROC optimal cutoff measure; the RMSE; the PSNR; and the percentage of error for the foreground, background and the overall error compared to the silhouette motion pixels, Table B.1.

The first three rows show different scales of learning rates: 0.1, 0.01, and 0.001. Learning rate 0.01 gave the minimum RMSE, the highest PSNR, and

LR ¹	ROC Cutoff ²	RMSE	PSNR (dB)	FG Error ³	BG Error ⁴	Overall Error ⁵
0.1	0.309	0.277	11.164	65.510	3.563	114.800
0.01	0.338	0.224	12.986	65.526	0.686	75.044
0.001	0.000	0.966	0.302	0.002	100.000	1394.400
0.004	0.939	0.110	19.150	5.163	0.935	18.189
0.006	0.587	0.179	15.313	40.583	0.714	50.516
0.008	0.351	0.221	13.108	64.285	0.627	73.002
0.010	0.338	0.224	12.986	65.526	0.686	75.044
0.012	0.338	0.225	12.950	65.466	0.736	75.668
0.014	0.338	0.227	12.880	65.425	0.829	76.907

1 LR Learning Rate

2 ROC Optimal Cutoff Measure

3 FG Error Foreground Percentage Error

4 BG Error Background Percentage Error

5 The percentage of the overall error compared to the motion pixels only

TABLE B.1: Assessment tests on the MOG using different learning rates on an indoor motion sequence with a $c_1c_2c_3$ colour model

the minimum overall error. 0.01 learning rate also gave the minimum background error. Learning rate 0.001 gave the minimum foreground error, but this is due to the fact that with this learning rate the system did not adapt to the background (notice 100% background error and see Figure B.1(a)). Thus the system with 0.001 was labelling erroneously all the scene as foreground which led to the minimum foreground error (almost zero).

Then we started investigating learning rates near 0.01 with smaller learning steps. The values were bounded between 0.004 and 0.014. We choose not to go below 0.004 because this adaption rate is already so slow that it took the system 39 frames to adapt. Also using value above than 0.014 will give a fast adaption leading to increase in the foreground error due to its being adapted to the background (notice when we increased the learning rate reaching up to 0.014 the foreground error had already increased to 65.43%, Table B.1). Learning rate 0.004 gave the best result in most of the shown columns. Learning rate 0.004 gave the highest ROC optimal cutoff value, minimum RMSE, highest PSNR, minimum foreground error and the minimum overall error, 18.19%. The background error was small with a value less than 1% but the it was the highest value among all the small steps tested. The results scored by learning rate 0.004 led us to choose it for indoor motion sequences testing.

For the outdoor assessment of 50 background frames for the learning rate parameter different scales of learning rates were used. Testing learning rates: 0.1, 0.01, and 0.001 resulted in Figure B.2(a). Using learning rate 0.001

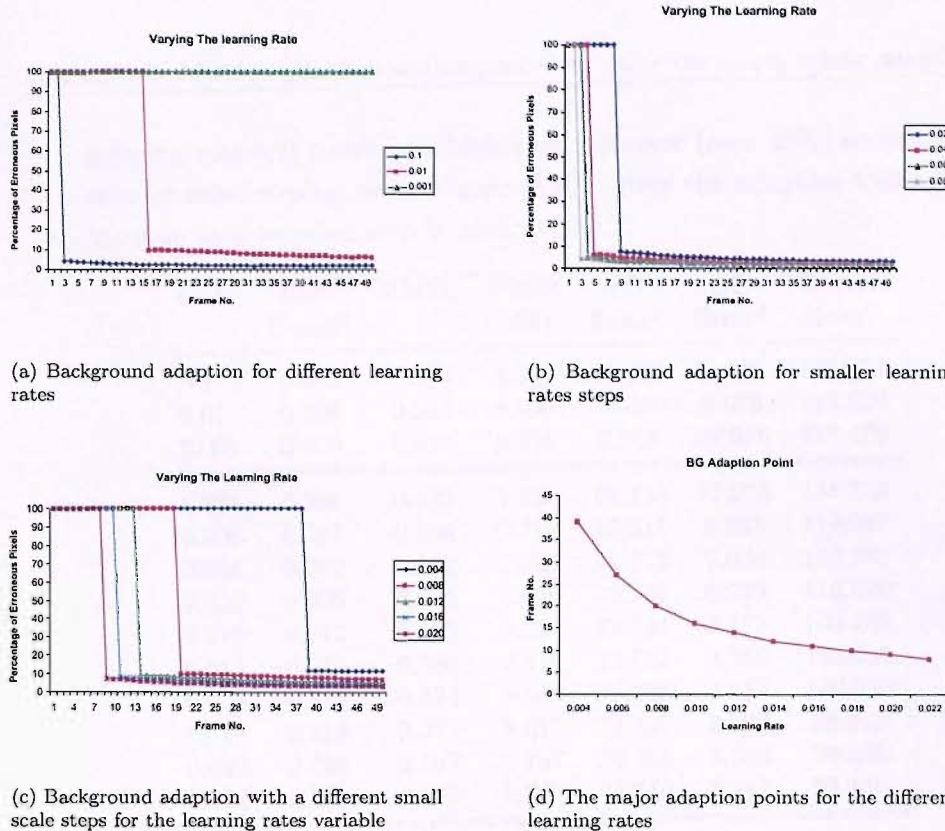


FIGURE B.2: The effect of varying the learning rate on background adaption for MOG with a $c_1c_2c_3$ colour model using an outdoor motion sequence

resulted on the system finishing all the background frames without adapting to the background. When a fast adaption rate of 0.1 was used, the system adapted to the background in 2 frames while maintaining a minimal error in the rest of the tested background frames. The adaption speed of 0.01 is more appropriate than the other two learning rates, 0.1 and 0.001. Then we tested with two different step scales of learning rates one with 0.02 starting from 0.02 to 0.08, Figure B.2(b), and another with 0.004 step value starting from 0.004 to 0.20, Figure B.2(c).

In Figure B.2(b) we notice the system is adapting quickly with the values used, 0.2 - 0.8. The system adapted in 9 frames in 0.02 while in the 0.08 it adapted in 3 frames.

More testing values are used in Figure B.2(c) with slower adapting rates. The adapting rates started adapting in 9 frames in 0.02 reaching up to 39 frames in 0.004. We notice that the slower the adaption time, the higher the error after adaption. 0.004 which is the slowest learning error used in this figure gave an error around 10% after adaption. Continuing with slower

learning rate will result in a higher overall error (over 10%) so we choose to stop at this learning rate. Figure B.2(d) gives the adaption frame for each learning rate (smaller step is used, 0.002).

LR ¹	ROC Cutoff ²	RMSE	PSNR (dB)	FG Error ³	BG Error ⁴	Overall Error ⁵
0.1	0.246	0.386	8.278	69.282	6.106	106.664
0.01	0.206	0.394	8.098	73.338	6.073	110.820
0.001	0.000	0.927	0.658	0.003	99.985	617.476
<hr/>						
0.004	0.246	0.432	7.288	64.134	11.273	133.723
0.006	0.257	0.408	7.789	65.611	8.658	119.053
0.008	0.212	0.402	7.924	71.712	7.074	115.363
0.010	0.206	0.394	8.098	73.338	6.073	110.820
0.012	0.212	0.385	8.287	73.534	5.279	106.102
0.014	0.215	0.380	8.418	73.727	4.736	102.937
0.016	0.219	0.374	8.542	73.905	4.237	100.030
0.018	0.218	0.371	8.617	74.327	3.893	98.306
0.020	0.220	0.367	8.707	74.468	3.546	96.285
0.022	0.221	0.365	8.760	74.540	3.347	95.128
<hr/>						
0.02	0.220	0.367	8.707	74.468	3.546	96.285
0.04	0.234	0.365	8.767	72.990	3.597	95.026
0.06	0.244	0.372	8.613	71.176	4.476	98.575
0.08	0.248	0.380	8.426	69.785	5.436	103.041

1 LR Learning Rate

2 ROC Optimal Cutoff Measure

3 FG Error Foreground Percentage Error

4 BG Error Background Percentage Error

5 The percentage of the overall error compared to the motion pixels only

TABLE B.2: Assessment tests on the MOG using different learning rates on an outdoor motion sequence with a $c_1c_2c_3$ colour model

Testing the learning rate with the large variations resulted in the upper three rows of Table B.2. The table shows that learning rate 0.001 classified most of the pixels as foreground (foreground error close to 0 and the background error close to 100%). This learning rate gave the worst error in most of the measurement (ROC optimal cutoff value, RMSE, PSNR, background error and the overall error). Learning rates 0.1 and 0.01 gave more reasonable values than 0.001 but the error is still high.

When we tested the values around 0.01 the middle rows of the table are obtained. The overall error decreases as we increase the learning rate value. The RMSE, the PSNR and the background error are all improving as we increase the learning rate. The foreground error behaves in an opposite way were it increases as we increase the learning rate. 0.022 gave the minimum RMSE, the highest PSNR, the minimum background error, and the minimum

overall error compared to the silhouette's motion pixels. However 0.022 gave the highest foreground error.

When changing the step scale to a larger step value, the bottom part of Table B.2) is produced. At learning rate 0.04 the optimal cutoff measure is not the best nor the foreground error but the RMSE, the PSNR, and the overall error are giving the best values at this learning rate. The background error is the second best in 0.04. Therefore we will use 0.04 for testing outdoor motion sequences.

- **Initial weight**

The initial weight can be set to any value larger than zero. We will discover the effect of changing the initial weight through testing different initial weights on an indoor sequence first.

