

# Video Monitoring and Analysis of Human Behavior for Diagnosis of Obstructive Sleep Apnoea

Ching-Wei Wang  
Department of Computing and Informatics  
University of Lincoln

March 2009

A thesis submitted to the University of Lincoln  
for the degree of Doctor of Philosophy.  
The work described in this thesis is entirely my own except where otherwise  
mentioned.

The research programme was carried out with a scholarship supported by  
the University of Lincoln and United Lincolnshire Hospitals NHS Trust.

## Abstract

This thesis investigates the use of the computerized video monitoring in support of the diagnosis of obstructive sleep apnoea, which is characterized by repetitive obstruction of the upper airways during sleep and resulted in arterial oxyhaemoglobin desaturation, excessive arousals, unrefreshing sleep, excessive daytime sleepiness, poor health-related quality of life, hypertension and severe life-threatening complications. According to recent research findings, the best predictors of morbidity are nocturnal oxygen saturation and movements during sleep. Although pulse oximetry is a well-established technique to analyze oxygen saturation, video monitoring and interpretation is less well developed due to the technical challenges of persistent occlusion, obscuration of the body by the bedding, variation of human behavior and the large volume of video data.

This work introduces a new automatic video monitoring technique for breathing behavior anomaly detection and assisting in diagnosis of obstructive sleep apnoea. The algorithm utilizes infrared video information, imposes few positional constraints on the patient, and deals with fully or partially covered bodies. A new motion detection model is presented to capture subtle and cyclical breathing signals. A novel action template is introduced to capture the dynamic spatial-temporal shape of normal breathing activities for action recognition, and adapts as the subject's pose changes. The online-constructed action template is used to classify an action as a normal breathing episode, an apnoea episode or a body movement episode. Although the presented approach is designed for diagnosis of obstructive sleep apnoea, it could be utilized in other applications that require the analysis of breathing behavior or monitoring subtle and cyclical activity.

This work also introduces two novel monocular video approaches (MatchPose and RTPose) for pose recognition of the covered human body. They are recommended for different purposes: RTPose provides coarse pose estimation and is computationally efficient; MatchPose produces fine pose estimation but takes 0.4 seconds to process a  $320 \times 240$  frame. If full body pose estimation is desirable, we recommend MatchPose. On the other hand, in the interests of computational speed, we recommend incorporating RTPose with motion information. The methods assume subjects lying horizontally. In addition, a low variance error boosting algorithm is developed for training head and upper leg pose templates.

In evaluation, we demonstrate that the breathing monitoring algorithm achieves high accuracy using confusion matrix in recognizing abnormal breathing activities and body movements and in classification of symptomatic and non-symptomatic subjects, and that the two pose estimation algorithms are able to identify human configurations with various poses and occlusion levels, and they are not particularly sensitive to environmental settings, including illumination and camera angle.

# Publications

Some of the work described in this thesis has been presented in journals, conferences and book chapters. Below is a complete list of publications arising during the course of this PhD study.

## Journal Articles

- Wang C.-W., , Hunter A. A Low Variance Error Boosting Algorithm, Journal of Applied Intelligence, Springer, Published online: 21 Feb 2009.
- Wang, C.-W., , Hunter A. A Novel Approach to Detect the Obscured Upper Body in application to Diagnosis of Obstructive Sleep Apnoea. IAENG International Journal of Computer Science, vol. 35, pp. 110-118, 2008.

## Conference Proceedings

- Wang, C.-W., , Hunter, A.. A Robust Pose Matching Algorithm for Covered Body Analysis for Sleep Apnoea. Proceedings of the 8th International Conference of IEEE BioInformatics and BioEngineering, 2008.
- Wang, C.-W., , Hunter, A.. A Simple Sequential Pose Recognition Model for Sleep Apnoea. Proceedings of the 8th International Conference of IEEE BioInformatics and BioEngineering, 2008.
- Wang, C.-W., , Ahmed, A., , Hunter, A.. Locating the Upper Body of Covered Humans in application to Diagnosis of Obstructive Sleep Apnoea. Proceedings of World Congress on Engineering, vol. 2, pp. 662-667, 2007. (*Best Student Paper of World Congress on Engineering*)
- Wang, C.-W., , Ahmed, A., , Hunter, A.. Vision Analysis in Detecting Abnormal Breathing Activity in application to Diagnosis of Obstructive

Sleep Apnoea. Proceedings of the 28th International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 4469-4473, 2006.

- Wang C.-W. Real Time Sobel Square Edge Detector for Night Vision Analysis, Proceedings of International Conference on Image Analysis and Recognition, Lecture Notes in Computer Science, LNCS 4141, pp 404-413, 2006.

### **Book Chapters**

- Wang, C.-W., , Hunter, A.. The Detection of Abnormal Breathing Activity by Vision Analysis in application to Diagnosis of Obstructive Sleep Apnoea. Encyclopedia of Healthcare Information Systems, Medical Information Science Reference, vol. 1, pp. 416-424, 2008.

# Acknowledgements

I would like to thank my supervisor, Prof. Andrew Hunter, for his invaluable advice, support and inspiration. I would also like to thank my co-supervisor Dr. Shigang Yue. It has been a great privilege to work under their guidance.

I would also like to thank Dr. Neil Gravill, Dr. Chris Hacking, Dr. Simon Matusiewicz, John Liversidge and Rosemary Barnes with all of whom I enjoyed very fruitful and stimulating research collaborations. This research was carried out through the generous financial support of United Lincolnshire Hospitals NHS Trust and the University of Lincoln. I would also like to thank Prof. Roy Davis and Dr. Grzegorz Cielniak for their valuable feedbacks in the Viva examination, Dr. Tom Duckett for his kind help in the Mock Viva, and Prof. David Walker for being the moderator in the Viva examination.

The faculty administration team and all members of the Vision and Robotics Research Centre have helped in making these past few years a thoroughly enjoyable experience for which I am truly grateful.

Finally, I am forever indebted to my parents for their love and unfailing faith in me.

# Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	Motivation . . . . .	2
1.1.1	Difficulties for Computer Vision Approaches . . . . .	3
1.2	Aim and Objectives . . . . .	4
1.3	Contributions and Proposed Methods . . . . .	5
1.4	Ethical Validity, Systems and Tools . . . . .	6
1.5	Thesis Outline . . . . .	6
<b>2</b>	<b>Background</b>	<b>8</b>
2.1	Obstructive Sleep Apnoea . . . . .	8
2.1.1	Prevalence - Underestimated and Undertreated . . . . .	9
2.1.2	Consequences of OSA Syndrome . . . . .	9
2.1.3	Definition of the Syndrome . . . . .	10
2.1.4	Existing Techniques for Diagnosis . . . . .	11
2.1.5	Limitations of Existing Techniques . . . . .	13
2.2	Monitoring of Breathing Activities . . . . .	14
2.2.1	Contact Type Techniques to Monitor Breathing . . . . .	14
2.2.2	Existing Non-invasive Techniques and Drawbacks . . . . .	15
2.2.3	Related Computer Vision Methods . . . . .	17
2.3	Monitoring of Covered Body Activities . . . . .	21
2.3.1	Pose Estimation Techniques . . . . .	21
2.3.2	Related Computer Vision Approaches . . . . .	22
2.4	Conclusion . . . . .	26
<b>3</b>	<b>Abnormal Breathing Detection</b>	<b>27</b>
3.1	Analysis of Breathing Behavior . . . . .	27
3.2	Related Work . . . . .	28
3.2.1	Motion Detection . . . . .	29
3.2.2	Motion Quantization . . . . .	31
3.2.3	Activity Recognition . . . . .	32

3.2.4	Online Training and Adaptive Action Template . . . .	33
3.3	Proposed Method . . . . .	34
3.3.1	Motion Detector for Breathing Analysis . . . . .	34
3.3.2	Adaptive Action Template to Capture Normal Breathing Patterns . . . . .	39
3.3.3	State Algorithm for Action Segmentation . . . . .	42
3.3.4	Action Recognition by Template Matching . . . . .	43
3.3.5	Parameter Definition and Deviation Detector . . . . .	46
3.4	Evaluation . . . . .	48
3.4.1	Experiments on Simulated Data . . . . .	49
3.4.2	Experiments on Clinical Data . . . . .	55
3.4.3	Model Parametrization and Sensitivity . . . . .	58
3.5	Conclusion . . . . .	61
<b>4</b>	<b>Pose Recognition of Covered Human Body</b>	<b>62</b>
4.1	Introduction . . . . .	62
4.1.1	Relevant Work . . . . .	63
4.1.2	Proposed Methods . . . . .	64
4.2	Ramanan Approaches . . . . .	65
4.2.1	Pictorial Structure for Pose Recognition . . . . .	66
4.2.2	Shape Matching: Chamfer Matching . . . . .	67
4.2.3	Experimental Results on Covered Human Data . . . .	70
4.3	The Weak Human Model . . . . .	73
4.3.1	Feature Extraction . . . . .	73
4.3.2	Obscured Head Detection . . . . .	76
4.3.3	Obscured Torso Detection . . . . .	83
4.3.4	Use Temporal Coherence . . . . .	85
4.3.5	Search Method: Greedy Search with Jumping . . . . .	87
4.4	Robust Pose Matching Method (MatchPose) . . . . .	88
4.4.1	Modified Pose Matching Algorithm (cwPose) . . . . .	88
4.4.2	The Integration Framework of MatchPose . . . . .	90
4.5	Real Time Simple Pose (RTPose) . . . . .	91
4.5.1	Reinforcement by the Linking Parameters . . . . .	92
4.5.2	Coarse Upper Leg Pose Recognition . . . . .	96
4.5.3	The Integration Framework of RTPose . . . . .	97
4.6	Evaluation . . . . .	98
4.6.1	Experimental Setup and Data . . . . .	98
4.6.2	Experimental Results . . . . .	100
4.6.3	Statistical Significance Test . . . . .	100
4.7	Conclusion . . . . .	108

<b>5</b>	<b>Conclusions</b>	<b>110</b>
5.1	Summary of Contributions . . . . .	110
5.1.1	Monitoring of Breathing Activity . . . . .	110
5.1.2	Pose Estimation of Covered Human Body . . . . .	111
5.2	Future work . . . . .	112
	<b>Bibliography</b>	<b>128</b>
<b>A</b>	<b>Terms and Definition</b>	<b>129</b>
<b>B</b>	<b>3–4 DT Algorithms</b>	<b>130</b>
<b>C</b>	<b>A Low Variance Error Boosting Algorithm</b>	<b>131</b>
C.1	Introduction . . . . .	131
C.1.1	Related Work . . . . .	132
C.1.2	Motivation . . . . .	133
C.2	Analyses of Benchmark Ensemble Learning Algorithms . . . . .	135
C.2.1	Bagging . . . . .	135
C.2.2	AdaBoost (AdaBoostM1) . . . . .	135
C.2.3	Modified AdaBoostM1 and MultiBoost . . . . .	136
C.2.4	Arcing . . . . .	140
C.3	Proposed Modification: cw-resampling . . . . .	141
C.3.1	cw-AdaBoost Algorithm . . . . .	144
C.3.2	cw-Arcing Algorithm . . . . .	144
C.3.3	cw-MultiBoost . . . . .	145
C.3.4	cw-Resample Algorithm . . . . .	146
C.4	Experiments . . . . .	146
C.4.1	Variance and Bias . . . . .	149
C.5	Conclusion . . . . .	149
C.6	Full Experimental Results . . . . .	153
<b>D</b>	<b>Full Statistical Test Results</b>	<b>156</b>
<b>E</b>	<b>Auxiliary Forms and Documents</b>	<b>161</b>

# List of Figures

2.1	Traditional Diagnosis Tools (a) PSG, (b) Portable PSG . . . .	12
2.2	Existing Video Monitoring System [158] using Patterned Sheet and Motion Detection . . . . .	14
2.3	Positional limitations of thermal imaging techniques: (a) the nasal area and the Segment of Carotid Vessel Complex [28]; (b) the subject's side face from a distance of six to eight feet [111]; (c) the frontal view of a face for the periorbital regions and the nose region [176]) . . . . .	16
2.4	Unconstrained poses in monitoring breathing activity (for illustration purpose, images are modified by adding brightness). . . . .	17
2.5	Motion History Images [17] for Action Recognition . . . . .	19
2.6	Action represented by Space-Time Shape Features [66]: (a) input; (b) degree of frequency; (c) plateness and stickness. Fast moving hands are identified as plates and appear in blue in (c); slow moving legs are identified as vertical sticks in the temporal direction and appear in green in (c); ball structure does not have any principal direction. . . . .	20
2.7	Thermal imaging fails to locate a true body posture because the heat remains on the bed after movements. #71: an unoccluded leg and a covered leg appear; #76: the image of the leg on the right indicates the covered leg's position, but the ghost on the left indicates the previous location of the unoccluded leg; #77 and #78: strong noises appear due to the remaining heat on the bed after movements of the occluded leg . . . . .	22
2.8	Fragmented Motion Data (upper column) and Raw Image (lower column): motion of the covered subject is fragmented and noisy as movement of the occluded subject also causes motion of the surface around rather than the exact area of the object, making object segmentation more challenging. . . . .	23

2.9	Person Model in [128] (a) the tree pictorial structure, (b) the edge template of the lateral walking pose pictorial structure, (c) a learned appearance template . . . . .	25
3.1	Cyclical Moving Flow in a Breath . . . . .	28
3.2	Top row: original images; middle row: DOF Results using $\alpha = 1$ ; bottom row: DOF Results using $\alpha = 2$ : with a low frame rate (7fps) and lowest possible $\alpha$ value, DOF can detect breathing movement, but the output signal is swamped by noise.	35
3.3	Activity maps by the proposed PLIM, effectively rendering clean breathing movements across a breathing cycle. . . . .	37
3.4	$e_t$ spectrum and characteristics for action recognition: the duration $d$ and peak value $h$ of each hill shape can be used to distinguish breathing action from large body action like body rotation (see section 3.3.5 for a simple action recognition model). However, small body movements can still be confused with apnoea episodes, and a more sophisticated action recognition model is introduced. . . . .	38
3.5	Normal Breathing Template Construction: a simple temporal aggregation model to combine spatial shapes for a continuous normal breathing period. . . . .	40
3.6	Motion captured with sensor noise: a simple temporal aggregation model suffers from high level of infrared sensor noise, and therefore a more sophisticated adaptive template model is introduced. . . . .	41
3.7	An adaptive action template in a clinical video data: as the normal breathing template adapts over time, the template at four different times, appearing differently, is selected for illustration purpose. . . . .	42
3.8	Various measurements ( $T$ : template and $A$ current activity). similarity ( $T \wedge A$ ) using the overlap 3; similarity ratio ( $\frac{T \wedge A}{T \vee A} = \frac{3}{2+3+4}$ ); dissimilarity ( $\sim T \wedge A = 4$ ) and ( $T \wedge \sim A = 2$ ). . . . .	44
3.9	Illustration of the Important State Switching Points: $t_s^1, t_s^2$ are the previous and new starting points to construct templates of normal breathing action; $t_e$ is the time to terminate the construction process; $t_m$ is the time to process template matching by comparing the action template $T(x, y, t_s^1, t_e)$ and the spatial shape $A(x, y, t_m)$ of the action at time $t_m$ . . . . .	47