Initial Weight	ROC Cutoff ¹	RMSE	PSNR (dB)	FG Error ²	BG Error ³	Overall Error ⁴
0.5	0.835	0.129	18.092	15.736	0.779	26.582
0.05	0.835	0.129	18.092	15.736	0.779	26.582
0.005	0.835	0.129	18.092	15.736	0.779	26.582
0.0005	0.835	0.129	18.092	15.736	0.779	26.582

1 ROC Optimal Cutoff Measure

2 FG Error Foreground Percentage Error

3 BG Error Background Percentage Error

4 The percentage of the overall error compared to the motion pixels only

TABLE B.3: Assessment tests on the MOG using different Initial Weigh on an indoor motion sequence with a $c_1c_2c_3$ colour model

The indoor motion sequence assessment result is shown in Table B.3. The change of the initial weight has no effect on the performance of the system extraction. We will set the initial weight to 0.0005 for extracting indoor motion sequences.

Initial Weight	ROC Cutoff ¹	RMSE	PSNR (dB)	FG Error ²	BG Error ³	Overall Error ⁴
0.5	0.206	0.394	8.098	73.338	6.073	110.820
0.05	0.206	0.394	8.098	73.338	6.073	110.820
0.005	0.206	0.394	8.098	73.338	6.073	110.820
0.0005	0.206	0.394	8.098	73.338	6.073	110.820

1 ROC Optimal Cutoff Measure

2 FG Error Foreground Percentage Error

3 BG Error Background Percentage Error

4 The percentage of the overall error compared to the motion pixels only

TABLE B.4: Assessment tests on the MOG using different Initial Weigh on an outdoor motion sequence with a $c_1c_2c_3$ colour model

The tests on the outdoor motion sequence is given in Table B.4. The table shows that varying the MOG initial weight is not significant on extracting outdoor motion sequences. The initial weight will be set the same initial weight as in indoor sequence, 0.0005.

- **The background threshold (T)**

The background threshold, T , can have a value of $0 < T \leq 1$. The larger T the more chance for more Gaussians to be considered as part of the background model. Smaller value of T will allow only fewer number of Gaussians to be in the background model.

T^1	ROC Cutoff ²	RMSE	PSNR (dB)	FG Error ³	BG Error ⁴	Overall Error ⁵
0.2	0.941	0.105	19.548	5.047	0.829	16.599
0.4	0.942	0.104	19.707	5.047	0.787	16.011
0.6	0.835	0.129	18.092	15.736	0.779	26.582
0.8	0.335	0.224	12.995	65.893	0.647	74.897

1 T Background Threshold

2 ROC Optimal Cutoff Measure

3 FG Error Foreground Percentage Error

4 BG Error Background Percentage Error

5 The percentage of the overall error compared to the motion pixels only

TABLE B.5: Assessment tests on the MOG using different background thresholds for an indoor motion sequence with a $c_1c_2c_3$ colour model

For the indoor sequences, Table B.5 shows an assessment for varying the background threshold parameter on the MOG algorithm. We notice that the best performance for the algorithm is obtained in background threshold, T , of 0.2 and 0.4. 0.4 is slightly better than 0.2 in the ROC optimal cutoff measure, the RMSE, the PSNR, the background error, and the overall error compared to the silhouette's motion pixels. In the foreground error both thresholds gave the same minimum error. Thus the extraction of indoor motion sequences will be done through using 0.4 as a background threshold.

Table B.6 gives the assessment result for varying the background threshold on an outdoor motion sequence. 0.4 and 0.8 are the best performing thresholds in this table. 0.8 is slightly better in the RMSE, the PSNR, the background error, and the overall error. On the other hand, 0.4 is better in the ROC optimal cutoff measure. Also it is better in a reasonable percent in the foreground error (almost 10%). Therefore 0.4 will be used as a background threshold for extracting outdoor sequences.

- **Initial variance**

T^1	ROC	RMSE	PSNR	FG Error ³	BG Error ⁴	Overall Error ⁵
	Cutoff ²		(dB)			
0.2	0.297	0.379	8.438	64.019	6.240	102.544
0.4	0.295	0.375	8.526	64.660	5.799	100.470
0.6	0.206	0.394	8.098	73.338	6.073	110.820
0.8	0.211	0.373	8.583	74.998	3.900	99.114

1 T Background Threshold

2 ROC Optimal Cutoff Measure

3 FG Error Foreground Percentage Error

4 BG Error Background Percentage Error

5 The percentage of the overall error compared to the motion pixels only

TABLE B.6: Assessment tests on the MOG using different background thresholds for an outdoor motion sequence with a $c_1c_2c_3$ colour model

The initial variance is expected to be large enough to accommodate a normal background pixel variations. The effect of changing the initial variance will be shown through testing variable initial variances on an indoor and outdoor motion sequences.

Init. Var. ¹	ROC Cutoff ²	RMSE	PSNR (dB)	FG Error ³	BG Error ⁴	Overall Error ⁵
0.5	0.000	0.259	11.735	100.000	0.000	100.000
0.05	0.409	0.202	13.902	58.927	0.171	61.194
0.005	0.860	0.156	16.300	12.193	1.789	37.098
0.01	0.835	0.129	18.092	15.736	0.779	26.582
0.03	0.729	0.136	17.582	26.921	0.143	28.873
0.05	0.409	0.202	13.902	58.927	0.171	61.194
0.07	0.222	0.232	12.699	77.644	0.191	80.162
0.09	0.126	0.245	12.240	87.261	0.136	89.051
0.001	0.861	0.286	10.879	5.557	8.352	122.070
0.003	0.872	0.179	15.038	10.132	2.721	48.033
0.005	0.860	0.156	16.300	12.193	1.789	37.098
0.007	0.850	0.141	17.221	13.797	1.239	31.049
0.009	0.840	0.132	17.862	15.156	0.898	27.660

1 Init. Var. Initial Variance

2 ROC Optimal Cutoff Measure

3 FG Error Foreground Percentage Error

4 BG Error Background Percentage Error

5 The percentage of the overall error compared to the motion pixels only

TABLE B.7: Assessment tests on the MOG using different Initial Variances for an indoor motion sequence with a $c_1c_2c_3$ colour model

The assessment table for the initial variance, Table B.7, was done in three parts. In part one, the test was performed using the following initial variances: 0.5, 0.05, and 0.005. We notice that in initial variance 0.5, all the

pixels are classified as background pixels. The smallest initial variances, 0.005, gave the best values in all the measurement except for the foreground error where it has increased slightly. The results indicates that an optimal value, if it exists, will be smaller than 0.05. The value might fall between 0.05 and 0.005 or even smaller than 0.005.

In the second part we tested initial variance between 0.01 and 0.09 in a step of 0.2. 0.01 gave the best result in all the measurement except the background error where it gave the worst error among the other tested initial variance but with a small error, 0.78%.

In the third part we used smaller values to test whether the error decrementing rate will continue. The values were set between 0.001 and 0.009 with a step of 0.002. The error in this stage changed the trend and increased as we went in decreasing the initial variance from 0.009 to 0.001. The change in trend can be noticed in the optimal ROC cutoff measure, the RMSE, the PSNR, the background error, and the overall percentage of error compared to the silhouette's motion pixels. The foreground error is the only measurement that continued to decrease but it is negligible when comparing this decrement with the amount of the overall error increment. From the given table we choose 0.01 as an initial variance to extract indoor motion sequences.

Init. Var. ¹	ROC Cutoff ²	RMSE	PSNR (dB)	FG Error ³	BG Error ⁴	Overall Error ⁵
0.5	0.000	0.374	8.543	100.000	0.000	100.000
0.05	0.040	0.383	8.357	94.344	1.653	104.423
0.005	0.145	0.373	8.571	82.822	2.685	99.363
0.01	0.110	0.375	8.535	86.810	2.184	100.192
0.03	0.069	0.386	8.288	90.597	2.551	106.185
0.05	0.040	0.383	8.357	94.344	1.653	104.423
0.07	0.025	0.379	8.447	96.528	0.938	102.254
0.09	0.017	0.378	8.468	97.606	0.680	101.751
0.001	0.266	0.429	7.358	62.175	11.238	131.497
0.003	0.175	0.381	8.386	78.441	4.093	103.692
0.005	0.145	0.373	8.571	82.822	2.685	99.363
0.007	0.125	0.373	8.577	85.235	2.277	99.212
0.009	0.113	0.374	8.548	86.471	2.191	99.890

¹ Init. Var. Initial Variance

² ROC Optimal Cutoff Measure

³ FG Error Foreground Percentage Error

⁴ BG Error Background Percentage Error

⁵ The percentage of the overall error compared to the motion pixels only

TABLE B.8: Assessment tests on the MOG using different Initial Variances for an outdoor motion sequence with a $c_1c_2c_3$ colour model

Table B.8 gives the evaluation tables for three stages of changing the initial variance for an outdoor motion sequence. In the first stage we tested the following initial variance: 0.5, 0.05, and 0.005. In initial variance 0.5 all the pixels are classified as background which means the initial variance size for the Gaussians is very large. In 0.05 and 0.005 the foreground error decreased while the background error increased. 0.005 gave the best values in all the measurements except in the background error where it had increased. The best setting for the initial variance in outdoor sequences is expected to be in between 0.05 and 0.005 or smaller than 0.005.

In the second stage the values between 0.01 and 0.09 are tested with a step of 0.2. We notice that the best values are scored by 0.01 in the ROC optimal cutoff measure, the RMSE, the PSNR, the foreground error, and the overall error compared to the silhouette's motion pixels. The minimum background error is scored by initial variance 0.09.

In the third stage we covered the values from 0.001 to 0.009 with a step of 0.002. In this stage, 0.007 scored the best values in the RMSE, the PSNR, the overall error when compared to the silhouette's motion pixels. 0.007 also scored the second best value in the foreground error. In all the three stages 0.007 scored the best value in RMSE, the PSNR, and the overall error. Therefore we will use this initial variance to extract outdoor motion sequences.