3.10	Experimental Results of 11 simulated video clips: constant differences occur at the end points of individual abnormal events between the proposed method and the reference standard because of the stabilizing factor $n$ designed for switching status from “not normal breathing ”to “normal breathing ”. . . . .	51
3.11	Experimental Results of four simulated video clips: constant differences occur at the end points of individual abnormal events between the proposed method and the reference standard because of the stabilizing factor $n$ designed for switching status from “not normal breathing” to “normal breathing”. . .	52
3.12	Confusion Matrix of Action Classification on Simulated Data: the rows are ground truth; the columns are the proposed model results; the main diagonal shows the fraction of frames correctly classified for each class, and each row represents the probabilities of that class being confused with all the other classes. . . . .	52
3.13	Illustration of one experimental analysis output. . . . .	53
3.14	Illustration of two experimental analysis outputs. . . . .	54
3.15	Confusion Matrix on Clinical Data. . . . .	56
3.16	Template Matching Screenshots: given the timing $t$ of template matching, each row contains raw image $I$ , current motion vector $M$ , online constructed action template $T$ , matching result at time $t$ (yellow: $\sim T \wedge M$ ; blue: $T \wedge \sim M$ ; white: $T \wedge M$ ); the upper 3 rows are apnoea episodes and the lower 3 rows are body movement episodes. . . . .	57
4.1	Limited Motion and Raw Image: the motion data of a covered sleeping subject is irregular, partial, fragmented and noisy. . .	64
4.2	Diagram of the two proposed pose recognition methods. . . .	65
4.3	Person Model in Ramanan’s method: (a) a tree pictorial structure parameterized by probability distributions capturing the geometric arrangement of parts (the directed arrows between parts) and local part appearances (the vertical arrows into the shaded nodes) (b) the edge template of a generic lateral walking pose pictorial structure (c) a learned appearance template	66
4.4	Edge Orientation Cues and Distance-Transformed Images: (a) raw image, (b) combined edge orientation, (c–f) individual edge orientation vectors, (g) DT vector based on vector(e), (h) DT vector based on vector(f) . . . . .	69

4.5	Results of Ramanan <i>et al.</i> 's Method [128]: (a) input images, (b) torso pixels, (c) lower arm pixels, (d) lower leg pixels, (e) posterior, (f) mode of posterior. . . . .	71
4.6	Comparison of occluded and un-occluded image feature: (a) Un-occluded human used in [128], (b) Un-occluded edge cue, (c) Covered human, (d) Noisy and obscured edges . . . . .	72
4.7	Results of Iterative Parsing Technique [131]: (a) inputs (b) the first edge-based parse (c) the second parse using region features from first parse (d) the best-scoring pose . . . . .	72
4.8	Comparison of general edge detectors and the proposed coarse horizontal oriented edge detector . . . . .	74
4.9	Edge Box Maps for Shape Abstraction: the torso can be detected under occlusion from the subject's arms, hands and the cover by utilizing the edge box maps. . . . .	75
4.10	Cascade Structure: (a) Boosted Cascade with Single Model [157], (b) Proposed Cascade with Diverse Models . . . . .	77
4.11	Image observations for head detection: (a) Horizontal oriented edge image $I_1$ (b) image by Prewitt kernels and a contrast enhancement filter $I_2$ (c) Vertical oriented edge image $I_3$ (d) Definition of potential areas of shoulders $S_1, S_2, S_3, S_4$ , and the top of the head $R_t$ to the head region $R_h$ . . . . .	79
4.12	Initial training data for $\mathcal{M}_1$ : the data format is a $14 \times 14$ matrix using the intensity values of edge images $I_1$ . . . . .	79
4.13	Iterative Construction of Boosting Cascade . . . . .	81
4.14	Edge box maps: regions with high $w(I_{tor}, X^{tor})$ are highlighted . . . . .	85
4.15	Estimated Pose and Image Observation for Hip Joint Detectors . . . . .	86
4.16	Temporal coherence on patterns. . . . .	86
4.17	Results of the search algorithm . . . . .	87
4.18	Pose Pictorial Structure: (a) Ramanan's Template Representation (b) Modified Template Representation . . . . .	88
4.19	Results of Improved Pose Matching Model: (a) inputs (b) edge orientations (c) the best scoring pose (d) top 25 poses are highlighted in white, with the best scoring pose highlighted in red . . . . .	89
4.20	Relationship between two model parameters: (a) Markov Network (b) Reinforcement network . . . . .	92
4.21	(a) input raw images (b) images by a convolution filter (c) DOF outputs using the processed images by convolution filter (d) DOF output using raw images. . . . .	94

4.22	(a) raw image (b) system output (c) image observation $I_1^h$ by Prewitt edge detector (d) reinforced feature space $I_1^h$ with the associated hypothesis $X_1^h$ (e) motion data $M_t$ (f) reinforced motion data $M_t'$ with the associated hypothesis $X_2^h$ (g) auxiliary image observation $I_2$ (h) reinforced auxiliary image observation $I_2^h$ with its associated hypothesis $X_3^h$ . . . . .	95
4.23	(a) edge box map with $X_t^h$ (b) reinforced edge box map . . . .	95
4.24	(a) Representation of upper-legs pose model (b) Some edge box maps. . . . .	96
4.25	10 upper-legs pose templates. . . . .	96
4.26	(a) Head hypotheses $\{X^h\}$ by head detectors (b) Head to torso search: $\{X^h, X^{tor}\}$ (c) Compare $\{X^{tor}\}$ and choose the strongest one as $X^{tor*}$ . (d) Torso to head search: output hypothesis $(X^{h*}, X^{tor*})$ . . . . .	98
4.27	RTPose Outputs: (a) various poses with the same subject and environment setting as the training data (b) various occlusion levels with different environment setting from the training data	102
4.28	RTPose Outputs of eight different subjects, excluding the one used in the training video clip. . . . .	103
4.29	MatchPose Outputs of various poses are highlighted with white rectangles (and red rectangles if multiple configurations are obtained with the minimum chamfer matching cost), using the same subject as in the training video clip . . . . .	104
4.30	MatchPose Outputs on eight subjects, different from the one in the training video clip. . . . .	105
4.31	Edge Orientations of Heavily Obscured Data and Erroneous Detection by Ramanan. . . . .	106
4.32	Misdetctions by (a) MatchPose and (b) RTPose. . . . .	106
C.1	Fast convergence of cw-AdaBoost and cw-Arcing . . . . .	139

C.2	Illustration of the proposed design and low generalization issues of existing boosting methods: If $C_{13}$ is an error free classifier, boosting methods in the first group will produce only 13 classifiers no matter how large the number of base models originally specified, and as $C_{13}$ gains infinite decision power, the decision of these ensembles is dominated by one base classifier; boosting methods in the second group assign considerably high decision power to the two error free models, $C_{13}, C_{15}$ , and thus the decision is dominated by these two base classifiers; boosting methods in the third group continuously produce identical classifiers $C_{13}$ , and the decision of such ensemble models is dominated by this one classifier; the proposed structure generate diverse classifiers and an effective ensemble with 30 different and 18 mature decision makers. . . .	142
C.3	Re-sampling Scheme . . . . .	143
C.4	Comparison on $Bias^2$ , $Variance$ and Error between the original algorithms and the algorithms with the proposed modifications. . . . .	151

# List of Tables

3.1	Measurements for Abnormal Events . . . . .	44
3.2	VAHI Values . . . . .	59
3.3	Values of Model Parameters . . . . .	60
4.1	Experimental Results on Shoulder Detection . . . . .	84
4.2	Evaluation Data Distribution . . . . .	99
4.3	Recognition Rates . . . . .	101
4.4	Significance Performance Test Results . . . . .	109
C.1	MultiBoost Performance with different $B_i$ . . . . .	139
C.2	10-fold Cross Validation Accuracy% for single dataset(Breast Cancer) . . . . .	147
C.3	Performance Index $P_m$ on 12 Gene Expression Datasets . . . .	148
C.4	$Bias^2$ , $Variance$ and $Error$ . . . . .	150

# Chapter 1

## Introduction

This thesis investigates automated video monitoring of breathing activity invariant to pose and occlusion, and pose estimation of the covered human body, which has not been widely studied. The aim of this research is to develop automated video approaches for diagnosis of obstructive sleep apnoea, which requires analysis of breathing behavior and body activity during sleep. In this introductory chapter, we present the motivation of this research, and then give an outline of the contributions and structure of the thesis.

### 1.1 Motivation

Obstructive Sleep Apnoea (OSA) syndrome was first identified only 43 years ago [59, 63]; its clinical importance is increasingly recognized. OSA is characterized by repetitive obstruction of the upper airways during sleep, resulting in oxygen de-saturation and frequent arousals. Importantly, OSA is a condition that not only presents with symptomatology troubling to the patient and their family, but also has severe complications which may be life-threatening. Reduction in cognitive function, cardiovascular diseases, stroke, decreased quality of life, fatigue and excessive day time sleepiness are common among OSA patients.

Although OSA is acknowledged as a worldwide problem, which in Western countries affects around 4% of men and 2% of women [49, 63], the majority of affected individuals remain undiagnosed. Some studies have suggested that figures are much higher [112, 126, 145, 154]. Due to lack of awareness among the general population and physicians, Hossain and Shapiro [72] suggested that an estimated 80% to 90% of OSA sufferers have not received a clinical diagnosis.

The standard diagnostic tool for sleep apnoea is Polysomnography (PSG),

which measures a wide range of parameters, including brain waves, eye movements, muscle activity or skeletal muscle activation, heart rhythm, airflow, respiratory effort, and blood oxygen saturation using a range of sensors. However, PSG requires costly measurement devices and labor-intensive work for the electrode hook-ups. Worse, it disturbs sleep and compromises results. Thus, there is growing interest in alternative approaches to replace PSG in the diagnostic assessment of patients with suspected sleep apnoea.

Instead of using the entire set of PSG, pulse oximetry (which measures blood oxygen saturation levels) in conjunction with noninvasive video monitoring has been utilized for diagnosis of OSA in the hospital. In principle, the medical doctor identifies doubtful areas on the pulse oximetry trace and reviews the video data during these identified periods. However, in practice, the pulse oximetry traces of some OSA patients do not show any abnormalities, and the medical doctor has to review the overnight video. Sivan *et al.* [144] indicate that the results from traditional PSG are highly correlated with the manually marked video test results. In addition, according to recent research findings [15, 89, 123], the best predictors of morbidity in individual patients, as assessed by improvements with CPAP (continuous positive airway pressure therapy), are nocturnal oxygen saturation levels of the blood and movement during sleep.

However, video monitoring and interpretation of OSA is not well developed due to the technical challenges, including heavy occlusion and obscuration of the human subject by bedding, variations in human size, sleeping posture and breathing behavior, changes of the subject’s facing with respect to the camera and difficulty in detecting breathing from video. Existing video monitoring techniques [158] utilize patterned sheets and infrared light to compute gross degrees of motion. However, gross motion suggests only periods of time with movements rather than identifying what the activities are, and therefore requires clinicians to review substantial amounts of video data manually – a time-consuming and expensive process.

As a result, there is a practical demand for automated methods that objectively and reliably detect OSA from video. There are two major activities of interest: breathing activities, and body movements such as limb movements.

### 1.1.1 Difficulties for Computer Vision Approaches

Automated video monitoring of covered human body movements and breathing activity are both challenging tasks. For monitoring breathing activity, existing computer vision approaches utilize thermal imaging [28, 111, 176] to capture a breathing signal. However, there are *strict positional constraints* as

these methods target the face. In addition, the regions of interest for these methods must be unoccluded. Moreover, the expense of thermal imaging technology is high.

Conventional motion detection approaches prove unsuitable for the capture of breathing movements. Background differencing fails because the salient motion operates so slowly that it is swamped by noise; optical flow has weaknesses with respect to cyclical movements, as an object that moves in a straight line but oscillates forwards and backwards has low salience. To the author’s best knowledge, there is no existing method to automatically analyze human breathing behavior from video, robust to occlusion and without positional limitations.

Further difficulties include the varying appearance of breathing activity due to changes of the pose and occlusion level, and variation in the region of interest, since breathing movements may occur in different regions such as the jaw, the chest area, the abdominal area or the shoulders. A detailed discussion is given in chapter 3.

Regarding the monitoring of covered body activities, it is difficult to obtain full body silhouettes due to partial and irregular movement during sleep. Thus, the model-based framework that identifies body pose first and analyzes activity using the estimated pose and detected motion is adopted. The first task is therefore to estimate the covered body pose, which however remains a challenging task. Many existing approaches to pose estimation make simplifications of the measurement problem, either using motion data (e.g. [1, 45, 67, 146]) to extract silhouettes, or assuming knowledge of appearance or color (e.g. [40, 95, 128, 136]), and the subjects tend to wear close-fitting clothing (or even to be unclothed [40]) in order to extract such information more easily. These methods are too restrictive and not applicable to the problem in the field of this study. Although there is some published research investigating the monitoring of partially occluded humans [69, 129, 153, 170], the methods examined do not deal with pose estimation of consistently and almost wholly occluded subjects.

## 1.2 Aim and Objectives

The aim of this research is to investigate the use of computerized video monitoring in support of the diagnosis of OSA. The primary objective is to detect abnormal breathing episodes based on video analysis. This requires the model also to distinguish breathing movements from other body movements. The secondary is to perform on pose estimation of covered humans to allow future work to recognize human activities during sleep.

### 1.3 Contributions and Proposed Methods

The first contribution of this work is a novel automatic video monitoring technique for breathing behavior anomaly detection, assisting in the diagnosis of OSA. The algorithm utilizes infrared video information, avoids imposing positional constraints on the patient, and deals with a fully or partially obscured patient body. A new motion detection model is introduced to capture subtle and cyclical breathing signals. Moreover, regarding action as a spatial-temporal shape, a normal breathing action template is created to model the shapes of normal breathing activities for action recognition. As the subject changes pose over time, the shapes of normal breathing activities change, which requires accompanying changes of the template. A dynamic action template model of normal breathing behavior is presented, which is used both to distinguish body movements from breathing episodes, and to classify breathing episodes as either a normal breathing episode or an apnoea episode. Although the approach is designed for diagnosis of OSA, it could also be utilized in other applications that require the analysis of breathing behavior or monitoring subtle and cyclical activity.

Apart from breathing activity, body movements such as limb movements are also used for diagnosis of sleep apnoea. For example, periodic limb movements during sleep are a common finding in patients with OSA [68, 82].

The second contribution is two novel monocular video algorithms, MatchPose and RTPose, to robustly locate a persistently and (fully or partially) occluded prone human body pose, which allows future work to recognize human activities during sleep. They are recommended for different purposes: if full body pose estimation is desirable, we recommend MatchPose; in the interests of computational speed, we recommend incorporating RTPose with motion information. The two pose estimation methods assume subjects lying horizontally. They are demonstrated to recognize poses of different obscured subjects with various occlusion levels and poses, and the methods are not particularly sensitive to changes of IR illumination and camera angle.

A robust Weak Human Model (WHM), which combines a number of obscured part detectors to accommodate various levels of occlusion and body postures, is introduced to effectively and efficiently identify a few upper body poses. In MatchPose, WHM is combined with an improved pose matching model (cwPose) to recover the full body pose of covered subjects. In RTPose, WHM is integrated with an upper leg pose estimator (cwULeg), which uses a novel representation to capture latent image features, and a new reinforcement tracker to reinforce both feature space and model parameter space. Although the methods are designed for sleep study, they could also be utilized in other applications that require the analysis of human poses and

behavior in situations with heavy obscuration or persistent occlusion.