- **Number of Gaussians per mixture**

The number of Gaussians was varied from 2 to 9 Gaussians per pixel to test the effect of such change on the system performance. Bearing in mind that more Gaussians means more speed degrade. Also we started with two Gaussians (not one), since this is the minimum possible number of Gaussians where one will be used for the background model and the other is used for motion pixels. The test was performed with the background threshold set to 0.2.

The evaluation of using different number of Gaussians on an indoor motion sequence is shown in Table B.9. Using 7 mixture of Gaussians gave the best result in the optimal ROC cutoff measure, the RMSE, the PSNR, the foreground error, and the overall error when compared to the to the silhouette's motion pixels. Also the 7 mixture of Gaussians scored close to the best performer in the background error. 7 Gaussians per pixels will be used for indoor motion sequences.

Table B.10 shows the evaluation of changing the number of Gaussians for the MOG algorithm using an outdoor motion sequence. 7 Gaussians scored the second best in the ROC optimal cutoff measure, and the background error.

No. of Gauss. ¹	ROC Cutoff ²	RMSE	PSNR (dB)	FG Error ³	BG Error ⁴	Overall Error ⁵
2	-0.274	0.814	1.824	60.359	67.086	999.741
3	0.257	0.340	9.695	65.231	9.076	190.564
4	0.620	0.182	15.103	36.901	1.077	51.900
5	0.835	0.129	18.092	15.736	0.779	26.582
7	0.941	0.088	21.181	5.426	0.431	11.428
9	0.903	0.095	20.562	9.452	0.291	13.492

1 No. of Gauss. Number of Gaussians

2 ROC Optimal Cutoff Measure

3 FG Error Foreground Percentage Error

4 BG Error Background Percentage Error

5 The percentage of the overall error compared to the motion pixels only

TABLE B.9: Assessment tests on the MOG using different number of Gaussians per pixel for an indoor motion sequence with a $c_1c_2c_3$ colour model

No. of Gauss. ¹	ROC Cutoff ²	RMSE	PSNR (dB)	FG Error ³	BG Error ⁴	Overall Error ⁵
2	0.108	0.494	6.130	72.583	16.643	174.846
3	0.259	0.422	7.502	63.811	10.282	127.275
4	0.282	0.396	8.045	63.945	7.833	112.285
5	0.297	0.379	8.438	64.019	6.240	102.544
7	0.288	0.363	8.799	66.752	4.466	94.316
9	0.241	0.364	8.792	72.260	3.598	94.461

1 No. of Gauss. Number of Gaussians

2 ROC Optimal Cutoff Measure

3 FG Error Foreground Percentage Error

4 BG Error Background Percentage Error

5 The percentage of the overall error compared to the motion pixels only

TABLE B.10: Assessment tests on the MOG using different number of Gaussians per pixel for an outdoor motion sequence with a $c_1c_2c_3$ colour model

Also 7 mixture of Gaussians scored the best result in the RMSE, the PSNR, and the overall error compared to the silhouette's motion pixels. Therefore 7 mixture of Gaussians will be used for outdoor motion sequences extraction.

Appendix C

Combined Classifiers Detailed Assessment results

In Chapter 6, the overall combination results are presented for two and three classifiers combination. In this appendix we will present the detailed assessment tables for those combinations.

C.1 Two Classifiers

C.1.1 Indoor motion sequences

Sequence Number	No. of Frames	RMSE	PSNR (dB)	FG Error ¹	BG Error ²	Overall Error ³	σ_{RMSE}^2
008a013s00L	178	0.074	22.677	3.818	0.286	7.407	3.930E-05
009a017s00L	169	0.067	23.487	3.252	0.260	7.048	2.321E-05
010a024s08L	187	0.068	23.441	3.782	0.233	7.269	2.451E-05
013a037s00L	114	0.070	23.144	4.409	0.274	9.310	5.052E-05
013a040s00L	184	0.072	22.924	4.437	0.271	8.702	5.501E-05
017a054s00L	188	0.071	23.029	3.176	0.312	7.523	3.230E-05
017a055s00R	162	0.069	23.217	2.300	0.357	7.505	3.132E-05
018a059s00L	188	0.065	23.785	2.634	0.266	6.445	2.091E-05
018a060s00L	179	0.070	23.182	3.286	0.283	7.204	2.340E-05
019a063s00L	186	0.065	23.784	1.865	0.326	6.801	2.092E-05
Average		0.069	23.267	3.296	0.287	7.521	3.214E-05

¹ FG Error Foreground Percentage Error ² BG Error Background Percentage Error

³ The percentage of the overall error compared to the motion pixels only

TABLE C.1: Combining the UC and the MOG classifiers for indoor motion sequences

Sequence Number	No. of Frames	RMSE	PSNR (dB)	FG Error ¹	BG Error ²	Overall Error ³	σ_{RMSE}^2
008a013s00L	178	0.073	22.793	3.883	0.266	7.220	4.262E-05
009a017s00L	169	0.069	23.194	2.622	0.337	7.537	2.272E-05
010a024s08L	187	0.071	23.023	3.252	0.317	7.989	1.997E-05
013a037s00L	114	0.063	23.995	3.808	0.214	7.646	3.681E-05
013a040s00L	184	0.075	22.561	3.473	0.381	9.490	6.665E-05
017a054s00L	188	0.078	22.189	2.028	0.509	9.127	3.284E-05
017a055s00R	162	0.066	23.592	1.423	0.376	6.912	3.332E-05
018a059s00L	188	0.064	23.949	2.701	0.245	6.203	1.932E-05
018a060s00L	179	0.073	22.735	2.804	0.373	7.982	2.221E-05
019a063s00L	186	0.063	23.979	2.053	0.295	6.510	2.230E-05
Average		0.070	23.201	2.805	0.331	7.661	3.188E-05

1 FG Error Foreground Percentage Error

2 BG Error Background Percentage Error

3 The percentage of the overall error compared to the motion pixels only

TABLE C.2: Combining the UC and the SBD classifiers for indoor motion sequences

Sequence Number	No. of Frames	RMSE	PSNR (dB)	FG Error ¹	BG Error ²	Overall Error ³	σ_{RMSE}^2
008a013s00L	178	0.075	22.498	3.309	0.354	7.732	4.768E-05
009a017s00L	169	0.068	23.463	5.316	0.127	7.157	5.130E-05
010a024s08L	187	0.071	22.986	6.996	0.076	8.123	5.162E-05
013a037s00L	114	0.063	24.099	3.285	0.233	7.458	3.111E-05
013a040s00L	184	0.070	23.142	6.172	0.135	8.296	5.882E-05
017a054s00L	188	0.071	23.055	5.739	0.126	7.487	3.432E-05
017a055s00R	162	0.065	23.808	4.187	0.163	6.574	2.979E-05
018a059s00L	188	0.068	23.356	4.444	0.187	7.117	2.582E-05
018a060s00L	179	0.067	23.535	3.573	0.222	6.647	2.103E-05
019a063s00L	186	0.064	23.906	4.850	0.123	6.705	5.939E-05
Average		0.068	23.385	4.787	0.174	7.330	4.109E-05

1 FG Error Foreground Percentage Error

2 BG Error Background Percentage Error

3 The percentage of the overall error compared to the motion pixels only

TABLE C.3: Combining the MOG and the SBD classifiers for indoor motion sequences

C.1.2 Outdoor motion sequences

Sequence Number	No. of Frames	RMSE	PSNR (dB)	FG Error ¹	BG Error ²	Overall Error ³	σ_{RMSE}^2
008e013s00L	100	0.201	13.968	21.191	1.366	30.051	2.056E-04
009e017s01L	96	0.211	13.525	9.207	3.780	35.826	1.359E-04
010e024s00L	94	0.222	13.092	18.898	2.660	35.172	1.932E-04
013e037s00L	158	0.153	16.307	12.249	0.933	18.815	6.633E-05
013e040s00L	151	0.228	12.844	10.349	4.513	44.232	2.249E-05
017e054s00L	112	0.214	13.422	12.273	3.304	32.604	2.022E-04
017e055s00R	88	0.206	13.732	27.163	0.918	33.499	1.366E-04
018e059s01L	104	0.159	16.005	8.170	1.708	19.937	1.195E-04
018e060s00L	88	0.186	14.606	18.419	1.111	25.454	1.690E-04
019e063s05L	112	0.154	16.299	18.308	0.743	25.539	1.492E-04
Average		0.193	14.380	15.623	2.104	30.113	1.400E-04

1 FG Error Foreground Percentage Error

2 BG Error Background Percentage Error

3 The percentage of the overall error compared to the motion pixels only

TABLE C.4: Combining the UC and the MOG classifiers for outdoor motion sequences

Sequence Number	No. of Frames	RMSE	PSNR (dB)	FG Error ¹	BG Error ²	Overall Error ³	σ_{RMSE}^2
008e013s00L	100	0.177	15.167	11.819	1.810	23.576	8.785E-04
009e017s01L	96	0.199	14.044	15.900	2.259	31.708	1.253E-04
010e024s00L	94	0.208	13.633	7.249	3.872	30.795	4.144E-05
013e037s00L	158	0.141	17.030	8.053	1.134	16.045	1.118E-04
013e040s00L	151	0.218	13.236	29.124	1.503	40.404	9.486E-06
017e054s00L	112	0.219	13.236	18.583	2.548	34.215	4.856E-04
017e055s00R	88	0.202	13.903	17.716	2.118	32.319	7.977E-05
018e059s01L	104	0.141	17.048	5.813	1.427	15.645	6.880E-05
018e060s00L	88	0.169	15.454	7.160	2.176	20.912	5.147E-05
019e063s05L	112	0.144	16.854	10.166	1.255	22.397	8.352E-05
Average		0.182	14.961	13.158	2.010	26.801	1.936E-04