The third contribution is a low variance error boosting algorithm that uses weight perturbation to reduce variance error, and is particularly effective when dealing with data sets, which have large numbers of features and small number of instances. The algorithm is used to train the obscured head template of WHM and upper leg pose templates of cwULeg.

The proposed methods for breathing monitoring and pose estimation overcome the difficulties caused by poor quality image cues due to heavy occlusion and obscuration, and to large variances in image features due to unpredictable human behavior (e.g. when removing or pulling back the cover).

## 1.4 Ethical Validity, Systems and Tools

**Research Ethics.** Two NHS research ethics applications were made for the project. The first one in March 2007, with REC reference number 07/Q2403/38 was not approved; the subsequent application in January 2008 was accepted, and in February 2008 Research Ethics approval was gained from Derbyshire Research Ethics Committee (REC number 08/H0401/12). Symptomatic and non-symptomatic volunteers were recruited to participate the study; informed consent forms and related documents such as the information sheets and GP letters are attached in Appendix E.

**Systems and Tools.** A video monitoring system has been installed in the sleep study room situated on Carlton Coleby Ward in the United Lincolnshire Hospital, UK in May 2008; testing of the system ends in September 2008. Prior to the research ethics application approval, a temporary video monitoring system was installed in the author's bedroom using camcorders and tripods. All software is implemented in C#, Microsoft Visual Studio .Net. Microsoft Directshow (DirectX) was used to decode video data, which was recorded and compressed into the WMV9 format.

## 1.5 Thesis Outline

The remainder of this thesis is arranged as follows:

In **chapter 2**, the medical background of OSA is given, with a review of current diagnostic techniques, followed by further discussion of existing approaches and potential techniques to monitor human breathing activities and covered human body activities.

In **chapter 3**, a new noninvasive real-time video monitoring technique is introduced for detecting abnormal breathing activities and for assisting in the diagnosis of OSA.

In **chapter 4**, an existing stylized pose detector [128] is reviewed and applied to the covered human video data; the performance is shown to be extremely poor due to the heavily obscured image features. Two markerless pose estimation approaches without manual initialization are introduced to estimate the pose of covered human subjects from image sequences. In evaluation, the techniques are used to estimate the covered body pose with various postures and obscuration levels.

In **chapter 5**, the contributions of this thesis are reviewed, and future areas for consideration are discussed.

# Chapter 2

## Background

This chapter describes the medical background of OSA and current diagnostic techniques in section 2.1, and discusses techniques to monitor breathing activities and the covered human body in section 2.2 and 2.3 respectively.

### 2.1 Obstructive Sleep Apnoea

OSA was first identified only **43 years ago** [63] and its clinical importance is increasingly recognized. OSA is one of the most common sleep disorders and occurs with similar frequency to Type 1 diabetes and twice that of severe asthma. It affects an estimated 4% of males and 2% of females in the UK, although the prevalence is thought to be considerably higher in specific groups and occupations, where the consequences can be fatal or lead to serious injury if left undiagnosed and un-treated. It is now well established that OSA is associated with an increased risk of cardiovascular disease, and patients with sleep apnoea have a high prevalence of the risk factors that comprise the metabolic syndrome, namely: central adiposity, dyslipidaemia, high blood pressure, insulin resistance, and hyperglycaemia [32, 113, 125].

There are two types of apnoea: *obstructive*, in which air flow ceases but movement of the chest wall (rib cage and abdomen) persists, implying respiratory effort in the face of a closed upper airway; and *central*, in which both flow and movement cease, apparently because of cessation of the drive to breathe. The primary focus of this research is on OSA, but this does not mean that our algorithm is unsuitable for central apnoea or other breathing disorder syndromes. The rest of this section describes the prevalence, consequences and syndromes of OSA and existing OSA diagnostic tools.

### 2.1.1 Prevalence - Underestimated and Undertreated

Although OSA is acknowledged as a worldwide problem, which in Western countries affects around 4% of men and 2% of women [49, 63], the majority of affected individuals remain undiagnosed. In the USA, about 70 million Americans suffer from a sleep problem, nearly 60% of them have a chronic disorder, and the second most common sleep disorder is OSA, which affects about 18 million Americans; additionally, other sleep disorders add an estimated \$15.9 billion to the National Health Care Bill [137]. In the UK, sleep disorders affect about 770,000 people [122].

Some studies have suggested that figures are much higher. Sjostrom *et al.* [145] estimate 24% of men and 9% of women in the middle aged population suffer from OSA, whilst Neven *et al.* [112] estimate that at least 45% of men aged 35 and over suffer from clinically significant OSA. In Asia, a study from Singapore [126] indicates that prevalence of OSA to be around 15% in the country. Another study in India [154], the researchers found that the prevalence of OSA was 7.5% in *healthy urban Indian males* between 35–65 years of age.

Due to lack of awareness among the general population and physicians, Hossain and Shapiro [72] suggested that an estimated 80–90% of OSA sufferers have not received a clinical diagnosis, and in the Wisconsin sleep cohort study [173], 93% of women and 82% of men with moderate-to-severe sleep apnoea did not receive diagnoses.

Thus, there is a growing interest in alternative approaches to the diagnosis of OSA as substitutes for labor-intensive and time-consuming PSG.

### 2.1.2 Consequences of OSA Syndrome

**Increases Risk of Heart Attack and Death.** OSA increases a person's risk of having a heart attack or dying by 30% over a period of four to five years, according to a new study [139], which includes 1123 patients referred for sleep apnoea evaluation. All patients underwent an overnight sleep study to determine if they had OSA, and over the next four to five years, they were followed to see how many had any heart disease events (heart attack, coronary angiography or bypass surgery) or died. The researchers indicated that sleep apnoea triggers the body's "fight or flight" mechanism, which decreases the amount of blood pumped to the heart; consequently, repeated episodes every night over several years can starve the heart of enough oxygen, when combined with the body's decreased oxygen intake due to the frequent breathing stoppages. Another recent study [164] compared mortality in three groups: 113 patients with heart failure, but little or no OSA; 37 such patients with

untreated moderate to severe OSA; and 14 with OSA treated with continuous positive airway pressure, and the presence of untreated OSA seemed to double mortality from heart failure over five years from 12% to 24%; there were no deaths in the small group treated with continuous positive airway pressure. Furthermore, there is evidence that treatment of OSA improves cardiac function [86], and untreated OSA patients have a higher risk of recurrence of atrial fibrillation after successful cardioversion than patients without known OSA [85].

**Earlier Death in Stroke Patients.** Sleep related breathing disorders (SRBDs), including Sleep Apnea, Cheyne-Stokes Respiration and Alveolar Hypoventilation Syndrome, are both a risk factor for and are common in patients with stroke, and it is found that SRBDs may have adverse impact on survival and prognosis [93]. Stroke victims who have OSA die sooner than stroke victims who do not have OSA, according to [140]. The researchers followed 132 stroke patients over 10 years, and 23 of those patients had OSA. The findings indicate the importance of a clinical trial for stroke patients with OSA to see whether treating the sleep disorder will extend their lives.

**Association with Eye Disease.** Multiple studies have identified OSA as an independent risk factor for the development of several medical conditions, including high blood pressure, which are related to impairments or alterations in a person's vascular (circulatory) system. With their own complex and sensitive vascular system, the eyes can be affected by systemic vascular problems. According to [141], a variety of ophthalmologic conditions are associated with OSA, including floppy eyelid syndrome, glaucoma (the second most common cause of blindness and the most common cause of irreversible blindness), Nonarteritic anterior ischemic optic neuropathy (NAION), and Papilledema.

### 2.1.3 Definition of the Syndrome

An Apnoea-Hypopnoea index (AHI) is generally used for evaluation of the severity of OSA and is calculated as the average number of apnoeas plus hypopnoeas, per hour of sleep. From [93], the clinical and research definitions of apnoea (and associated hypopnoea) are described below:

**Clinical Definition:** Apnoea is defined as a cessation of airflow for  $> 10$  seconds. The event is obstructive if during apnoea there is effort to breathe; the event is central if during apnoea there is no effort to breathe. Several

clinical definitions of hypopnoea are in clinical use and there is no clear consensus. An approved definition of hypopnoea is an abnormal respiratory event with at least 30% reduction in thoracoabdominal movement or airflow as compared to baseline lasting at least 10 seconds, and with  $\geq 4\%$  oxygen desaturation.

**Research Definition:** Apnoea is defined as a clear decrease ( $> 50\%$ ) from baseline in the amplitude of a valid measure of breathing during sleep lasting at least 10 seconds (note, there is little differentiation between OSA or hypopnoea). Hypopnoea is defined as a clear decrease ( $< 50\%$ ) from baseline that is associated with an oxygen desaturation of  $> 3\%$  or an arousal.

The apnoea often ends with a loud snore or gasp, along with movements of the whole body. This awakening is sufficient to make the patient's throat opening muscles work so (s)he can breathe in again, but (s)he usually falls asleep again so quickly that (s)he does not remember it happening. In OSA, this cycle repeats itself throughout the night as the muscles relax and the throat blocks off again. During sleep, the intervals between the breaths (apnoeic spells) or the reduction of the depth of breathing (hypopnoea) lead to a decrease of the oxygen in the blood and will cause the afflicted person to wake up many times during the night.

#### 2.1.4 Existing Techniques for Diagnosis

The standard diagnostic tool in sleep medicine is Polysomnography (PSG), as displayed in Figure 2.1. PSG measures a wide range of variables and monitors body functions, including brain waves by electroencephalography (EEG), eye movements by electrooculography (EOG), skeletal muscle activation by electromyography (EMG), heart rhythm by electrocardiography (ECG), airflow by thermistor or pressure transducer, respiratory effort by thoracic-abdominal bands, and blood oxygen saturation by pulse oximetry.

Flemons *et al.* [49] categorize general sleep monitoring techniques into four types, including: Type 1 Monitoring (standard PSG), Type 2 Monitoring that incorporates sleep staging and respiratory measures with a minimum of seven channels, Type 3 Monitoring using at least three respiratory channels (ventilation or airflow, heart rate or ECG, oxygen saturation) and Type 4 Monitoring utilizing at least one respiratory channel, usually either oxygen saturation or airflow.

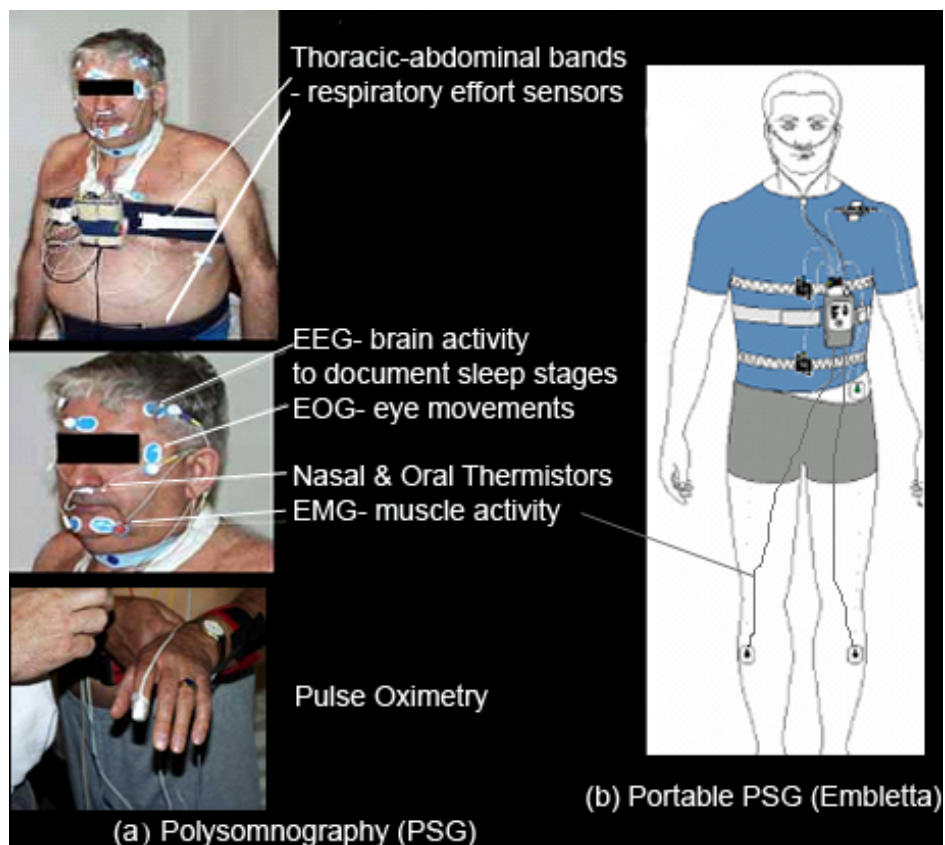


Figure 2.1: Traditional Diagnosis Tools (a) PSG, (b) Portable PSG

### 2.1.5 Limitations of Existing Techniques

Flemons *et al.* [49] point out two drawbacks of PSG. Firstly, thermistors may not be significantly sensitive to detect hypopneas. Secondly, nasal pressure can produce poor results if patients are mouth breathing. Furthermore, PSG requires labor-intensive work to attach the sensors onto the patient's body and expensive equipment for diagnosis. It also provides an intrusive monitoring environment, which disturbs sleep and compromises results; see Figure 2.1. Thus, there is growing interest in alternative approaches to PSG.

Video monitoring in conjunction with pulse oximetry has been adopted to assist diagnosis of OSA. According to recent findings [15, 89, 123, 144], the best predictors of morbidity in individual patients, as assessed by improvements with Continuous Positive Airway Pressure (CPAP) therapy, are nocturnal oxygen saturation and movement during sleep, rather than the Apnoea-Hypopnoea Index, which is calculated as the average number of apnoeas plus hypopnoeas per hour of sleep. In [144], Sivan *et al.* showed that the results from traditional PSG are highly correlated with the manually analyzed video test results. The protocol is that the medical doctor identifies doubtful areas on the pulse oximetry trace and reviews the video data during these identified periods. However, the pulse oximetry traces of some OSA patients show no abnormality. For example, the lung of the athlete works so efficiently that the blood oxygen saturation sensed by the pulse oximetry appears normal even though the subject suffers from OSA. In such cases, the medical doctor has to review the overnight video.

Automated video monitoring and interpretation of OSA is under development due to the computational complexity of video analysis. The current approach for video analysis is for clinicians to review substantial amounts of video data manually. Existing video systems in the sleep lab in United Lincolnshire Hospital, United Kingdom [158] utilize motion sensors, patterned sheets and infrared light to detect gross degrees of motion, which suggest periods of activity but do not identify what the activities are. Moreover, as Matusiewicz and Gravill [107] pointed out, once the patterned cover is removed by the patient the system fails to detect human activities, and produces useless motion information. Hence, clinicians still have to analyze substantial amounts of video data, which is a time-consuming and expensive process. In addition, the diagnosis is subject to human error and to the uncertainty of subjective judgments. Hoffstein *et al.* [70] studied 25 cases, all of whom had full nocturnal PSG, including the measurement of snoring, to compare the subjective snoring count by two listeners during a 20 minute segment. In 7 out of 25 patients, the difference in subjective snore counts perceived by the listeners was larger than 25%.



Figure 2.2: Existing Video Monitoring System [158] using Patterned Sheet and Motion Detection

As a result, there is a practical demand for automated methods that objectively and reliably analyze human action from video. The clinical experts from Lincoln County Hospital [107] identified two major themes for monitoring: human breathing activities and body movements such as limb movements. Both are challenging since the subject tends to be heavily and persistently occluded. In the following two sections, potential or existing techniques, including non-video approaches, are discussed for their suitability to analyze human breathing activity and covered body activity during the subject's sleep.