1 FG Error Foreground Percentage Error

2 BG Error Background Percentage Error

3 The percentage of the overall error compared to the motion pixels only

TABLE C.5: Combining the UC and the SBD classifiers for outdoor motion sequences

Sequence Number	No. of Frames	RMSE	PSNR (dB)	FG Error ¹	BG Error ²	Overall Error ³	σ_{RMSE}^2
008e013s00L	100	0.178	15.115	12.401	1.761	23.854	8.764E-04
009e017s01L	96	0.200	13.988	15.260	2.411	32.129	1.338E-04
010e024s00L	94	0.201	13.948	10.651	2.944	28.665	2.106E-05
013e037s00L	158	0.148	16.670	10.818	0.951	17.560	2.451E-04
013e040s00L	151	0.215	13.372	29.946	1.228	39.162	6.440E-06
017e054s00L	112	0.220	13.199	19.193	2.504	34.561	5.645E-04
017e055s00R	88	0.216	13.309	26.814	1.477	37.080	6.504E-05
018e059s01L	104	0.145	16.808	7.937	1.243	16.502	3.771E-05
018e060s00L	88	0.173	15.287	12.859	1.417	21.808	2.005E-04
019e063s05L	112	0.143	16.925	13.234	0.909	22.046	7.168E-05
Average		0.184	14.862	15.911	1.685	27.337	2.222E-04

1 FG Error Foreground Percentage Error

2 BG Error Background Percentage Error

3 The percentage of the overall error compared to the motion pixels only

TABLE C.6: Combining the MOG and the SBD classifiers for outdoor motion sequences

C.2 Three Classifiers with a Non-weighted Combination

C.2.1 Indoor motion sequences

Sequence Number	No. of Frames	RMSE	PSNR (dB)	FG Error ¹	BG Error ²	Overall Error ³	σ_{RMSE}^2
008a013s00L	178	0.132	17.628	3.238	1.612	23.632	2.756E-05
009a017s00L	169	0.116	18.743	2.355	1.275	21.041	4.385E-05
010a024s08L	187	0.125	18.066	2.484	1.503	25.056	2.480E-05
013a037s00L	114	0.106	19.472	5.305	0.900	21.617	3.133E-05
013a040s00L	184	0.115	18.836	2.958	1.212	22.241	5.452E-05
017a054s00L	188	0.115	18.780	1.572	1.320	20.134	6.416E-05
017a055s00R	162	0.099	20.136	2.449	0.869	15.172	1.027E-05
018a059s00L	188	0.105	19.575	4.154	0.893	16.954	2.186E-05
018a060s00L	179	0.108	19.360	2.123	1.096	17.401	3.048E-05
019a063s00L	186	0.107	19.446	3.084	1.012	18.439	2.631E-05
Average		0.113	19.004	2.972	1.169	20.169	3.351E-05

1 FG Error Foreground Percentage Error

2 BG Error Background Percentage Error

3 The percentage of the overall error compared to the motion pixels only

TABLE C.7: Combining the UC, the SBD, and the MOG classifiers for indoor motion sequences using the Maximum combination rule

Sequence Number	No. of Frames	RMSE	PSNR (dB)	FG Error ¹	BG Error ²	Overall Error ³	σ_{RMSE}^2
008a013s00L	178	0.072	22.909	4.344	0.216	7.038	4.503E-05
009a017s00L	169	0.065	23.752	3.296	0.231	6.657	2.935E-05
010a024s08L	187	0.066	23.633	3.651	0.221	6.952	2.335E-05
013a037s00L	114	0.068	23.336	6.650	0.125	8.889	4.368E-05
013a040s00L	184	0.070	23.112	4.024	0.276	8.384	6.739E-05
017a054s00L	188	0.069	23.225	2.672	0.326	7.211	3.629E-05
017a055s00R	162	0.064	23.921	3.354	0.209	6.399	3.432E-05
018a059s00L	188	0.065	23.773	4.576	0.130	6.445	1.494E-05
018a060s00L	179	0.067	23.538	3.169	0.252	6.661	2.900E-05
019a063s00L	186	0.061	24.264	3.479	0.173	6.100	2.389E-05
Average		0.067	23.546	3.922	0.216	7.073	3.472E-05

1 FG Error Foreground Percentage Error

2 BG Error Background Percentage Error

3 The percentage of the overall error compared to the motion pixels only

TABLE C.8: Combining the UC, the SBD, and the MOG classifiers for indoor motion sequences using the Median combination rule

Sequence Number	No. of Frames	RMSE	PSNR (dB)	FG Error ¹	BG Error ²	Overall Error ³	σ_{RMSE}^2
008a013s00L	178	0.134	17.470	7.221	1.364	24.515	3.405E-05
009a017s00L	169	0.116	18.692	5.763	1.058	21.278	4.343E-05
010a024s08L	187	0.125	18.089	6.138	1.248	24.898	2.907E-05
013a037s00L	114	0.113	18.963	9.834	0.795	24.242	2.721E-05
013a040s00L	184	0.115	18.840	6.859	0.961	22.163	4.090E-05
017a054s00L	188	0.104	19.693	5.788	0.740	16.181	2.080E-05
017a055s00R	162	0.102	19.853	5.873	0.704	16.181	1.750E-05
018a059s00L	188	0.109	19.257	6.811	0.798	18.251	2.977E-05
018a060s00L	179	0.108	19.328	5.199	0.884	17.542	3.489E-05
019a063s00L	186	0.108	19.336	5.940	0.855	18.918	3.279E-05
Average		0.113	18.952	6.543	0.940	20.417	3.104E-05

1 FG Error Foreground Percentage Error

2 BG Error Background Percentage Error

3 The percentage of the overall error compared to the motion pixels only

TABLE C.9: Combining the UC, the SBD, and the MOG classifiers for indoor motion sequences using the Product combination rule

Sequence Number	No. of Frames	RMSE	PSNR (dB)	FG Error ¹	BG Error ²	Overall Error ³	σ_{RMSE}^2
008a013s00L	178	0.132	17.628	3.238	1.612	23.632	2.756E-05
009a017s00L	169	0.116	18.743	2.355	1.275	21.040	4.385E-05
010a024s08L	187	0.125	18.066	2.484	1.503	25.056	2.480E-05
013a037s00L	114	0.106	19.472	5.304	0.900	21.617	3.133E-05
013a040s00L	184	0.115	18.836	2.958	1.212	22.241	5.452E-05
017a054s00L	188	0.115	18.780	1.572	1.320	20.134	6.416E-05
017a055s00R	162	0.099	20.136	2.449	0.869	15.172	1.026E-05
018a059s00L	188	0.105	19.575	4.154	0.893	16.954	2.186E-05
018a060s00L	179	0.108	19.360	2.123	1.096	17.401	3.048E-05
019a063s00L	186	0.107	19.446	3.084	1.012	18.439	2.631E-05
Average		0.113	19.004	2.972	1.169	20.168	3.351E-05

1 FG Error Foreground Percentage Error

2 BG Error Background Percentage Error

3 The percentage of the overall error compared to the motion pixels only

TABLE C.10: Combining the UC, the SBD, and the MOG classifiers for indoor motion sequences using the Sum combination rule

C.2.2 Outdoor motion sequences

Sequence Number	No. of Frames	RMSE	PSNR (dB)	FG Error ¹	BG Error ²	Overall Error ³	σ_{RMSE}^2
008e013s00L	100	0.198	14.129	13.535	2.435	29.351	5.840E-04
009e017s01L	96	0.234	12.626	23.555	2.892	43.968	8.674E-05
010e024s00L	94	0.212	13.496	10.171	3.543	31.797	2.396E-05
013e037s00L	158	0.184	14.753	12.018	2.184	27.353	4.772E-04
013e040s00L	151	0.250	12.051	28.538	3.274	53.126	3.965E-05
017e054s00L	112	0.233	12.684	19.809	3.066	38.670	3.675E-04
017e055s00R	88	0.234	12.628	12.267	4.516	43.566	2.431E-04
018e059s01L	104	0.159	15.988	10.814	1.331	19.986	1.034E-04
018e060s00L	88	0.173	15.269	9.053	2.021	21.825	9.887E-05
019e063s05L	112	0.155	16.197	12.211	1.414	26.047	6.373E-05
Average		0.203	13.982	15.197	2.668	33.569	2.088E-04

1 FG Error Foreground Percentage Error

2 BG Error Background Percentage Error

3 The percentage of the overall error compared to the motion pixels only

TABLE C.11: Combining the UC, the SBD, and the MOG classifiers for outdoor motion sequences using the Maximum combination rule

Sequence Number	No. of Frames	RMSE	PSNR (dB)	FG Error ¹	BG Error ²	Overall Error ³	σ_{RMSE}^2
008e013s00L	100	0.178	15.077	13.570	1.602	23.967	8.364E-04
009e017s01L	96	0.208	13.658	22.434	1.750	34.677	1.666E-04
010e024s00L	94	0.203	13.867	9.886	3.158	29.214	2.502E-05
013e037s00L	158	0.144	16.910	11.312	0.750	16.627	2.578E-04
013e040s00L	151	0.215	13.369	29.699	1.265	39.191	8.182E-06
017e054s00L	112	0.218	13.262	19.846	2.305	34.012	4.992E-04
017e055s00R	88	0.192	14.328	20.438	1.278	29.293	3.881E-05
018e059s01L	104	0.149	16.547	9.628	1.153	17.566	8.971E-05
018e060s00L	88	0.165	15.666	10.070	1.560	19.930	1.016E-04
019e063s05L	112	0.142	16.977	12.332	0.977	21.801	1.121E-04
Average		0.181	14.966	15.921	1.580	26.628	2.135E-04

1 FG Error Foreground Percentage Error

2 BG Error Background Percentage Error

3 The percentage of the overall error compared to the motion pixels only

TABLE C.12: Combining the UC, the SBD, and the MOG classifiers for outdoor motion sequences using the Median combination rule

Sequence Number	No. of Frames	RMSE	PSNR (dB)	FG Error ¹	BG Error ²	Overall Error ³	σ_{RMSE}^2
008e013s00L	100	0.198	14.134	13.739	2.399	29.320	5.958E-04
009e017s01L	96	0.235	12.582	24.348	2.841	44.402	7.291E-05
010e024s00L	94	0.214	13.394	11.929	3.385	32.564	5.898E-05
013e037s00L	158	0.189	14.504	14.071	2.123	28.971	4.751E-04
013e040s00L	151	0.250	12.059	29.020	3.198	53.035	4.617E-05
017e054s00L	112	0.234	12.647	20.783	2.964	39.025	4.169E-04
017e055s00R	88	0.238	12.499	13.718	4.481	44.779	1.985E-04
018e059s01L	104	0.163	15.785	12.133	1.289	21.012	1.701E-04
018e060s00L	88	0.179	14.987	11.086	1.941	23.355	1.915E-04
019e063s05L	112	0.157	16.089	13.118	1.388	26.703	7.882E-05
Average		0.206	13.868	16.395	2.601	34.317	2.305E-04

1 FG Error Foreground Percentage Error

2 BG Error Background Percentage Error

3 The percentage of the overall error compared to the motion pixels only

TABLE C.13: Combining the UC, the SBD, and the MOG classifiers for outdoor motion sequences using the Product combination rule

Sequence Number	No. of Frames	RMSE	PSNR (dB)	FG Error ¹	BG Error ²	Overall Error ³	σ_{RMSE}^2
008e013s00L	100	0.198	14.129	13.535	2.435	29.351	5.840E-04
009e017s01L	96	0.234	12.626	23.555	2.892	43.968	8.674E-05
010e024s00L	94	0.211	13.500	10.046	3.559	31.782	2.628E-05
013e037s00L	158	0.184	14.754	11.997	2.183	27.330	4.658E-04
013e040s00L	151	0.250	12.052	28.535	3.273	53.113	3.980E-05
017e054s00L	112	0.233	12.696	19.725	3.061	38.557	3.547E-04
017e055s00R	88	0.234	12.630	12.254	4.516	43.554	2.445E-04
018e059s01L	104	0.159	15.987	10.859	1.326	19.996	1.088E-04
018e060s00L	88	0.173	15.262	9.094	2.021	21.863	1.022E-04
019e063s05L	112	0.155	16.216	12.100	1.413	25.931	6.124E-05
Average		0.203	13.985	15.170	2.668	33.545	2.074E-04

1 FG Error Foreground Percentage Error

2 BG Error Background Percentage Error

3 The percentage of the overall error compared to the motion pixels only

TABLE C.14: Combining the UC, the SBD, and the MOG classifiers for outdoor motion sequences using the Sum combination rule

C.3 Three Classifiers with a Weighted Combination

C.3.1 Indoor motion sequences

Sequence Number	No. of Frames	RMSE	PSNR (dB)	FG Error ¹	BG Error ²	Overall Error ³	σ_{RMSE}^2
008a013s00L	178	0.078	22.217	3.228	0.401	8.248	4.832E-05
009a017s00L	169	0.071	23.002	2.338	0.385	7.954	4.603E-05
010a024s08L	187	0.075	22.574	2.473	0.428	8.884	2.348E-05
013a037s00L	114	0.069	23.249	5.287	0.209	9.043	3.063E-05
013a040s00L	184	0.077	22.328	2.933	0.450	10.064	7.576E-05
017a054s00L	188	0.090	20.990	1.556	0.758	12.194	7.711E-05
017a055s00R	162	0.068	23.446	2.429	0.323	7.153	3.499E-05
018a059s00L	188	0.068	23.384	4.124	0.204	7.057	1.631E-05
018a060s00L	179	0.074	22.687	2.111	0.432	8.105	3.103E-05
019a063s00L	186	0.067	23.493	3.063	0.279	7.283	2.477E-05
Average		0.074	22.737	2.954	0.387	8.598	4.084E-05

1 FG Error Foreground Percentage Error

2 BG Error Background Percentage Error

3 The percentage of the overall error compared to the motion pixels only

TABLE C.15: Combining the UC, the SBD, and the MOG classifiers for indoor motion sequences using a weighted Maximum combination rule

Sequence Number	No. of Frames	RMSE	PSNR (dB)	FG Error ¹	BG Error ²	Overall Error ³	σ_{RMSE}^2
008a013s00L	178	0.074	22.616	3.234	0.345	7.537	5.172E-05
009a017s00L	169	0.068	23.375	2.345	0.340	7.311	4.760E-05
010a024s08L	187	0.070	23.150	2.480	0.355	7.790	2.799E-05
013a037s00L	114	0.067	23.559	5.296	0.175	8.431	3.456E-05
013a040s00L	184	0.073	22.781	2.946	0.389	9.102	8.223E-05
017a054s00L	188	0.087	21.222	1.561	0.713	11.567	7.678E-05
017a055s00R	162	0.064	23.864	2.438	0.279	6.521	4.149E-05
018a059s00L	188	0.065	23.736	4.135	0.166	6.512	1.706E-05
018a060s00L	179	0.069	23.272	2.117	0.359	7.097	3.418E-05
019a063s00L	186	0.064	23.912	3.074	0.235	6.626	2.832E-05
Average		0.070	23.149	2.963	0.336	7.849	4.419E-05

1 FG Error Foreground Percentage Error

2 BG Error Background Percentage Error

3 The percentage of the overall error compared to the motion pixels only

TABLE C.16: Combining the UC, the SBD, and the MOG classifiers for indoor motion sequences using a weighted Median combination rule

Sequence Number	No. of Frames	RMSE	PSNR (dB)	FG Error ¹	BG Error ²	Overall Error ³	σ_{RMSE}^2
008a013s00L	178	0.132	17.628	3.238	1.612	23.632	2.756E-05
009a017s00L	169	0.116	18.743	2.354	1.275	21.040	4.385E-05
010a024s08L	187	0.125	18.066	2.484	1.503	25.056	2.480E-05
013a037s00L	114	0.106	19.472	5.304	0.900	21.617	3.133E-05
013a040s00L	184	0.115	18.836	2.958	1.212	22.241	5.452E-05
017a054s00L	188	0.115	18.780	1.572	1.320	20.134	6.415E-05
017a055s00R	162	0.099	20.136	2.449	0.869	15.172	1.026E-05
018a059s00L	188	0.105	19.575	4.154	0.893	16.954	2.186E-05
018a060s00L	179	0.108	19.360	2.123	1.096	17.401	3.047E-05
019a063s00L	186	0.107	19.446	3.084	1.012	18.439	2.631E-05
Average		0.113	19.004	2.972	1.169	20.168	3.351E-05

1 FG Error Foreground Percentage Error

2 BG Error Background Percentage Error

3 The percentage of the overall error compared to the motion pixels only

TABLE C.17: Combining the UC, the SBD, and the MOG classifiers for indoor motion sequences using a weighted Product combination rule

Sequence Number	No. of Frames	RMSE	PSNR (dB)	FG Error ¹	BG Error ²	Overall Error ³	σ_{RMSE}^2
008a013s00L	178	0.088	21.167	3.238	0.576	10.472	3.966E-05
009a017s00L	169	0.079	22.043	2.352	0.515	9.880	4.046E-05
010a024s08L	187	0.086	21.368	2.484	0.615	11.707	1.857E-05
013a037s00L	114	0.076	22.432	5.304	0.310	10.902	2.625E-05
013a040s00L	184	0.086	21.366	2.956	0.601	12.492	6.522E-05
017a054s00L	188	0.095	20.460	1.570	0.867	13.752	7.584E-05
017a055s00R	162	0.076	22.461	2.447	0.442	8.925	2.413E-05
018a059s00L	188	0.076	22.457	4.151	0.319	8.730	1.638E-05
018a060s00L	179	0.084	21.503	2.122	0.611	10.628	2.703E-05
019a063s00L	186	0.075	22.534	3.083	0.395	9.064	2.131E-05
Average		0.082	21.779	2.971	0.525	10.655	3.548E-05

1 FG Error Foreground Percentage Error

2 BG Error Background Percentage Error

3 The percentage of the overall error compared to the motion pixels only

TABLE C.18: Combining the UC, the SBD, and the MOG classifiers for indoor motion sequences using a weighted Sum combination rule

C.3.2 Outdoor motion sequences

Sequence Number	No. of Frames	RMSE	PSNR (dB)	FG Error ¹	BG Error ²	Overall Error ³	σ_{RMSE}^2
008e013s00L	100	0.189	14.557	13.532	2.033	26.745	6.763E-04
009e017s01L	96	0.226	12.908	23.017	2.578	41.229	9.572E-05
010e024s00L	94	0.206	13.711	9.753	3.352	30.258	2.276E-05
013e037s00L	158	0.166	15.636	11.657	1.498	22.241	2.974E-04
013e040s00L	151	0.231	12.743	28.869	2.185	45.265	1.198E-05
017e054s00L	112	0.227	12.928	19.512	2.784	36.619	4.189E-04
017e055s00R	88	0.203	13.869	17.983	2.105	32.629	7.477E-05
018e059s01L	104	0.156	16.179	10.477	1.255	19.128	9.993E-05
018e060s00L	88	0.167	15.543	9.519	1.743	20.522	1.297E-04
019e063s05L	112	0.148	16.606	11.963	1.205	23.715	6.816E-05
Average		0.192	14.468	15.628	2.074	29.835	1.896E-04