## 2.2 Monitoring of Breathing Activities

Monitoring of breathing has broad applications such as polygraph (popularly referred to as a lie detector), sleep studies, sport training, early detection of sudden infant death syndrome in neonates, and patient monitoring. Current breathing monitoring techniques can be categorized into two types: invasive and non-invasive.

### 2.2.1 Contact Type Techniques to Monitor Breathing

The contact type approaches include thoracic-abdominal bands [80, 150], which track changes in the body circumference during the respiratory cycle, stick-on electrodes such as the Electrocardiogram (ECG) method [110], the nasal temperature probe [149] and contact-type microphone for audio analysis to monitor tidal volumes from human breathing activity [5, 78]. In

a typical breath monitor, a thermistor, an accelerometer, or a contact-type microphone must be attached directly to the person’s body, near the nose and the mouth, or over the chest wall and the trachea.

The main disadvantage of all the aforementioned technologies is that they require close contact with the subject, which in certain cases may be quite uncomfortable and not practical. Apart from the invasive nature of these monitoring equipments, which disturbs sleep and therefore compromises results, the thermistors used in PSG sense differences in temperature, but they do not have a linear relationship with true airflow, and consequently may not be sufficiently sensitive to detect hypopneas [49]. Furthermore, nasal pressure gives a linear approximation to airflow but can produce false-positive events and low quality signals if patients are mouth breathing [49]. Regarding the thoracic-abdominal bands, if the tension on the strap is not calibrated, the system will not track the respiration motion correctly, so that adjustment may be necessary. In addition, measurements on patients with shallow and abdominal breathing patterns may fail because the sensor cannot track adequately in a reproducible manner if the chest displacements during normal breathing and breath-hold are not distinctly different. More seriously, the invasive approaches very often fail to monitor continuously because the devices can be pulled off by the subject during sleep unconsciously. For example, in 48 percent of the cases (21 of 43 cases) in the clinical study conducted in [118], the monitoring system failed to measure the physiological values continuously enough to diagnose disease.

### **2.2.2 Existing Non-invasive Techniques and Drawbacks**

Published non-invasive techniques include non-contact type audio analysis [29, 120], vibration sensors [104, 134], and thermal imaging [111, 28, 176]. A major challenge for non-contact type audio analysis is the extraction of breathing sounds from sensor signals contaminated by environmental noise. Cheng *et al.* [29] developed a portable device (SOD) to detect snores, but the SOD is not intended to be a diagnosis device for OSA. Instead, the device is to be used as a precautionary measure for monitoring snoring at home, and subjects whose snoring patterns are classified as possible OSA symptoms by the device are suggested to consult doctors for further diagnosis. Although the preliminary study [120] suggests that the bispectral analysis of snore signals might be useful to distinguish apneic patients from benign patients, it has not been proved that diagnosis can be obtained from the analysis of audio monitoring of snoring.

A preliminary study of vibration sensors to monitor breathing activities was conducted in [104]. The subject had the vibration sensor pad placed

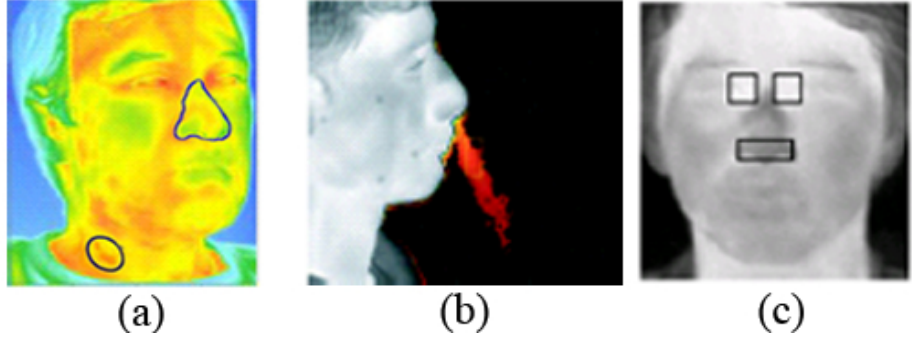


Figure 2.3: Positional limitations of thermal imaging techniques: (a) the nasal area and the Segment of Carotid Vessel Complex [28]; (b) the subject’s side face from a distance of six to eight feet [111]; (c) the frontal view of a face for the periorbital regions and the nose region [176])

underneath their calf while they were lying on their back on a bed, and with the subject lying on their stomach on the bed the sensor was placed under their chest. Due to the expensive hardware, the positional and postural constraints and the physicians’ preferences for video monitoring, this technique is not investigated further in this research.

Concerning thermal imaging techniques [28, 111, 176], the researchers utilize thermal imaging to capture the human breathing signal. In [111], the system captures the profile view of the subject’s side face from a distance of 1.83 to 2.44 meter to monitor the air flow through the nose and mouth. Chekmenev *et al.* [28] monitor the nasal area and the Segment of Carotid Vessel Complex, and indicate that the temperature is relatively high around the eye region, especially periorbit, a small area between the eye and bridge of the nose. They measure the subject’s face and neck from a distance of one meter due to the limitations of the existing optics of the camera. In [176], the frontal view of the subject’s face is captured to monitor the periorbital regions and the nose region. However, there are *strict positional limitations* for targeting faces, as shown in Figure 2.3. Moreover, the regions of interest for these methods need to be visible without occlusion. These requirements are not easily fulfilled when monitoring humans during sleep; see Figure 2.4. As a result, thermal imaging does not appear to be a suitable option.



Figure 2.4: Unconstrained poses in monitoring breathing activity (for illustration purpose, images are modified by adding brightness).

### 2.2.3 Related Computer Vision Methods

One of the objectives of this research is to detect abnormal breathing episodes from video. However, this is a challenging task. To the author’s best knowledge, barring the thermal imaging approaches discussed previously, there is no existing method identified to deal with capturing and analyzing breathing behavior from video. This is due to two major technical challenges: breathing movements are so subtle that they are difficult even for human eyes to observe; and motion self-occlusion of the cyclical breathing movements, which further increases the difficulty in action recognition (e.g., in hand gesture recognition, waving the hand is often confused with moving the hand from left to right only; in breathing monitoring, head movements can be confused with apnoea over-breathing actions).

**Capturing Breathing Signals.** The first technical challenge is to capture the breathing signals from video. Conventional motion detection approaches such as differences of frames (DOF) have been proven to be unsatisfactory in [169] and in our experiments (see chapter 3.2), because the frame to frame motion of non-salient objects may be larger than that of salient objects, especially if the salient object is moving relatively slowly. In our case, the salient object moves so slowly that noise generated by the sensor is comparatively large.

Another traditional technique – optical flow is, however, extremely computational expensive. In addition, Lipton [101] indicated that most optical flow algorithms fail in largely homogeneous regions (ie. regions lacking texture). Furthermore, Wixson [169] stated that an object that moves in a straight line but oscillates forwards and backwards, would have low salience. These weaknesses of optical flow applied to large homogeneous regions, cyclical movements and real-time performance make it unsuitable for this prob-

lem domain as the subject is largely occluded by the blank cover, breathing movement is cyclical, and the amount of video data to process is substantial.

**Action Recognition.** Assuming that breathing signals have been successfully captured, the next problem is to recognize abnormal breathing actions. Over the past two decades, the number of papers within the field of action recognition using computer vision has grown significantly. An important distinction is to look at whether the recognition is static or dynamic, i.e., whether the recognition is based on one or more frames. The simple static recognition approach is used mainly to recognize various postures, e.g., pointing, standing and sitting, or specially defined postures used in interfaces. As there is no distinctive posture, which can represent an abnormal breathing action like an apnoea episode, the simple static recognition approach is inapplicable.

On the other hand, the approaches of dynamic recognition use temporal characteristics in the recognition task. In 1975, Johansson [84] showed in his moving lights displays (MLD) experiments that the actions of a human may be recognized solely from sparse motion signals (of the lights). Recent successful work in the area of action recognition [17, 43, 66, 88, 116, 172] has shown that it is useful to analyze actions by treating a video sequence as a three dimensional space-time volume (of intensities, gradients, optical flow or other local features). These techniques are categorized and discussed below.

**Action Recognition using Optical Flow.** Efros *et al.* [43] perform action recognition by correlating optical flow measurements from low resolution videos. However, the weaknesses of optical flow applied to large homogeneous regions, cyclical movements and real-time performance make it unsuitable for this problem domain.

**Action Recognition using Silhouettes.** Bobick and Davis [17] propose a static vector-image as a temporal template to represent human movement, where the vector value at each point is a function of the motion properties at the corresponding spatial location in an image sequence, and they introduced the global descriptors Motion History Image (MHI) and Motion Energy Image (MEI) as spatiotemporal templates to be matched to stored models of known actions. To construct action templates, MEI and MHI of pre-recorded actions are collected to produce statistical models using 7 Hu moments [73]; to recognize an input action, a Mahalanobis distance is calculated between the moment description of the input and each of the known movements.

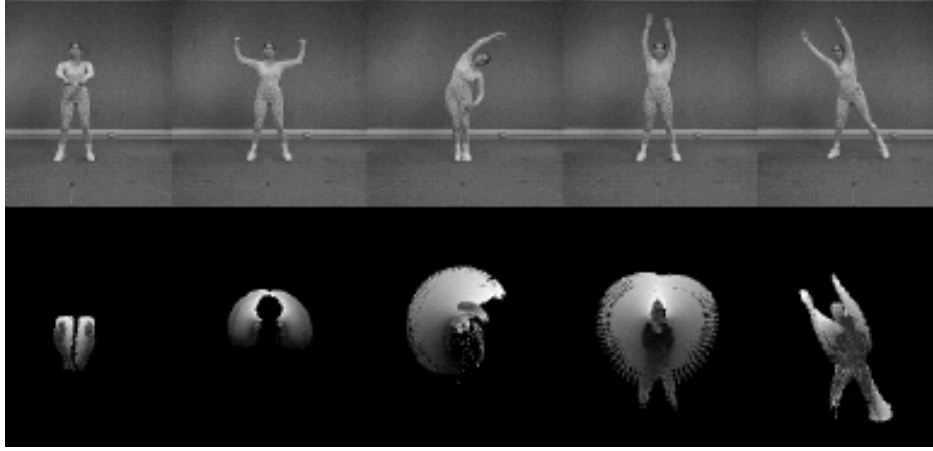


Figure 2.5: Motion History Images [17] for Action Recognition

However, the technique is view-sensitive, requiring the shapes of actions in the same category to be similar and the shapes of actions in different categories to be distinctive. In our domain, there is little constraint on the subject’s sleeping posture and the “shape” of breathing varies. Moreover, MEI and MHI are derived from DOF, which performs poorly in detecting breathing signals because the movements are so subtle that noise generated by the sensor are comparatively large (see section 3.2.1). Hence, Bobick and Davis [17] suggested that a more robust motion detection mechanism is required in situations where the test subject moves slowly.

In addition, a shortcoming of MHI is its lack of robustness against spatial motion self-occlusion occurring during the same temporal window due to overwriting. Valstar *et al.* [155] developed an extension of MHI – MMHI, which aims at handling motion self-occlusion, by recording motion history at multiple time intervals. However, their experimental results do not clearly demonstrate the superior performance of MMHI with respect to MHI.

Gorelick *et al.* [66] also uses spatiotemporal volumes for action recognition, seeing human action as silhouettes of a moving torso and protruding limbs undergoing articulated motion and computing the space-time saliency, and three types of space-time structures are defined, including plateness, stickness and ballness to represent actions; see Figure 2.6. The space-time shapes are able to discriminate between different actions like dancing, jumping or walking, but do not allow for detecting abnormalities of the same action. Albu and Beugeling [3] built a 3D extension of MHI – VMHI as motion representation to handle motion self-occlusion issue. However, as the VMHI is still built based on the DOF technique, VMHI is not suitable for

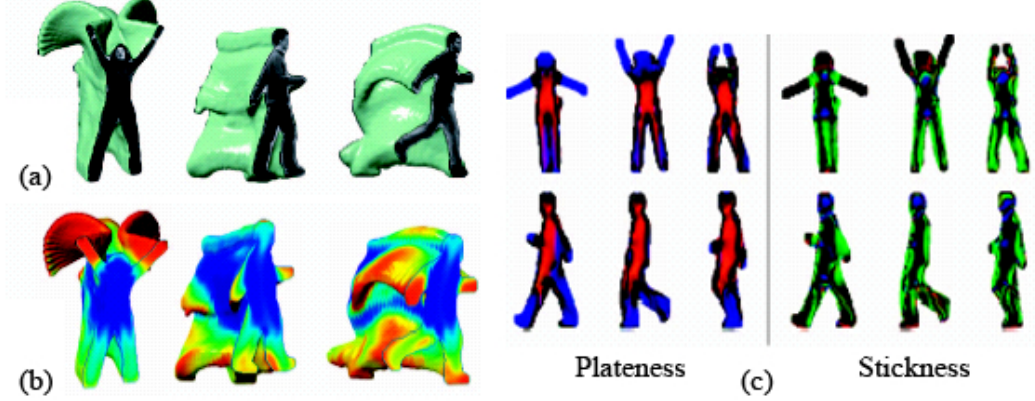


Figure 2.6: Action represented by Space-Time Shape Features [66]: (a) input; (b) degree of frequency; (c) plateness and stickness. Fast moving hands are identified as plates and appear in blue in (c); slow moving legs are identified as vertical sticks in the temporal direction and appear in green in (c); ball structure does not have any principal direction.

breathing monitoring; see section 3.2 for more details.

**Action Recognition by Tracking Space-Time Interest Points.** Another technique is to track space-time interest points [116] to generate spatial-temporal “words” (using the bag of words representation originally developed for text analysis domain). The geometric arrangement between visual features is ignored, and a histogram of the number of occurrences of particular visual patterns in a given image is computed to represent actions. In the context of human action classification, the bag of words assumption – the order of words in a text document can be neglected – translates into a video representation that ignores the positional arrangement, in space and time, of the spatial-temporal interest points. To represent motion patterns, Niebles *et al.* [116] first extract local space-time regions using the space-time interest point detector [42], and these local regions are then clustered into a set of spatial-temporal words, called codebook. Then, probability distributions are learned using Probabilistic Latent Semantic Analysis (pLSA) [71] or Latent Dirichlet Allocation (LDA) [16] to recognize and localize human action classes in video sequences.

Regions with spatially distinguishing characteristics undergoing a complex motion can induce a strong response to generate an interest point. However, this is unsuitable for our research, as generally the hospital cover/sheet

does not contain distinctive patterns. A patterned sheets could be used, but this makes the system less robust because if the cover is removed by the patient during sleep, the system fails.

## 2.3 Monitoring of Covered Body Activities

Recognition of covered human body activity is a challenging task. Some research work analyzes human activity without *a priori* shape models. In monitoring covered body activities, it is difficult to determine the activity based on pure motion because it is difficult to obtain full body silhouettes. Thus, the common model-based framework that identifies body pose first and analyzes the human activity using the estimated pose and detected motion is adopted here.