1 FG Error Foreground Percentage Error

2 BG Error Background Percentage Error

3 The percentage of the overall error compared to the motion pixels only

TABLE C.19: Combining the UC, the SBD, and the MOG classifiers for outdoor motion sequences using a weighted Maximum combination rule

Sequence Number	No. of Frames	RMSE	PSNR (dB)	FG Error ¹	BG Error ²	Overall Error ³	σ_{RMSE}^2
008e013s00L	100	0.179	15.044	13.555	1.629	24.132	8.245E-04
009e017s01L	96	0.210	13.589	22.530	1.812	35.223	1.558E-04
010e024s00L	94	0.203	13.845	9.896	3.181	29.358	2.359E-05
013e037s00L	158	0.145	16.827	11.369	0.786	16.939	2.582E-04
013e040s00L	151	0.216	13.331	29.600	1.324	39.529	7.439E-06
017e054s00L	112	0.219	13.244	19.829	2.333	34.162	5.079E-04
017e055s00R	88	0.192	14.324	20.252	1.309	29.321	3.628E-05
018e059s01L	104	0.150	16.496	9.723	1.169	17.773	8.862E-05
018e060s00L	88	0.165	15.651	10.010	1.582	20.003	1.073E-04
019e063s05L	112	0.143	16.928	12.283	1.006	22.043	1.073E-04
Average		0.182	14.928	15.905	1.613	26.848	2.117E-04

¹ FG Error Foreground Percentage Error² BG Error Background Percentage Error³ The percentage of the overall error compared to the motion pixels only

TABLE C.20: Combining the UC, the SBD, and the MOG classifiers for outdoor motion sequences using a weighted Median combination rule

Sequence Number	No. of Frames	RMSE	PSNR (dB)	FG Error ¹	BG Error ²	Overall Error ³	σ_{RMSE}^2
008e013s00L	100	0.198	14.129	13.535	2.435	29.351	5.840E-04
009e017s01L	96	0.234	12.626	23.555	2.892	43.968	8.674E-05
010e024s00L	94	0.212	13.496	10.171	3.543	31.797	2.396E-05
013e037s00L	158	0.184	14.753	12.018	2.184	27.353	4.772E-04
013e040s00L	151	0.250	12.051	28.538	3.274	53.126	3.965E-05
017e054s00L	112	0.233	12.684	19.809	3.066	38.670	3.675E-04
017e055s00R	88	0.234	12.628	12.267	4.516	43.566	2.431E-04
018e059s01L	104	0.159	15.988	10.814	1.331	19.986	1.034E-04
018e060s00L	88	0.173	15.269	9.053	2.021	21.825	9.887E-05
019e063s05L	112	0.155	16.197	12.211	1.414	26.047	6.373E-05
Average		0.203	13.982	15.197	2.668	33.569	2.088E-04

¹ FG Error Foreground Percentage Error² BG Error Background Percentage Error³ The percentage of the overall error compared to the motion pixels only

TABLE C.21: Combining the UC, the SBD, and the MOG classifiers for outdoor motion sequences using a weighted Product combination rule

Sequence Number	No. of Frames	RMSE	PSNR (dB)	FG Error ¹	BG Error ²	Overall Error ³	σ_{RMSE}^2
008e013s00L	100	0.194	14.282	13.540	2.279	28.363	5.820E-04
009e017s01L	96	0.231	12.753	23.271	2.754	42.720	9.798E-05
010e024s00L	94	0.208	13.638	9.757	3.438	30.773	2.101E-05
013e037s00L	158	0.174	15.227	11.685	1.822	24.504	3.982E-04
013e040s00L	151	0.240	12.397	28.691	2.710	49.030	2.554E-05
017e054s00L	112	0.230	12.807	19.582	2.934	37.616	3.853E-04
017e055s00R	88	0.226	12.930	14.034	3.834	40.592	1.991E-04
018e059s01L	104	0.158	16.058	10.613	1.313	19.661	9.995E-05
018e060s00L	88	0.169	15.455	9.289	1.845	20.937	1.249E-04
019e063s05L	112	0.151	16.417	11.950	1.310	24.765	5.434E-05
Average		0.198	14.196	15.241	2.424	31.896	1.988E-04

¹ FG Error Foreground Percentage Error

² BG Error Background Percentage Error

³ The percentage of the overall error compared to the motion pixels only

TABLE C.22: Combining the UC, the SBD, and the MOG classifiers for outdoor motion sequences using a weighted Sum combination rule

References

Al-Mazeed, A., M. Nixon, and S. Gunn (2003). Fusing complementary operators to enhance foreground/background segmentation. In *the British Machine Vision Conference (BMVC'2003)*, pp. 501–510.

Al-Mazeed, A. H., M. S. Nixon, and S. R. Gunn (2004). Classifiers combination for improved motion segmentation. In A. Campilho and M. Kamel (Eds.), *Proceedings of International Conference on Image Analysis and Recognition, Lecture Notes in Computer Science*, Volume LNCS 3212, pp. 363–371. Springer.

Alexandre, L., A. Campilho, and M. Kamel (2001). On combining classifiers using sum and product rules. *Pattern Recognition Letters* 22(12), 1283–1289.

Alkoot, F. and J. Kittler (1999). Experimental evaluation of expert fusion strategies. *Pattern Recognition Letters* 20, 1361–1369.

Amari, S. and S. Wu (1999). Improving support vector machine classifiers by modifying kernel functions. *Neural Networks* 12(6), 783–789.

Andreadis, I., M. Browne, and J. Swift (1990). Image pixel classification by chromaticity analysis. *Pattern Recognition Letters* 11(1), 51–58.

Arseneau, S. and J. Cooperstock (1999). Real-time image segmentation for action recognition. In *Proceedings of the IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM'99)*, pp. 86–89.

Bazin, A., L. Middleton, and M. S. Nixon (2005). Probabilistic fusion of gait features for biometric verification. In *Proc. of Eighth International Conference of Information Fusion*.

Bishop, C. (1996). *Neural Networks for Pattern Recognition*. Oxford University Press.

Bobick, A. and A. Johnson (2001). Gait recognition using static, activity-specific parameters. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Volume 1, pp. 423–430.

Boinovic, N. and J. Konrad (2005). Motion analysis in 3d dct domain and its application to video coding. *Signal Processing: Image Communication* 20(6), 510–528.

Bowden, R. (1999). *Learning Nonlinear Models of Shape and Motion*. Thesis, Brunel University, Dep. of Systems Eng.

Buhmann, M. D. (2000). Radial basis functions. *Acta Numerica* 9, 1–38.

Carron, T. and P. Lambert (1994). Color edge detector using jointly hue, saturation and intensity. In *IEEE International Conference on Image Processing*, Volume 3, pp. 977–981.

Chang, I. and C. Huang (2000). The model-based human body motion analysis system. *Image and Vision Computing* 18(14), 1067–1083.

Chang, K., K. Bowyer, S. Sarkar, and B. Victor (2003). Comparison and combination of ear and face images in appearance-based biometrics. *IEEE Trans. Pattern Anal. and Mach. Intel.* 25(9), 1160–1165.

Chen, K., L. Wang, and H. Chi (1997). Methods of combining multiple classifiers with different features and their applications to text-independent speaker identification. *International Journal of Pattern Recognition and Artificial Intelligence* 11(3), 417–445.

Chen, Y., X. Zhou, and T. Huang (2001). One-class SVM for learning in image retrieval. In *Proc. IEEE Inter. Cofn. on Image Processing*.

Cheng, H., X. Jiang, Y. Sun, and J. Wang (2001). Color image segmentation: advances and prospects. *Pattern Recognition* 34(12), 2259–2281.

Collins, R., A. Lipton, H. Fujiyoshi, and T. Kanade (2001, October). Algorithms for cooperative multisensor surveillance. *Proceedings of the IEEE* 89(10), 1456–1477.

Cucchiara, R., C. Grana, M. Piccardi, and A. Prati (2000). Statistic and knowledge-based moving object detection in traffic scenes. In *Proceedings of the 3rd IEEE Intelligent Transportation System Conference (ITSC'2000)*, pp. 27–32.

Cucchiara, R., M. M. Piccardi, and P. Mello (2000). Image analysis and rule-based reasoning for a traffic monitoring system. *IEEE Transactions on Intelligent Transportation Systems* 1(2), 119–130.

Cutler, R. and L. Davis (1998). View-based detection and analysis of periodic motion. In *the 14th International Conference on Pattern Recognition*, pp. 495–500.

Cwik, J. Koronacki, J. (1996). Probability density estimation using a Gaussian clustering algorithm. *Neural Computing and Applications* 4(3), 149–160.

Czyz, J., J. Kittler, and L. Vandendorpe (2004). Multiple classifier combination for face-based identity verification. *Pattern Recognition* 37(7), 1459–1469.

Dagless, E., A. Ali, and J. Cruz (1993). Visual road traffic monitoring and data collection. In *Proceedings of the IEEE-IEE Vehicle Navigation and Information Systems Conference (VNIS'93)*, pp. 146–149.

Davis, J. and A. Bobick (1997). The representation and recognition of action using temporal templates. *Proc. of IEEE CVPR* 6(2), 928–934.

Duin, R. (2002). The combining classifier: to train or not to train? In *Proc. IEEE International Conference on Pattern Recognition*, Volume 2, pp. 765–770.

Eckert, M. P. and A. P. Bradley (1998). Perceptual quality metrics applied to still image compression. *Signal Processing* 70, 177–200.

Elgammal, A., R. Duraiswami, D. Harwood, and L. S. Davis (2002). Background and foreground modeling using non-parametric kernel density estimation for visual surveillance. *Proceedings of the IEEE* 90(7), 1151–1163.

Elgammal, A., D. Harwood, and L. S. Davis (2000). Nonparametric background model for background subtraction. In *Proceedings of the 6th Eur. Conference on Computer Vision*, Volume 2, pp. 751–767.

Eskicioglu, A. M. and P. S. Fisher (1995). Image quality measures and their performance. *IEEE Trans. Communications* 43, 2959–2965.

Etemadnia, H. and M. Alsharif (2003). Automatic image shadow identification using lpf in homomorphic processing system. In *Proceedings of the VIIth Digital Image Computing Techniques and Applications (DICTA '2003)*.

Eveland, C., K. Konolige, and R. Bolles (1998). Background modeling for segmentation of video-rate stereo sequences. In *Proceedings of the IEEE Conference on Vision and Pattern Recognition (CVPR'98)*, pp. 266–271.

Friedman, N. and S. Russell (1997). Image segmentation in video sequences: a probabilistic approach. In *Proceedings of the 13th Conference on Uncertainty in Artificial Intelligence*, pp. 175–181.

Gauch, J. and C. Hsia (1992). A comparison of three-color image segmentation algorithms in four color spaces. *SPIE, Visual Communications and Image Processing* 1818, 1168–1181.

Gevers, T. and A. Smeulders (1996). A comparative study of several color models for color image invariant retrieval. In *International Workshop on Image Databases and Multi-Media Search*, pp. 17–26.

Gevers, T. and A. Smeulders (1999). Color based object recognition. *Pattern Recognition* 32(3), 453–464.

Gevers, T. and A. Smeulders (2000). Pictoseek: Combining color and shape invariant features for image retrieval. *IEEE Transactions on Image Processing* 9(1), 102–119.

Golland, P. and A. Bruckstein (1996). Why RGB? or how to design color displays for martians. *Graphical Models and Image Processing* 58(5), 405–412.

Gonzalez, R. and R. Woods (1992). *Digital Image Processing*. Addison-Wesley.

Grant, M., M. Nixon, and P. Lewis (2002). Extracting moving shapes by evidence gathering. *Pattern Recognition* 35(5), 1099–1114.

Grimson, W., C. Stauffer, R. Romano, and L. Lee (1998). Using adaptive tracking to classify and monitor activities in a site. In *Proceedings of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'98)*, pp. 22–29.

Grzymala-Busse, J., L. Goodwin, and X. Zhang (2003). Increasing sensitivity of preterm birth by changing rule strengths. *Pattern Recognition Letters* 24(6), 903–910.

Grzymala-Busse, J., W. Grzymala-Busse, and L. Goodwin (2002). A comparison of three closest fit approaches to missing attribute values in preterm birth data. *International Journal of Intelligent Systems* 17(2), 125–134.

Hayfron-Acquah, J., M. Nixon, and J. Carter (2001). Automatic gait recognition by symmetry analysis. In *Proceedings of the International Conference on Audio-and-Video-Based Biometric Person Authentication*, pp. 272–277.

Ho, T., J. Hull, and S. Srihari (1994). Decision combination in multiple classifier systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 16(1), 66–75.

Horn, B. (1986). *Robot Vision*. MIT Press.

Horprasert, T., D. Harwood, and L. Davis (1999). A statistical approach for real-time robust background subtraction and shadow detection. In *Proceedings of IEEE ICCV'99*, pp. 1–19.

Horprasert, T., D. Harwood, and L. Davis (2000). A robust background subtraction and shadow detection. In *Proceedings of the Asian Conference on Computer Vision ACCV'2000*, Taipie, Taiwan.

Huang, Y. and C. Suen (1995). Method of combining multiple experts for the recognition of unconstrained handwritten numerals. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 17(1), 90–94.

Javed, O., K. Shafique, and M. Shah (2002). A hierarchical approach to robust background subtraction using color and gradient information. In *IEEE Workshop on Motion and Video Computing*, pp. 22–27.

Javed, O. and M. Shah (2002). Tracking and object classification for automated surveillance. In *the seventh European Conference on Computer Vision (ECCV'2002)*, Volume 4, pp. 343–357.

Jing, X. and D. Zhang (2003). Face recognition based on linear classifiers combination. *Neurocomputing* 50, 485–488.

Jing, X., C. Zhu, and L. Chau (2003). Smooth constrained motion estimation for video coding. *Signal Processing* 83(3), 677–680.

Johnson, R. and D. Wichern (2002). *Applied Multivariate Statistical Analysis* (5 ed.). Prentice Hall.

Jones, D. and J. Malik (1992). Computational framework for determining stereo correspondence from a set of linear spatial filters. *Image and Vision Computing* 10(10), 699–708.

KaewTraKulPong, P. and R. Bowden (2001a). An adaptive visual system for tracking low resolution colour targets. In *Proc. of BMVC'01*, Volume 1, pp. 243–252.

KaewTraKulPong, P. and R. Bowden (2001b). An improved adaptive background mixture model for real-time tracking with shadow detection. In *Proc. 2nd European Workshop on Advanced Video Based Surveillance Systems*. Kluwer Academic Publishers.

KaewTrakulPong, P. and R. Bowden (2003). A real time adaptive visual surveillance system for tracking low-resolution colour targets in dynamically changing scenes. *Image and Vision Computing* 21(10), 913–929.

Kale, A., A. Chowdhury, and R. Chellappa (2004). Fusion of gait and face for human recognition. In *Inter. Conf. on Acoustics, Speech, and Signal Processing*.

Kameda, Y. and M. Minoh (1996). A human motion estimation method using 3-successive video frames. In *the International Conference on Virtual Systems and Multimedia*.

Kender, J. (1976). Saturation, hue and normalized color: Calculation, digitization effects, and use. Technical report, Carnegie-Mellon University.

Kilger, M. (1992). A shadow handler in a video-based real-time traffic monitoring system. In *Proceedings of IEEE Workshop on Applications of Computer Vision*, pp. 11–18.

Kim, J. and H. Kim (2003). Efficient region-based motion segmentation for a video monitoring system. *Pattern Recognition Letters* 24(1), 113–128.

Kim, K., C. Lee, E. Lee, and Y. Ha (1996). Color image quantization using weighted distortion measure of HVS color activity. In *IEEE International Conference on Image Processing*, Volume 3, pp. 1035–1039.

Kim, W. and R. Park (1996). Color image palette construction based on the HSI color system for minimizing the reconstruction error. In *IEEE International Conference on Image Processing*, Volume 3, pp. 1041–1044.

Kittler, J., M. Hatef, R. Duin, and J. Matas (1998). On combining classifiers. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 20(3), 226–239.

Kittler, J. and Hojjatoleslami (1998). A weighted combination of classifiers employing shared and distinct representations. In *CVPR*, pp. 924–29.

Kittler, J., A. Hojjatoleslami, and T. Windeatt (1997). Weighting factors in multiple expert fusion. In *Proc. British Machine Vision Conference*, pp. 41–50.

Kuncheva, L. (2002). A theoretical study on six fusion strategies. *IEEE Trans. Pattern Recognition Mach. Intell.* 24(2), 281–286.

Kuncheva, L., J. Bezdek, and R. Duin (2001). Decision templates for multiple classifier fusion: an experimental comparison. *Pattern Recognition* 34(2), 299–314.

Lai, Y. K. and C.-C. J. Kuo (2000). A haar wavelet approach to compressed image quality measurement. *Journal of Visual Communication and Image Understanding* 11, 17–40.

Leandro Rodrguez-Liales, L., C. Garca-Mateob, and J. Alba-Castrob (2003). On combining classifiers for speaker authentication. *Pattern Recognition* 36(2), 347–359.

Lee, J. (1998). Combining the evidence of different relevance feedback methods for information retrieval. *Information Processing & Management* 34, 681–691.

Lipton, A., H. Fujiyoshi, and P. R. (1998). Moving target classification and tracking from real-time video. In *Proc. of DARPA Image Understanding Workshop (IUIW'98)*.

Littmann, E. and H. Ritter (1997). Adaptive color segmentation - a comparison of neural and statistical methods. *IEEE Trans. Neural Networks* 8(1), 175–185.

Maes, P., T. Darrell, B. Blumberg, and A. Pentland (1997). The alive system: wireless, full-body interaction with autonomous agents. *ACM Multimedia Systems* 5(2), 105–112.

Manevitz, L. M. and M. Yousif (2001). One-class svms for document classification. *Journal of Machine Learning Research* 2, 139–154.

McKenna, S., S. Jabri, Z. Duric, A. Rosenfeld, and H. Wechsler (2000). Tracking groups of people. *Computer Vision and Image Understanding* 80(1), 42–56.

McKenna, S., Y. Raja, and G. S. (1998). Object tracking using adaptive colour mixture models. In *Proc. of the Third Asian Conference on Computer Vision (ACCV'98)*, Volume 1, pp. 615–622. Springer-Verlag.

McKenna, S., Y. Raja, and G. S. (1999). Tracking colour objects using adaptive mixture models. *Image and Vision Computing* 17(3-4), 225–231.

Michalopoulos, P. (1991). Vehicle detection video through image processing: the autoscope system. *IEEE Transactions on Vehicular Technology* 40(1), 21–29.

Mikic, I., M. Trivedi, E. . Hunter, and P. Cosman (2003). Human body model acquisition and tracking using voxel data. *International Journal of Computer Vision archive* 53(3), 199–223.

Mikic, I., M. Trivedi, E. Hunter, and P. Cosman (2001). Articulated body posture estimation from multi-camera voxel data. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Volume 1, pp. 455–460.

Mittal, A. and D. Huttenlocher (2000). Scene modeling for wide area surveillance and image synthesis. In *CVPR*, pp. 2160–2167.