### 2.3.1 Pose Estimation Techniques

A number of methods, which might be considered for pose estimation of sleeping patients, are discussed here. Laser rangefinders are commonly used in 3D object geometry capture. A barrier to adoption of this technology is the safety for patients' eyes, as lasers can be dangerous. Although some laser rangefinders claim to be eye-safe, a technique must be thoroughly tested before it is applied to patients. Apart from safety issues, the technique may disturb the patient, the cost of laser rangefinders is high, and the processing time to reconstruct 3D geometry is substantial.

The pressure sensitive mattress is an alternative non-intrusive approach to identify occurrence of movements, and the technique has been proposed for monitoring patients' respiratory activities [105]. However, to the authors' best knowledge, the pressure sensitive mattress approach has not been utilized to analyze body movements. In this project, it is not adopted because of the high development cost, which also requires additional hardware instead of utilizing existing measurement equipment, such as a video monitoring system.

An alternative approach is to investigate imaging modalities that might see through the bed covering. X-ray is clearly too expensive and dangerous. A thermal imaging system [135] has been evaluated for obtaining the covered body posture in this work. It is found that, due to heat retention properties of the bed clothes, the thermal imaging system often fails to locate a true human posture because the heat tends to remain on the sheet or over the bed after the body posture has changed. Figure 2.7 illustrates the issue, showing thermal images with a leg movement. New technology of 3D camera such

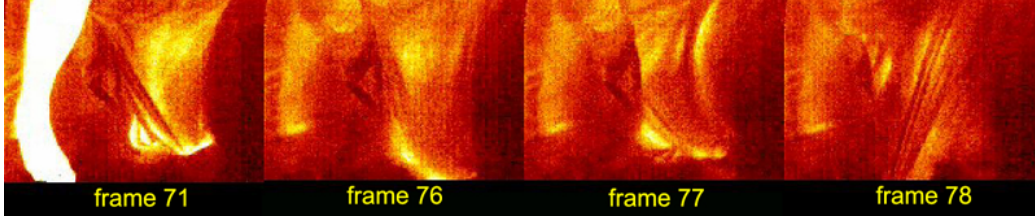


Figure 2.7: Thermal imaging fails to locate a true body posture because the heat remains on the bed after movements. #71: an un-occluded leg and a covered leg appear; #76: the image of the leg on the right indicates the covered leg’s position, but the ghost on the left indicates the previous location of the un-occluded leg; #77 and #78: strong noises appear due to the remaining heat on the bed after movements of the occluded leg

as Swiss Ranger, which is only recently introduced these two years, might be considered. However, as there was no such technique available in the beginning of the research project and the cost of the equipment is high, this technology is not adopted here.

This research addresses the problem of detecting and segmenting the covered human body using infrared vision. Difficulties arising from varying occlusion by the bedding, the shifting of the cover surface with movements, obscuration of the bodies’ edges by the cover, and wrinkle noises from the cover, are compounded by human articulated deformation. Traditional computer vision methods such as correlation, template matching, background subtraction, contour models and related techniques for object tracking become ineffective [22, 75] because of the large degree of occlusion for long periods.

### 2.3.2 Related Computer Vision Approaches

Popular human pose estimation algorithms utilize various tracking algorithms and sampling schemes such as mean field Monte Carlo [74] and annealed particle filter [40] in combination with simple detection models like rectangles of edges or motion. Assuming that there are clear image cues (for simple detection models) and clean full body motion data (for dynamical models in sampling), the subject pose can be estimated reliably in some systems. However, in real world applications, occlusion often occurs and these assumptions often fail. As a result, recognizing the pose of a person who is persistently under cover remains challenging.

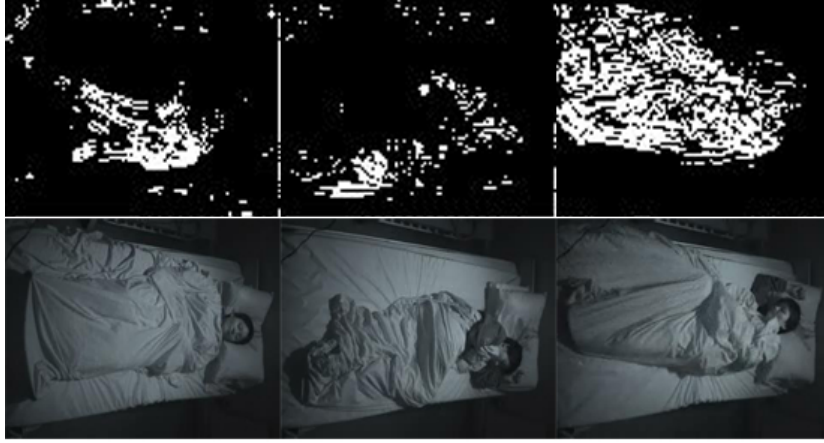


Figure 2.8: Fragmented Motion Data (upper column) and Raw Image (lower column): motion of the covered subject is fragmented and noisy as movement of the occluded subject also causes motion of the surface around rather than the exact area of the object, making object segmentation more challenging.

Many existing approaches to pose estimation simplify the measurement problem, either using motion data [1, 45, 67, 146] to extract silhouettes, or assuming knowledge of appearance or color [40, 95, 128, 136], and the subjects tend to wear close-fitting clothing (or even to be unclothed [40]) in order to extract such information more easily. These methods are too restrictive for this field of study. Although there is some published research investigating the monitoring of partially occluded humans [69, 129, 153, 170], the methods examined do not deal with pose estimation of consistently and almost wholly occluded subjects.

**Camouflaged Object Detection.** As the human is obscured by the bedding, to detect the pose of the subject under cover presents some similarity to camouflaged object detection. Camouflaged objects generally attempt to conceal themselves *within* the cover, and such targets are only visible while in motion. Boulton *et al.* [22] presented a surveillance system for perimeter security using adaptive multi-background modelling, temporal adaption and quasi-connected components techniques to detect the camouflaged targets. Similarly, Huang and Jiang [75] presented an iterative method of weighted region consolidation to track a camouflaged animal within an environment of similar colors. They first detect the full body motion of the target based on both spatial and intensity densities by locating pixels with high motion prob-

abilities, and then enhance the moving object iteratively, i.e. to consolidate the object region, by evaluating for each pixel its weighted overall neighborhood intensity based on the pixel distances and intensity. The contour of the object’s moving area is then constructed. However, in comparison with camouflaged objects, motion tends to be irregular, noisy and fragmented (see Figure 2.8) in our problem domain. Apart from partial and irregular movement, movement of the covered object causes motion of the surface around rather than the exact area of the object, making object segmentation more challenging.

**Dealing with Occlusion.** Current research [27, 31, 46, 62, 114, 168] for monitoring or tracking occluded human focuses on temporary rather than persistent occlusion. Jaeggli *et al.* [81] developed a learned statistical model to analyze human locomotion from a running or walking sequence when only a subset of the features of interest can be observed; the model is used to predict occluded features based on available features. However, the method requires 2D trajectories of a number of un-occluded locations on the human body, representing a period of specific types of actions, (i.e. running or walking), to initialize the model. This is not applicable in our case, because the body (barring the head) can be wholly covered, and un-occluded features can be fairly limited.

There is some research on human detection in crowded scenes, in which temporary and partial occlusion often occurs. Wu and Nevatia [170] combines edgelet based part detectors, including head-shoulder, torso and leg, and a full body detector to detect and track partially occluded humans. They learn tree structured multi-view part detectors by a boosting approach proposed by Huang *et al.* [76, 77], which is an enhanced version of Viola and Jones’s framework [157]. The real AdaBoost algorithm [138] is used to build each part detector, and they collect a large set of human samples, containing 1742 humans of frontal/rear view and 1120 side view, from which tree structured detectors for multi-view humans are learned. Liebe *et al.* [99] presented a framework to combine local and global cues for pedestrian detection in crowded scenes and utilize full body silhouettes for chamfer matching [18, 20]. By applying a Canny edge detector and distance transformation [19], they chamfer match the input silhouette area with 210 trained pedestrian silhouettes (plus their mirrored versions) to determine if a person is detected. However, the aforementioned approaches are used for human detection, and do not deal with pose estimation.

Ramanan *et al.* [128] presented an approach to both track people and to identify body poses on outdoor and indoor activity; the method can recover

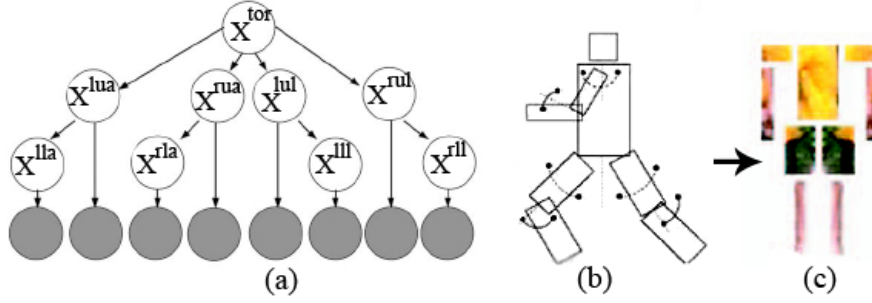


Figure 2.9: Person Model in [128] (a) the tree pictorial structure, (b) the edge template of the lateral walking pose pictorial structure, (c) a learned appearance template

when it loses track due to occlusion. The approach first builds an appearance model of each person in the video, then tracks by detecting those models in each frame. Two algorithms were developed to build appearance models: a bottom-up approach groups together candidate body parts found throughout a sequence; a top-down approach builds appearance models by detecting lateral walking poses, and the human body is modelled as an articulated set of rectangles, which is often called a pictorial structure [47, 48]; see Figure 2.9.

The bottom-up approach first detects candidate parts in each frame with an edge based part detector, clusters the resulting image patches to identify body parts that look similar across time, and then prunes clusters that move too fast in some frames. The clustering of part detectors works well when parts are reliably detected. However, building a reliable part detector is hard; a well-known drawback of bottom-up approaches. An alternative strategy is to look for an entire person in a single frame, but this is difficult because people are hard to detect due to variability in shape, pose, and clothing; a well-known drawback of top-down approaches. The top-down approach detects a lateral walking pose by convolving the distance-transformed edge image with a lateral walking pose edge template, as shown in Figure 2.9(b). The edge pixels are quantized into one of 12 orientations and the chamfer cost is computed separately for each orientation with the costs added together. As the target is covered by the bedding in this research, the appearance of body parts and the appearance of the cover are identical. Hence, appearance modelling is ineffective.

In this work, Ramanan’s method – the stylized pose detector– is tested on the covered human body sequences, but the results show that the method performs poorly on persistently-occluded subjects. The method and the ex-

perimental results are discussed in more detail in Chapter 4.

## 2.4 Conclusion

As the consequences of OSA can be fatal or lead to serious injury if left undiagnosed and untreated, there is a growing interest in alternative approaches to diagnosis as substitutes for labor-intensive and time-consuming PSG.

Video monitoring has been adopted to assist diagnosis of OSA, but is under developed due to its relative computational complexity. The current approach to video analysis is for clinicians to review substantial amounts of video data manually. As a result, there is a practical demand for automated methods that support OSA diagnosis from video. The two requirements are to monitor breathing and body movements.

Existing methods for action recognition using space-time shapes are insufficiently robust to analyze breathing activities. Another popular approach is to track body parts and then use the obtained motion trajectories to perform action recognition. Such approaches cannot be adopted for breathing monitoring in OSA analysis because prior robust identification of body parts is difficult to achieve and model-based approaches tend to be view dependent.

Traditional methods such as correlation, template matching, background subtraction, contour models and related techniques are ineffective because of the large degree of occlusion for long periods. Although there is some published research investigating the monitoring of partially occluded humans, the methods examined do not deal with pose estimation of consistently and almost wholly occluded subjects.

Having reviewed the action recognition and pose recognition literature, we conclude that no existing models are suitable for monitoring of breathing behavior or monitoring of human activity for persistently occluded subjects from video.

## Chapter 3

# Abnormal Breathing Detection

In this chapter, a new noninvasive real-time video monitoring technique is introduced for detecting abnormal breathing activities and assisting in diagnosis of OSA, using infrared video information. A novel motion model is presented to detect subtle and cyclical breathing signals from video, and an adaptive model of dynamic patterns is developed to construct an action template online. The classes of actions include normal breathing episodes, apnoea episodes, body movement episodes and deep breathing episodes. The proposed technique avoids imposing positional constraints on the patient, allowing patients to sleep on their back or side, with or without facing the camera, fully or partially occluded by the bed clothes. Furthermore, shallow and abdominal breathing patterns do not adversely affect the performance of the proposed approach.

The organization of the chapter is as follows. An analysis of breathing behavior is given in Section 3.1, followed by the discussion of potential relevant computer vision techniques in Section 3.2. The proposed algorithm is introduced in Section 3.3. Section 3.4 shows the experimental results on fifteen simulated video clips and four clinical video clips, which demonstrate that the model achieves high accuracy in recognizing abnormal breathing activities and other body movements. Section 3.5 concludes the chapter.

### 3.1 Analysis of Breathing Behavior

To recognize abnormal breathing activities, it is necessary to differentiate body movements (such as movements of the head, torso, arm or leg), from breathing activities, allowing further discrimination of normal and abnormal breathing status. Thus, we first analyze human breathing behavior in contrast to general body movement.

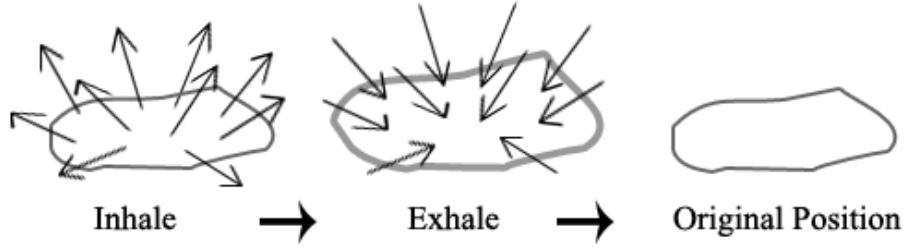


Figure 3.1: Cyclical Moving Flow in a Breath

**Subtle and Inconspicuous Movement.** Even for humans, it is difficult to observe human breathing movements in high resolution video sequences. Breathing is a subtle and relatively inconspicuous less noticeable movement compared to other body movements.

**Cyclical Movement.** Knowing that an object’s motion is periodic is a strong cue for object and action recognition [33]. A key characteristic of breathing activity is cyclical motion: the elements of the entire surface move forward and backward approximately to their previous position in a breathing cycle (see Figure 3.1). In contrast, elements tend to move toward different positions for general body movements.

Given the subtlety of breathing activities and the heavy level of motion self-occlusion, the proposed method needs to be capable of integrating minor movements across multiple frames of video.

**Changing Spatiotemporal Shape.** As the subject changes pose over time, the spatiotemporal shape of breathing movements changes, requiring the combination of adaptive patterns.

## 3.2 Related Work

While motion representation is identified by Moeslund and Granum [109] as an essential component of tracking, it can also serve higher-end purposes, such as activity recognition and abnormal event detection. Recent successful work in the area of action recognition [17, 43, 66, 88, 172] has shown that it is useful to analyze actions by treating a video sequence as a three dimensional space-time volume (of intensities, gradients, optical flow or other local features).

A crucial issue is how to choose the features to be employed in action recognition. Features may be based on only static cues (shape and appearance), or only dynamic cues (motion), or on both. Importantly, in our problem domain, breathing movements occur in different regions such as the jaw, the chest area, the abdominal area or the shoulders according to the individual breathing behavior and lying posture at the time, which makes the problem more challenging. In our experiments, we observe that some subjects tend to breathe with chest movements and others with tummy movements. In addition, when the subject lies on the side, breathing with shoulder movements often occurs. As there are large variances on individual's breathing behavior, the subject's appearance and the camera views to the subject, robust and unconstrained monitoring is required. Therefore, static cues do not appear as suitable choices. In this research, only dynamic cues (motion) are utilized for analyzing breathing patterns.

Three major issues are identified to investigate: motion capture of subtle and self-occluded breathing movements; motion quantization; adaptive action template model for dynamic breathing patterns and activity recognition. Relevant work for each issue is discussed below.

### 3.2.1 Motion Detection

#### 1 Difference of Frames

A generally adopted front end motion detection method is Difference of Frames (DOF) as formulated in Equation 3.1.

$$D(x, y, t) = |I(x, y, t) - I(x, y, t - k)| \quad (3.1)$$

where  $I(x, y, t)$  is the intensity of each pixel at location  $x, y$  at frame/time  $t$ ,  $k$  is the selected time interval among frames to compare, and  $D(x, y, t)$  is the difference of two frames representing pixels of motion. If  $k = 1$ ,  $D(x, y, t)$  is the difference of consecutive frames. A conventional method is to threshold the difference to produce a binary map.

$$B(x, y, t) = \begin{cases} 1 & \text{if } D(x, y, t) > \alpha \\ 0 & \text{otherwise} \end{cases} \quad (3.2)$$

where  $\alpha$  is a selected threshold.

DOF is commonly used as the front end method for motion detection and further motion models for activity recognition, such as spatiotemporal tem-

plates Motion History Image (MHI)  $T_1(x, y, t)$ , Motion Energy Image (MEI)  $T_2(x, y, t)$  [17] and Volumetric Motion History Image  $T_3(x, y, t)$  [3], which are described below.

## 2 Motion History Image Template

$$T_1(x, y, t) = \begin{cases} \tau & \text{if } B(x, y, t) = 1 \\ \max(0, T_1(x, y, t-1) - 1) & \text{otherwise} \end{cases} \quad (3.3)$$

where  $\tau$  is the length of the temporal window capturing the motion.  $T_1(x, y, t)$  is a function of motion history at  $(x, y)$  and the result is a scalar-valued image where more recently moving pixels are brighter.

## 3 Motion Energy Image Template

In contrast to MHI, MEI results in a binary image that highlights the regions in the image where any form of motion was present since the beginning of the action.

$$T_2(x, y, t) = \bigcup_{i=0}^{\tau-1} B(x, y, t-i) \quad (3.4)$$

where  $\tau$  is the specified length of the temporal window.

## 4 Volumetric Motion History Image Template

Volumetric Motion History Image (VMHI) [3] is a 3D extension of the MHI. Given sequences of binary silhouettes  $B$  as computed in Equation 3.2, VMHI is formulated as follows.

$$T_3(x, y, t) = \begin{cases} B(x, y, t) \Delta B(x, y, t+1) & \text{if } \text{cons}B(x, y, t) \neq \text{cons}B(x, y, t+1) \\ 1 & \text{otherwise} \end{cases} \quad (3.5)$$

where  $\text{cons}B(x, y, t)$  denotes the one pixel thick contour of the binary silhouette  $B(x, y, t)$ , and  $\Delta$  is the symmetric difference operator. (The symmetric difference of two sets  $S_1, S_2$  is the set of all  $x$  such that  $x \in S_1$  or  $x \in S_2$  but not both.)

The DOF technique was briefly evaluated for breathing detection, experimenting with different values of  $k$  (time interval) and  $\alpha$  (the pixel threshold). However, DOF fails to capture continuous motion of breathing due to occlusion by the bed cover, subtle changes and cyclical operations of breathing movements, and obtains a number of false positive detections from image noise, as illustrated in Figure 3.2. Only motion of large actions like deep breathing can be detected. Failure of the frontend motion capture approach renders further motion models like MHI, MEI or VMHI, feature extraction or activity recognition techniques ineffective.

Typical approaches for suppressing false positive detection are based on the aspect ratio, size, or magnitude of the frame-to-frame flow or normal flow. However, these approaches have been proven to be unsatisfactory in [169] because the frame to frame motion of the non-salient objects may be larger than that of the salient objects, especially if the salient object is moving relatively slowly. In our case, the salient object moves so slowly that noise generated by the sensor is comparatively large.

Optical flow is another common technique for motion detection, however, it is extremely computationally expensive. In the context of object detection, optical flow estimation requires hundreds or thousands of operations per pixel [157]. In addition, Lipton [101] indicated that most optical flow algorithms fail in largely homogeneous regions (i.e. regions lacking texture), and Gao *et al.* [58] also indicated that optical flow estimation in texture-less patches would be erroneous. As the subject is covered by the bed clothes, the regions of interest tend to be textureless.

Wixson [169] stated that an object that moves in a straight line but oscillates forwards and backwards, such as taking two steps forward and then one backward, would have low salience, and therefore optical flow is not suitable for detecting cyclical movements. Furthermore, optical flow approaches are highly susceptible to image noise [33], and video sensors in the infra-red band contain higher noise levels than in the visible band [133]. These weaknesses make optical flow unsuitable for this problem domain.

The first challenge in this research is to develop a suitable front end motion detection technique to capture continuous motion information of breathing activities, overcoming difficulties caused by heavy occlusion by the cover, inconspicuous frame to frame differences and motion self-occlusion. The proposed method is described in section 3.3.1.

### 3.2.2 Motion Quantization

The study of human motion from video sequences is mostly driven by applications in security. Some major surveillance-related themes address human

activity identification [43, 66], gait-based biometrics [21], object classification [33], and real-time abnormal event detection [58]. An emerging area of interest for vision-based human motion analysis is the field of perceptual human computer interfaces, addressing gesture recognition. The main goal in these applications is to detect and recognize either the motion, action or gesture performed by the human, or the human subject herself from her motion-based signature, and pre-trained motion descriptors or templates are commonly built for recognition.

The main goal of this chapter is not only to recognize the activity as a breathing action or a body action, but also to learn individual normal human breathing patterns online by quantifying and monitoring the subject's performance over time, and to detect the abnormal motion as a quantifiable deviation from the normal breathing patterns. However, to the best of the author's knowledge, motion analysis for performance quantification is still a fairly unexplored field in computer vision.

Cutler and Davis [33] introduce self-similarity matrices to detect and characterize the periodic motion for tracking and classifying objects. Given image sequences  $I_t(x, y)$ , they first produce a segmented foreground object  $O_t$  using motion, and compute the object's self-similarity  $S_{t_1, t_2}$  by comparing differences of images at time  $t_1$  and  $t_2$  within the bounding box  $B_{t_1}$  of the segmented object  $O_{t_1}$ ; see equation 3.6. They determine if an object exhibits periodicity based on the 1-D power spectrum of  $S_{t_1, t_2}$  for a fixed  $t_1$  and all values of  $t_2$ .

$$S_{t_1, t_2} = \sum_{(x, y) \in B_{t_1}} |I_{t_1}(x, y) - I_{t_2}(x, y)| \quad (3.6)$$

If the subject stays in the same position and thus we assume that the bounding box includes the entire image, the equation to generate self-similarity is equal to computing the degree of differences of frames with various time intervals  $k$ . However, in the previous section it has been found that DOF fails to detect breathing movements even with manipulation of  $k$  values. Similarly, VMHI [3] is inapplicable here due to failure of DOF.

### 3.2.3 Activity Recognition

In the specific problem domain of monitoring human breathing activities during sleep, one has to deal with heavy occlusion by the bed cover, high variability of appearance according to the occlusion status, large variances of human breathing behavior on areas of movements (chest movements versus abdominal movement), strength (shallow versus deep) and length of breathing periodicity, motion self-occlusion of breathing patterns and substantial

changes of camera views as the subject may sleep on his or her back or either side and face toward the camera or back toward the camera.

The aim of this chapter is to recognize and classify the motion event as a normal breathing event, a deep breathing event, an apnoea event or a body movement event. Due to the large variances in individual breathing patterns, subject appearance and camera view with respect to the subject, it is not practical to pre-train a general normal breathing template appropriate for everyone, and thus supervised approaches such as pre-training key pose templates [103] or motion descriptors [43] do not represent suitable options. Similarly, model-based approaches with *a priori* shape models [2, 130], focusing on the relative motion and prior identification of body parts, cannot be adopted because prior robust identification of body parts is difficult to achieve.

### 3.2.4 Online Training and Adaptive Action Template

Some recent research work has explored unsupervised model-free methods for motion analysis [66, 115, 116, 172]. These propose feature based methods, which extract space-time interest points from image sequences as a collection of spatial-temporal words to categorize human actions. Yilmaz and Shah [172] utilize a sequence of the points on the outer boundary of the object with respect to time to generate a spatiotemporal volume in  $(x, y, t)$ , differential geometric properties of which (e.g. peaks, pits, valleys and ridges) are used to generate actions descriptors. Gorelick *et al.* [66] extracts silhouettes of subjects over time to compute space-time saliency and orientation for action classification. Niebles and Li [115] use both static shape features by shape context [152] as well as space-time features by a space-time interest point detector [42] for human action categorization. In a later work [116], Niebles *et al.* abandon static shape features and extract space-time regions using only space-time interest points [42], and cluster these regions into a set of spatial-temporal words, called codebooks, to code actions. For the space-time interest point detector, any region with spatially distinguishing characteristics undergoing a complex motion can induce a strong response.

The aforementioned approaches do not deal well with persistent heavily occluded subjects because the point correspondence between consecutive frames or silhouette extraction are difficult and spatially distinguishing features are rarely available from occluded (texture-less) subjects in our problem domain. Furthermore, shapes of breathing actions appear different over time, requiring accompanying changes in the learned patterns. Therefore, there is a need for an adaptive approach to capture dynamic breathing patterns and deal with heavily occluded subjects.

### 3.3 Proposed Method

This section presents a new video monitoring approach for anomalous breathing behavior detection. Instead of using local features or static image signals such as edges or raw pixel values, the motion dynamics is selected as the fundamental cue providing robustness to large variances of poses and heavy occlusion; see Figure 2.4. Hence, no positional constraint on the patient is imposed (other than by the orientation and position of the bed), allowing patients to sleep on their back or side, with or without facing the camera. Moreover, no limitation on the level of occlusion is applied, and therefore the proposed technique also deals with the fully or partially occluded subject. Infrared video sensors are used to avoid disturbing the subject’s sleep.

The functionalities of the proposed model include: capturing continuous signals of breathing movements, continuously learning a normal breathing pattern online, detecting abnormal events when a deviation from the learned normal breathing patterns occurs, and classifying the activity as a body movement episode, a normal breathing episode, a deep breathing episode or an over-breathing episode (the latter is a typical event in the end of an apnoea episode and therefore is also referred to as an apnoea episode in this work).

#### 3.3.1 Motion Detector for Breathing Analysis

The first challenge in this research is to capture breathing movements; these are subtle and cyclical with complete motion self-occlusion, which makes them hard to detect. As the breathing movement is so subtle, the difference over consecutive frames for each pixel is so small that, if the DOF technique is used, the value of  $\alpha$  in Equation 3.2 must be decreased to such a small value (e.g.  $\alpha = 1$ ) to detect these differences, that noise detection becomes excessive particularly as infra-red sensors suffer from high noise levels [133]. In addition, the subject is occluded by the pattern-less hospital cover, which makes the problem more challenging. An illustration is given in Figure 3.2. A low frame rate of seven frames per second (fps) is used, as at a higher rate (15 fps), no difference can be detected. However, the frame rate must be sufficiently high to monitor the entire breathing cycle.

The design of the proposed motion detector is inspired by “the visual staying phenomenon” of human vision’. A persistent luminous impression is created, stored for a while and continuously but slowly updated. The persistent luminous impression is similar to the concept of background modelling. Background modelling has been mainly applied for foreground object segmentation rather than motion detection. As the output of background modelling

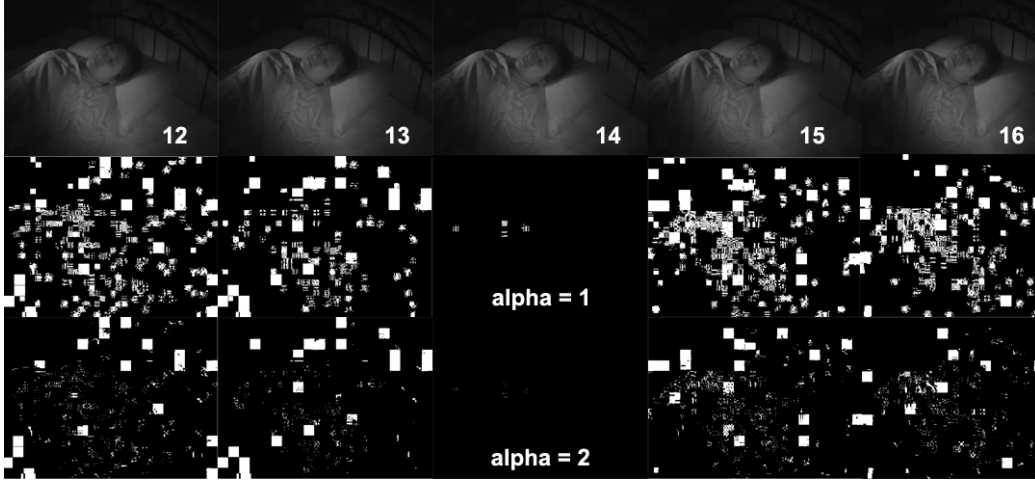


Figure 3.2: Top row: original images; middle row: DOF Results using  $\alpha = 1$ ; bottom row: DOF Results using  $\alpha = 2$ : with a low frame rate (7fps) and lowest possible  $\alpha$  value, DOF can detect breathing movement, but the output signal is swamped by noise.

is change detection rather than true motion detection, and is highly susceptible to non-motion pixel change due to noise and lighting changes, one effective method of foreground object extraction is to suppress the background points in the image frames [60]. Normally, the maintained background model is updated over time to avoid accumulated errors.

In contrast, a Persistent Luminous Impression Model (PLIM) is created in this research to extract real-time motion information rather than to segment the foreground object. The PLIM is designed to utilize accumulated errors to enhance breathing signals and to differentiate between breathing activity and body movement. Some existing background models are briefly discussed below, followed by the details of the proposed PLIM, motion detector and motion quantization models.

## 1 Existing Adaptive Background Models

A simple background model [106] checks whether each pixel remains unchanged for some time, and if a time threshold is exceeded, the observed pixel in the current frame is included into the background; otherwise the pixel is considered as either a moving object or noise. The background model  $I_B(x, y, t)$  is formulated below.

$$I_B(x, y, t) = \frac{p-c}{p}I_B(x, y, t-1) + \frac{c}{p}I(x, y, t) \quad (3.7)$$

where  $c$  is the number of consecutive frames during which a change is observed and is reset to zero each time the pixel becomes part of the background;  $p$  is the adaptation time or insertion delay constant. The moving object is extracted at pixels  $(x, y)$  for which the relationship  $|I(x, y, t) - I_B(x, y, t)| > L$  is satisfied, where  $L$  is the global threshold value.