Moeslund, T. (2000). Interacting with a virtual world through motion capture. In *Interaction in Virtual Inhabited 3D Worlds*. Springer-Verlag.

Nottelmann, H. and U. Straccia (2005). Information retrieval and machine learning for probabilistic schema matching. In *Proc. of the 14th International Conference on Information and Knowledge Management*.

Ohya, J., J. Kurumisawa, R. Nakatsu, K. Ebihara, S. Iwasawa, D. Harwood, and T. Horprasert (1999). Virtual metamorphosis. *IEEE Multi-media* 6(2), 29–39.

Park, D., J. Park, T. Kim, and J. Han (1999). Image indexing using weighted color histogram. In *Proceedings. International Conference on Image Analysis and Processing*, pp. 909–914.

Patias, P. (2002). Medical imaging challenges photogrammetry. *ISPRS Journal of Photogrammetry and Remote Sensing* 56(5–6), 295–310.

Phillips, P., S. Sarkar, I. Robledo, P. Grother, and K. Bowyer (2002). The gait identification challenge problem: data sets and baseline algorithm. In *Proc. IEEE International Conference on Pattern Recognition*, Volume 3644, pp. 1:385–388.

Pietikainen, M., S. Nieminen, E. Marszalec, and T. Ojala (1996). Accurate color discrimination with classification based on feature distributions. In *Proceedings of the 13th International Conference on Pattern Recognition*, Volume 3, pp. 833–838.

Platt, J. (1999). Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In A. Smola, P. Bartlett, B. Schölkopf, and D. Schuurmans (Eds.), *Advances in Large Margin Classifiers*. MIT press.

Polat, E., M. Yeasin, and R. Sharma (2003). Robust tracking of human body parts for collaborative human computer interaction. *Computer Vision and Image Understanding* 89(1), 44–69.

Prati, A., I. Mikic, M. Trivedi, and R. Cucchiara (2003). Detecting moving shadows: algorithms and evaluation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25(7), 918–923.

Rahman, A. and M. Fairhurst (1997). A new hybrid approach in combining multiple experts to recognise handwritten numerals. *Pattern Recognition Letters* 18, 781–790.

Rajan, V., S. Subramanian, D. Keenan, D. Johnson, A. Sandin, and T. DeFanti (2002). A realistic video avatar system for networked virtual environments. In *Proc. International Immersive Projection Technology*.

Roberts, S., D. Husmeier, I. Rezek, and W. Penny (1998). Bayesian approaches to gaussian mixture modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20(11), 1133–1142.

Rosin, P. (2002). Thresholding for change detection. *Computer Vision and Image Understanding* 86(2), 79–95.

Rui, Y., A. She, and T. Huang (1996). Automated region segmentation using attraction-based grouping in spatial-color-texture space. In *IEEE International Conference on Image Processing*, Volume 1, pp. 53–56.

Ruta, D. and B. Gabrys (2005). Classifier selection for majority voting. *Information Fusion* 6(1), 63–81.

Salvador, E., A. Cavallaro, and T. Ebrahimi (2001). Shadow identification and classification using invariant color models. In *In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '01)*, Volume 3, pp. 1545–1548.

Salvador, E., A. Cavallaro, and T. Ebrahimi (2004). Cast shadow segmentation using invariant colour features. *Computer Vision and Image Understanding* 95(2), 238–259.

Schölkopf, B., J. Platt, J. Shawe-Taylor, J. Smola, and R. Williamson (2001). Estimating the support of a high-dimensional distribution. *Neural Computation* 13(7), 1443–1472.

Schwarz, M., W. Cowan, and J. Beatty (1987). An experimental comparison of RGB, YIQ, LAB, HSV, and opponent color models. *ACM Transactions on Graphics* 6(2), 123–158.

Sebe, N. and M. Lew (2001). Color-based retrieval. *Pattern Recognition Letters* 22(2), 223–230.

Senior, A., A. Hampapur, Y.-L. Tian, L. Brown, S. Pankanti, and R. Bolle (2001). Appearance models for occlusion handling. In *Second International workshop on Performance Evaluation of Tracking and Surveillance systems*.

Shakhnarovich, G. and T. Darrell (2002). On probabilistic combination of face and gait cues for identification. In *Proc. of the Int. Conference on Automatic Face and Gesture Recognition*.

Shakhnarovich, G., L. Lee, and T. Darrell (2001). Integrated face and gait recognition with multiple views. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Volume 1, pp. 439–446.

Sharkey, A. and N. Sharkey (1997). Combining diverse neural nets. *The Knowledge Engineering Review* 12(3), 231–247.

Shutler, J., M. Grant, M. Nixon, and J. Carter (2002). On a large sequence-based human gait database. In *Proc. of RASC*, pp. 66–71.

Smith, A. (1978). Color gamut transform pairs. *Computer Graphics* 12(3), 12–19.

Smith, C., C. Richards, S. Brandt, and N. Papanikolopoulos (1996). Visual tracking for intelligent vehicle-highway systems. *IEEE Transactions for Vehicular Technology* 45(4), 744–759.

Stauffer, C. and W. E. L. Grimson (1999). Adaptive background mixture models for real-time tracking. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition, CVPR*, pp. 246–252.

Stauffer, C. and W. E. L. Grimson (2000). Learning patterns of activity using real-time tracking. *IEEE Trans. on PAMI* 22(8), 747–757.

Stokman, H. and T. Gevers (2001). Density estimation for color images. *Journal of Electronic Imaging* 10(1), 221–227.

Störring, M. (2004). *Computer Vision and Human Skin Colour*. Ph. D. thesis, Faculty of Engineering and Science, Aalborg University, Niels Jernes Vej 14, 9220 Aalborg, Denmark.

Su, C., H. Hang, and D. Lin (1999). Global motion parameter extraction and deformable block motion estimation. *IEICE Trans. Inf. and Syst.* E82-D(8), 1210–1218.

Tax, D. and R. Duin (1999). Data domain description by support vectors. In *Proc. ESANN*, pp. 251–256.

Tax, D., R. Duin, and M. Breukelen (1997). Comparison between product and mean classifier combination rules. In: Workshop on Statistical Techniques in Pattern Recognition. Prague, Czech Republic.

Teo, P. C. and D. J. Heeger (1994). Perceptual image distortion. In *Proc. IEEE Int. Conf. Image Processing*, pp. 982–986.

Tepichin-Rodriguez, E., J. Suarez-Romero, and G. Ramirez (1995). Hue, brightness, and saturation manipulation of diffractive colors. *Optical Engineering* 34(10), 2886–2890.

Terrillon, J., M. David, and S. Akamatsu (1998). Detection of human faces in complex scene images by use of a skin color model and of invariant fourier-mellin moments. In *Proceedings of the International Conference on Pattern Recognition*, Volume 3, pp. 1350–1355.

Toet, A. and M. Lucassen (2004). A new universal colour image fidelity metric. *Displays* 24(4–5), 197–207.

Traven, H. (1991). A neural network approach to statistical pattern classification by semiparametric estimation of probability density functions. *IEEE Trans. Neural Networks* 2(3), 366–377.

Tsang, P. and W. Tsang (1996). Edge detection on object color. In *IEEE International Conference on Image Processing*, Volume 3, pp. 1049–1052.

Tur, G., D. Hakkani-Tur, and R. Schapire (2005). Combining active and semi-supervised learning for spoken language understanding. *Journal of Speech Communication* 45(2), 171–186.

Valev, V. and A. Asaithambi (2001). Multidimensional pattern recognition problems and combining classifiers. *Pattern Recognition Letters* 22(12), 1291–1297.

Wahba, G. (1992). Multivariate function and operator estimation, based on smoothing splines and reproducing kernels. In M. Casdagli and S. Eubank (Eds.), *Nonlinear Modeling and Forecasting, SFI Studies in the Sciences of Complexity*, Volume XII, pp. 95–112. Addison-Wesley.

Wan, X. and C. Kuo (1996). Color distribution analysis and quantization for image retrieval. *Proceedings of SPIE* 2670, 8–16.

Wan, X. and C. Kuo (1998). A new approach to image retrieval with hierarchical color clustering. *IEEE Transactions on Circuits and Systems for Video Technology* 8(5), 628–643.

Wang, Z. and A. C. Bovik (2002). A universal image quality index. *IEEE Signal Processing Letters* 9(3), 81–84.

Wang, Z., H. R. Sheikh, and A. C. Bovik (2003). Objective video quality assessment. In B. Furht and O. Marqure (Eds.), *The Handbook of Video Databases: Design and Applications*, pp. 1041–1078. CRC Press.

Winkler, S. (1999a). Issues in vision modeling for perceptual video quality assessment. *Signal Processing* 78(2), 231–252.

Winkler, S. (1999b). A perceptual distortion metric for digital color video. In *Proc. SPIE*, Volume 3644, pp. 175–184.

Wren, C., A. Azarbayejani, T. Darrell, and A. Pentland (1997). Pfnder: real-time tracking of the human body. *IEEE Trans. on PAMI* 19(7), 780–785.

Xu, L., A. Krzyzak, and C. Suen (1992). Methods of combining multiple classifiers and their applications to handwriting recognition. *IEEE Trans. on Systems, Man and Cybernetics* 22(3), 418–435.

Yoo, J., M. Nixon, and C. Harris (2002). Model-driven statistical analysis of human gait motion. In *Proc. of the IEEE International Conf. Image Processing*, Volume 1, pp. 285–288.

Zheng, H. (2004). *Maximum entropy modeling for skin detection: With an application to Internet filtering*. Ph. D. thesis, ENIC - Telecom Lille 1 / Universit des Sciences et Technologies de Lille, Lille, France.