The background model can also be updated periodically using a temporal median filter [142] with the result that all static objects are eventually incorporated into the background. In [98], the background models a mean  $I_m(x, y, t)^*$  and standard deviation  $I_\sigma(x, y, t)^*$  for each pixel:

$$I_m(x, y, t)^* = \alpha I_m(x, y, t-1) + (1-\alpha)I_m(x, y, t) \quad (3.8)$$

$$I_\sigma(x, y, t)^* = \alpha I_\sigma(x, y, t-1) + (1-\alpha)I_\sigma(x, y, t) \quad (3.9)$$

where  $0 < \alpha < 1$  defines the speed of adaptation, and if the difference between the input frame  $I(x, y, t)$  and the mean  $I_m(x, y, t)^*$  is large compared with  $I_\sigma(x, y, t)^*$ , the probability of classifying the pixel  $(x, y)$  as background is decreased.

There are more complicated background modelling methods that model each pixel as a mixture of Gaussians [147] or as a mixture of uniform distributions [11]. This multi-modal background representation is commonly used for backgrounds with frequent changes, such as outdoor applications with vegetation and illumination issues, or indoor applications with light reflection problems. Such methods tend to be computationally expensive. As in this research the environment is indoor and there is no reflection or changing illumination problem for infrared video sensors, multi-modal background modelling approaches are not needed.

## 2 Persistent Luminous Impression Model

The proposed Persistent Luminous Impression Model (PLIM) is designed to reinforce cyclical breathing motion signals spatially and body motion signals temporally, while suppressing video sensor noise, and is able to capture continuous real-time breathing signal as displayed in Figure 3.3. In addition, a motion quantization index is created based on this model.



Figure 3.3: Activity maps by the proposed PLIM, effectively rendering clean breathing movements across a breathing cycle.

At frame/time 0, the PLIM  $P(x, y, t)$  is initiated with the image values of frame 0.

$$P(x, y, 0) = I(x, y, 0) \quad (3.10)$$

At time  $t$ , the updated PLIM  $P(x, y, t)$  is given by:

$$\Delta(x, y, t) = I(x, y, t) - P(x, y, t - 1) \quad (3.11)$$

$$P(x, y, t) = P(x, y, t - 1) + \begin{cases} 1 & , \Delta(x, y, t) > 0 \\ 0 & , \Delta(x, y, t) = 0 \\ -1 & , \Delta(x, y, t) < 0 \end{cases} \quad (3.12)$$

### 3 PLIM Activity Map

The PLIM activity map  $A(x, y, t)$  is defined as:

$$A(x, y, t) = \begin{cases} 1 & \text{if } I(x, y, t) - P(x, y, t) > \alpha \\ 0 & \text{otherwise} \end{cases} \quad (3.13)$$

where  $\alpha$  is the detection threshold, a parameter of the model.

The PLIM plays an important role in monitoring of human breathing activities because it allows subtle and cyclical breathing motion signals to be detected, and makes the spatial and temporal magnitude of non-cyclical motion signals significantly larger than cyclical motion ones, allowing body movements to be distinguished from breathing episodes. The proposed PLIM activity map is able to extract clean motion data and monitor entire breathing cycles; see Figure 3.3.

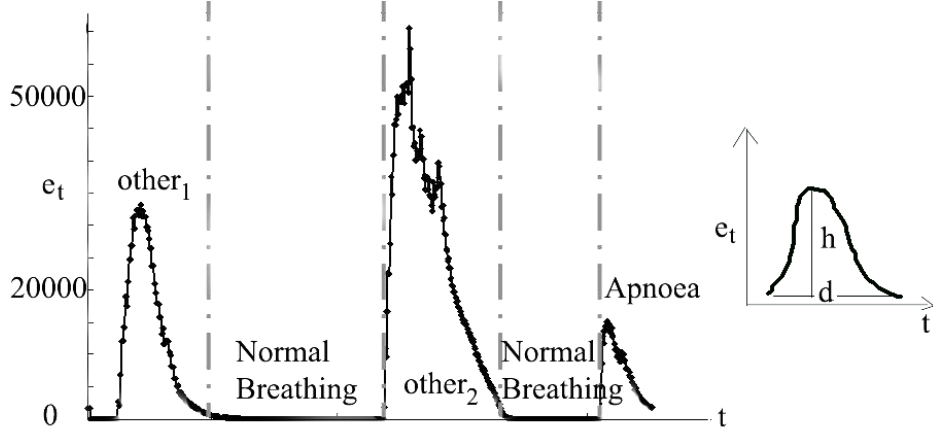


Figure 3.4:  $e_t$  spectrum and characteristics for action recognition: the duration  $d$  and peak value  $h$  of each hill shape can be used to distinguish breathing action from large body action like body rotation (see section 3.3.5 for a simple action recognition model). However, small body movements can still be confused with apnoea episodes, and a more sophisticated action recognition model is introduced.

#### 4 Motion Quantization

An activity index,  $e_t$ , is defined for motion quantization, as the number of set pixels in the activity map at time  $t$ :

$$e_t = \#(A(x, y, t) > 0) \quad (3.14)$$

An action is characterized by the shape of the spectrum of  $e_t$  values. Figure 3.4 illustrates a period of  $e_t$  values in an experimental video data. The duration,  $d$ , and the maximum value,  $h$  of  $e_t$ , for each hill shape are used as criteria to distinguish breathing motion from other motion. The  $d$  and  $h$  for an abnormal breathing event such as an apnoea episode is relatively small compared to other body movements because of the subtle trajectory of breathing motion, causing a small difference (small  $h$ ), and the cyclical breathing motion, reducing the duration of motion (small  $d$ ) due to the convergence of the current frame and the PLIM. In addition, from the shape of the spectrum, we can not only recognize  $other_2$  as a body movement, but also further identify  $other_2$  as a body rotation event rather than movement of a body part such as the head or the arm.

However, a simple action recognition model based on the analysis of the  $e_t$

spectrum cannot be fully relied on to distinguish abnormal breathing events from body movements, because small movements like slight head movements can cause similar  $e_t$  events to apnoea episodes. As a result, a more sophisticated online training model for building templates of normal breathing activities in real-time is further introduced in the next section.

### 3.3.2 Adaptive Action Template to Capture Normal Breathing Patterns

There are four statuses defined in the proposed model: *normal breathing* status, *deep breathing* status, *apnoea* status and *other body movement* status. While the status is “normal breathing”, according to the level of  $e_t$ , an adaptive action template is built to represent dynamic forms of normal breathing activities. The normal breathing action template is modified over time until an abnormal event occurs (high  $e_t$  detected). This novel technique is proposed to both assist classification of the current movement as a breathing action event or a body movement episode, and to classify the breathing activity as an extreme apnoea episode, a moderate deep breathing episode or a normal breathing episode.

## 1 Temporal Aggregation of Spatial Shapes

As shown in Figure 3.5, the real motion produced by breathing movements during normal breathing periods is comparatively low, and therefore the informative features for each spatial action shape are limited. Hence, the initial simple design is to aggregate spatiotemporal features by combining 2D motion shapes within normal breathing periods, and produce a spatiotemporal shape of normal breathing activities.

$$T(x, y, t_s, t_e) = \bigcup_{t=t_s}^{t_e} A(x, y, t) \quad (3.15)$$

where  $A(x, y, t)$  is the activity map defined in equation 3.13,  $t_s$  is the detected beginning of a period of continuous normal breathing cycles,  $t_e$  is the detected end of a period of continuous normal breathing cycles, and  $t_s, t_e$  are formulated in section 3.3.5; see Figure 3.5.

Although the Motion Detector using PLIM effectively filters out most of the sensor noise and captures real body movements, the level of noise in our clinical experimental data is still problematic; see Figure 3.6. The

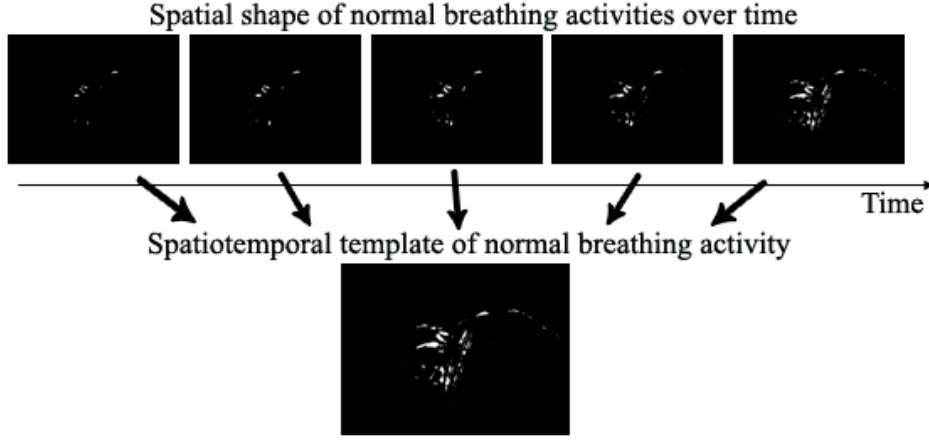


Figure 3.5: Normal Breathing Template Construction: a simple temporal aggregation model to combine spatial shapes for a continuous normal breathing period.

simple approach to template construction described above is susceptible to corruption by sensor noise and fails to generate an action template. A more sophisticated approach is described in the next section.

## 2 Adaptive Construction of Dynamic Patterns

As the subject tends to change his/her pose during sleep, the spatiotemporal shape of breathing actions change over time too. Therefore, the template model is adaptively modified over time to capture the dynamic patterns.

Firstly, a blank template is created, and when the status switches to *normal breathing*, the adaptive construction is triggered to update the template and proceeds until the status changes to any other status. Moreover, if the abnormal episode is a breathing event, the template is retained and is used for adaptive construction when the status changes back to *normal breathing*. On the other hand, if the abnormal episode is an *other body movement*, the template is discarded since the shapes of breathing action change in the mean time, and a new template is created when next entering *normal breathing* status. In contrast, for breathing events, the subject will return to his/her original position as analyzed in section 3.1.

The adaptive action template construction algorithm needs to not only capture limited and changing shapes of breathing activities, but also identify redundant information and discard noise. Figure 3.6 illustrates the captured

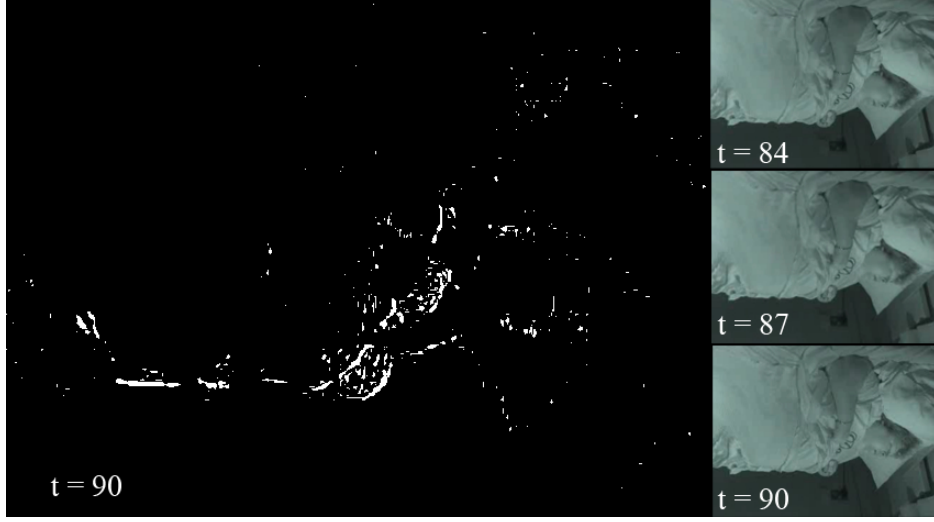


Figure 3.6: Motion captured with sensor noise: a simple temporal aggregation model suffers from high level of infrared sensor noise, and therefore a more sophisticated adaptive template model is introduced.

motion with some noise. As noise is random and less repetitious than breathing signals, signals appearing once are not included in the template, while repeating signals are retained. Signals that stop repeating are discarded gradually, and the template is updated with the newest repeating signals. Furthermore, if the data in the template is insufficient ( $q < \lambda$ :  $q$  is the template quality index, and  $\lambda$  is the template validity threshold criterion described in the next section), all data is retained as it is more important to accumulate data than to avoid noise. An adaptive action template in a clinical video sequence is shown in Figure 3.7.

A template with gradient values,  $T_g(x, y, t)$ , is built with the value of individual data point scaled to the range  $[0, 255]$ , and the final template is  $T(x, y, t)$ .

$$T_g(x, y, t) = \begin{cases} 255 & \text{if } A(x, y, t) = 1 \wedge T_g(x, y, t-1) > 0 \\ \delta & \text{if } A(x, y, t) = 1 \wedge T_g(x, y, t-1) = 0 \\ T_g(x, y, t-1) - \epsilon & \text{if } A(x, y, t) = 0 \wedge q_{t-1} > \lambda \wedge T_g(x, y, t-1) > 0 \\ 0 & \text{otherwise} \end{cases} \quad (3.16)$$

$$T(x, y, t) = \begin{cases} 1 & \text{if } T_g(x, y, t) > \delta \\ 0 & \text{otherwise} \end{cases} \quad (3.17)$$

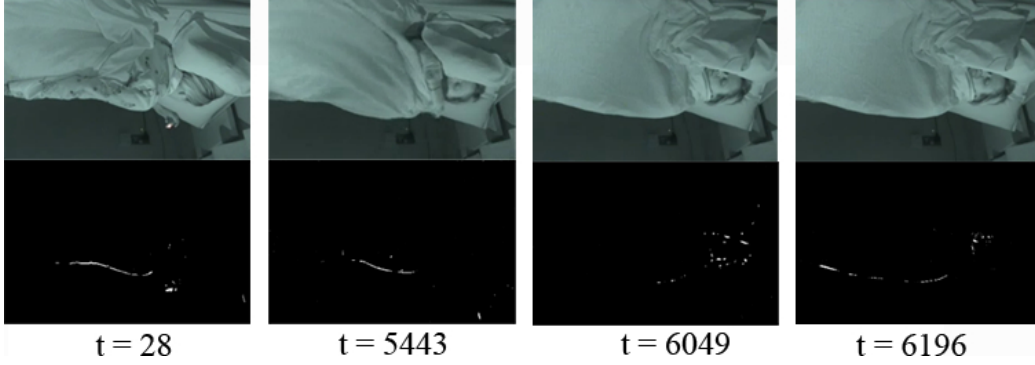


Figure 3.7: An adaptive action template in a clinical video data: as the normal breathing template adapts over time, the template at four different times, appearing differently, is selected for illustration purpose.

where  $\delta = 100, \epsilon = 4, \lambda = 0.00117WH$  are empirically determined. The template quality threshold  $\lambda$  is also ultimately used to determine whether the template contains sufficient information to be used for action classification.

### 3 Evaluation of Constructed Dynamic Templates

When the subject has a shallow breathing pattern, the action template generated has few features and provides a poor reference standard. A template quality index  $q$  is used to evaluate the effectiveness of a constructed template  $T(x, y, t)$  to determine if the template is usable for further action recognition;  $q$  being the number of features in the template:

$$q = \#(T(x, y, t) > 0) \quad (3.18)$$

If  $q < \lambda$ , the template is discarded ( $\lambda = 0.00117WH$  is determined empirically, where video frames were acquired with resolution of  $W \times H$ ), and the system automatically switches to an alternative novel action recognition model, which is described in section 3.3.5.

#### 3.3.3 State Algorithm for Action Segmentation

Although normal breathing may be barely perceptible in the  $e_t$  spectrum, motion events manifest as observable perturbations; see Figure 3.4. We segment motion events by identifying the start and end time,  $t_s$  and  $t_e$ , where the activity index  $e_t$  rises above and subsequently falls below thresholds,

described below. The sequences between motion events, where the activity index is very low, correspond to periods of normal breathing. We therefore use a two state algorithm, which switches between the *normal breathing state* and the *motion event state*. It is possible to classify motion events using only the duration and peak values from the corresponding section of the spectrum, but this is insufficient to distinguish some movements (e.g. slight head movements) from apnoea episodes. A more sophisticated approach using online breathing templates is introduced in the next section.

### 3.3.4 Action Recognition by Template Matching

When a large movement other than a normal breathing action is detected according to the  $e_t$  value (see section 3.3.5), the status is switched from a normal breathing episode to an unknown episode, the online construction process for the normal breathing template is terminated, and action recognition models are activated to recognize the new action. If the normal breathing template is usable according to the template quality index  $q$ , a template matching model is executed for action recognition.

A template matching model (cwMatch) is introduced below to classify the unknown action, using the recent normal breathing template, as a "Deep Breathing", "Apnoea" or "Body Movement" event. Below, we first discuss shape comparison methods and the reason why such techniques are inapplicable to this problem, and then introduce cwMatch.

## 1 Shape Comparison is Not Usable

Shape comparison has been used to classify distinctive actions by measuring similarity (or dissimilarity) between shapes of actions. Existing methods to measure similarity matrices include normal cross-correlation [65], the Hausdorff distance [14, 79], color indexing [12] and absolute correlation, as presented in equation 3.6. Cross-correlation gives a more robust measure of the correspondence between two matrices than absolute correlation. However, due to the high computational cost, normal cross-correlation is rarely adopted [117]. The Hausdorff distance measures the degree of mismatch between two sets  $(\mathcal{A}, \mathcal{B})$  by measuring the distance of the point of  $\mathcal{A}$  that is farthest from any point of  $\mathcal{B}$  and vice versa; unlike most methods of comparing shapes there is no explicit pairing of points of  $\mathcal{A}$  with points of  $\mathcal{B}$ . The time complexity of a Hausdorff distance measurements for two point sets of size  $p$  and  $q$  is  $O((p+q)\log(p+q))$ [79]. Regarding the color indexing technique, it is not suitable for night vision applications.

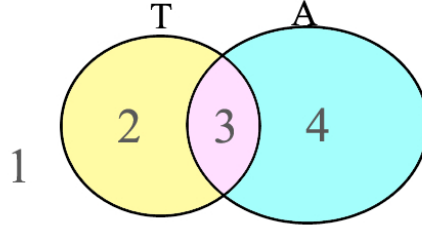


Figure 3.8: Various measurements ( $T$ : template and  $A$  current activity). similarity ( $T \wedge A$ ) using the overlap 3; similarity ratio ( $\frac{T \wedge A}{T \vee A} = \frac{3}{2+3+4}$ ); dissimilarity ( $\sim T \wedge A = 4$ ) and ( $T \wedge \sim A = 2$ ).

Table 3.1: Measurements for Abnormal Events

Abnormal Events	DBreath	Apnoea	Body(S)	Body(M)	Body(L)
Similarity $T \wedge A$	$\uparrow$	$\uparrow$	$\downarrow$	$\downarrow$	$\downarrow$
Dissimilarity $\sim T \wedge A$	$\downarrow$	$\uparrow$	$\uparrow$	$\uparrow$	$\uparrow\uparrow$
Dissimilarity $T \wedge \sim A$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$
Similarity Ratio $\frac{T \wedge A}{T \vee A}$	$\frac{\uparrow}{\uparrow+\downarrow+\downarrow}$	$\frac{\uparrow}{\uparrow+\uparrow+\downarrow}$	$\frac{\downarrow}{\downarrow+\uparrow+\uparrow}$	$\frac{\downarrow}{\downarrow+\uparrow+\uparrow}$	$\frac{\downarrow}{\downarrow+\uparrow+\uparrow}$
cwMatch: $w_1 = \frac{T \wedge A}{T}$	$\uparrow$	$\uparrow\uparrow$	$\downarrow$	$\downarrow$	$\downarrow$
cwMatch: $w_2 = \frac{\sim T \wedge A}{A}$	$\downarrow\downarrow$	$\uparrow$	$\uparrow$	$\uparrow$	$\uparrow\uparrow$
cwMatch: $s = \frac{w_2}{w_1}$	$\downarrow\downarrow$	$\downarrow$	$\uparrow$	$\uparrow$	$\uparrow\uparrow$

$T$ : action template of normal breathing;  $A$ : current movement; DBreath: Deep breathing episode; Body(S): minor body episode such as head movement; Body(M): medium body episode; Body(L): large body episode such as body rotation.

The aim here is to recognize the action rather than to simply measure the similarity between shapes, and the shapes between different action classes (e.g. “Deep Breathing” and “Apnoea”) may not appear markedly different. To illustrate that shape comparison is insufficient for this problem domain, an illustration is given in Figure 3.8 with Table 3.1 listing the orientations of different measurements, including a similarity measurement between the action template  $T$  and current activity  $A$ , two dissimilarity measurements and a measurement of the similarity degree, for five action types. It shows that action classification cannot be based on shape comparison measurements either using the similarity (as in correlation approaches) or using the dissimilarity (as in Hausdorff distance methods) because different actions are confused. We tested the degree of similarity, and using this measurement alone, there were many cases of confusion between “Deep breathing” and “Apnoea” and between “Body Movement” and “Apnoea”. In addition, the value fluctuates a lot for the same action type in individual video clip and among different video clips.

## 2 A Template Matching Model

A template matching method, `cwMatch`, is introduced here. Importantly, the output matching score  $s$  is designed to enhance the discriminative power of the action classifier. In our experiments,  $s$  is consistent over time for individual action classes and across different subjects and environmental settings (e.g. camera view angles and illuminations), and the values of  $s$  for the three action classes – “Deep Breathing”, “Apnoea” and “Body Movement” are distinct.

Given the latest normal breathing template trained  $T$ , and the current activity map  $A$  at time  $t$ , the template matching score  $s$  is formulated as the ratio  $s = w_2/w_1$ , where  $w_1$  is the ratio of the overlap between the action template and the current activity to the action template, and  $w_2$  is the ratio of the current activity not on the template to the entire current activity (see equation 3.19, 3.20 and 3.21). The matching score  $s$  is then utilized to determine the type of current event.

$$w_1 = \frac{\#(T \wedge A)}{\#T} \quad (3.19)$$

$$w_2 = \frac{\#(\sim T \wedge A)}{\#A} \quad (3.20)$$

where  $\sim T$  is the complement of  $T$ .

$$s = w_2/w_1 \quad (3.21)$$

The action is determined by the category  $s$  falls into. Two parameters  $\gamma_1, \gamma_2$  are used to separate “Deep breathing”, “Apnoea” and “Body Movement” events.

$$action = \begin{cases} o_1 & \text{if } s \geq \gamma_1 \\ o_2 & \text{if } \gamma_2 \leq s < \gamma_1 \\ o_3 & \text{if } s < \gamma_2 \end{cases} \quad (3.22)$$

where  $o_1$  represents a body movement event;  $o_2$  represents an apnoea event;  $o_3$  represents a deep breathing event;  $\gamma_1, \gamma_2$  are defined empirically ( $\gamma_1 = 0.03, \gamma_2 = 0.004$ ).

In this model,  $w_1$  represents the degree of the cyclical movement detected whereas  $w_2$  represents the degree of the non-recurrent movement. Table 3.1 lists rough scales of  $w_1, w_2, s$  for five action types, showing that  $s$  has different values for the three action classes–“Deep Breathing”, “Apnoea” and “Body Movement”. This model is insensitive to the parameter values. In addition, the time complexity of the method is  $\mathcal{O}(p)$  where  $p \geq q$ , which is much more efficient than cross-correlation or Hausdorff distance measurements.

### 3.3.5 Parameter Definition and Deviation Detector

This section defines the model parameters for status switching and the simple action recognition model. Furthermore, an alternative approach to eliminate the occurrences of other body movements when the breathing template is invalid (see Equation 3.18), is presented here.

#### 1 State Transition Rules

For set  $T$  and natural number  $n$ , define the order function  $\mathcal{O}(T, n)$  to be the  $n^{th}$  smallest element of  $T$ :

$$\mathcal{O}(T, n) = t_i : t_i \in T \wedge \#\{t : t \in T, t \leq t_i\} = n \quad (3.23)$$

The algorithm enters the normal breathing state, at time  $t_s$ , when the activity index  $e_t$  drops below the activity threshold  $\theta_s$  for  $n = 10$  (not necessarily consecutive) time-steps; see figure 3.9:

$$T_s = \{t : e_t \leq \theta_s \wedge t > t_m\} \quad (3.24)$$

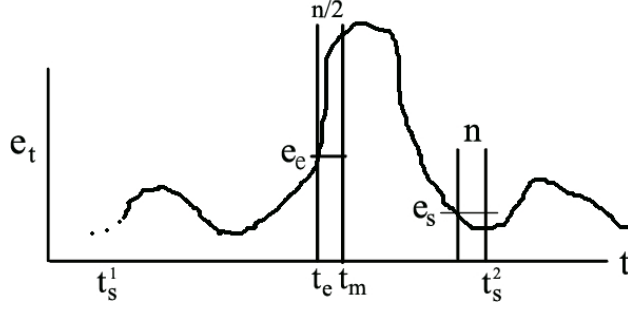


Figure 3.9: Illustration of the Important State Switching Points:  $t_s^1, t_s^2$  are the previous and new starting points to construct templates of normal breathing action;  $t_e$  is the time to terminate the construction process;  $t_m$  is the time to process template matching by comparing the action template  $T(x, y, t_s^1, t_e)$  and the spatial shape  $A(x, y, t_m)$  of the action at time  $t_m$ .

$$t_s^{new} = \mathcal{O}(T_s, n) \quad (3.25)$$

where  $t_m$ , representing the time at which the last motion event was analyzed, is initialized to zero and subsequently updated as below. The threshold  $\theta_s = 0.0033WH$  is determined empirically.

The algorithm moves from the normal breathing state to the motion event state at time  $t_e$ , when the activity index  $e_t$  exceeds the adaptive threshold  $\theta_e$ . This is initialized to  $\theta_e = 2.2\lambda$  at frame 0 and when a new template is constructed after a body movement, and then updated every  $25^{th}$  frame using:

$$\theta_e = \max(\nu q_t, 2.2\lambda) \quad (3.26)$$

where  $\nu = 1.3$  is empirically determined. Thus,  $\theta_e$  detects a significant rise (30%) above the periodically sampled expression of the online normal breathing template, with a minimum threshold of  $2.2\lambda$  to deal with the case of shallow breathing. Thus, the normal breathing state terminates when a sudden increase in activity with a rapid deviation from the learnt cyclical breathing patterns is detected.

The motion event is evaluated at time  $t_m$  on the fifth time step with the activity index above the threshold after the motion event state is entered:

$$T_m = \{t : e_t \geq \theta_e \wedge t > t_e\} \quad (3.27)$$

$$t_m = \mathcal{O}(T_m, n/2) \quad (3.28)$$

## 2 Simple Action Recognition Model

When the normal breathing patterns are insufficient for template matching (e.g. due to shallow breathing),  $q_t < \lambda$ , a simple alternative approach is used to identify body movements and apnoea episodes, based on the duration  $d$  of the episode and the activity index  $e_t$ . This model classifies abnormal events into three categories:  $o_1$  represents a body movement event;  $o_2$  represents an apnoea event;  $o_3$  represents a deep breathing event. The characteristics of  $e_t$  spectrum for other body movement events include relatively high peak values and long duration. Hence, the model is formulated as follows.

$$d = t_s^{new} - t_e^{old} \quad (3.29)$$

$$action = \begin{cases} o_1 & \text{if } e_{t_m} \geq \theta_m \vee d \geq \theta_d \\ o_2 & \text{if } d \geq \frac{\theta_d}{2} \\ o_3 & \text{otherwise} \end{cases} \quad (3.30)$$

where thresholds  $\theta_m, \theta_d$  are formulated as follows.

$$\theta_m = \kappa W H \quad (3.31)$$

$$\theta_d = \beta F \quad (3.32)$$

where video frames were acquired at  $F$  frames per second, with resolution of  $W \times H$  ( $\kappa = 0.26, \beta = 4.6$  are defined empirically).

## 3.4 Evaluation

The proposed action recognition technique is used to identify normal breathing episodes, over-breathing episodes and body movement episodes. Over-breathing episodes occur at the end of every apnoea episode, and consequently the proposed method uses over-breathing actions to identify apnoea episodes. Body movement is used as an indicator of waking-up by clinicians to assess sleeping quality. If a body movement follows an over-breathing event, it supplies additional evidence of an apnoea episode. Consequently, the

evaluation of the proposed technique is based on detection of over-breathing episodes, whether or not they are followed by the body movement episodes. However, body movements without over-breathing are not classified as apnoea episodes.

### 3.4.1 Experiments on Simulated Data

The simulated data allows us to test a number of scenario with various occlusion levels, body poses, body movements (e.g. minor head movement, limb movement, body rotation and slight torso movement), breathing behavior (e.g. shallow vs. heavy breathing, mouth breathing, chest breathing, and abdominal breathing) and sequences of linking events (e.g. apnoea–body movement and body movement–apnoea). To evaluate the proposed method, 15 video clips were filmed, each containing simulated apnoea episodes and other movements.

**Experimental Setup.** Two SONY infrared camcorders (DCR-HC-30E) were utilized, with three different shooting angles, at 15 frames per second and a resolution of  $320 \times 240$ . The video data was first captured with the WMP9 compression algorithm, to minimize storage size, and then decompressed for off-line analysis. In order to simulate the environment for diagnosis of sleeping disorders, there was no visible lighting in the filming room and the subjects were covered by a sheet. The experimental data was collected from two subjects with three main postures (i.e. lying on the back, sleeping on one side and facing the camera, sleeping on the other side with their back facing the camera). The data was also collected on different days, from multiple camera positions, with the subjects wearing different clothing. Activities, such as normal breathing, obstructive apnoea and body movement, were simulated by the subjects. Furthermore, one of the subjects has shallow breathing patterns.

To produce a reference standard, the experimental video contents were manually marked by human observation to define the events including the frame numbers of the beginning and end of each event. That is, the observer specifically identifies the start and end frame of abnormal movements, including apnoea episodes and body movement episodes. In addition, deep breathing action is marked as a normal breathing activity and not regarded as an abnormal action episode. The manually analyzed results are then used to compare with the outputs of the proposed method.

**Experimental Results.** The results of the method are lists of episodes with the associated beginning and end frame numbers for individual abnormal events—namely apnoea and body movement. These episodes are compared to the reference standard generated by human observation. Figure 3.10 and Figure 3.11 present the experimental results of the proposed method and manually observation on all fifteen video clips. The use of the stabilizing factor  $n$ , designed for switching status from “not normal breathing” to “normal breathing”, as illustrated in Figure 3.9, causes a fairly consistent “overshoot” of duration  $\approx n$ , at the ends of abnormal events. This is reasonable and consistent with the design of the proposed model.

Figure 3.12 shows the quantitative classification results in the form of a confusion matrix [90], which has been used to evaluate action recognition methods in recent research work [43, 100, 116]. The rows are ground truth; the columns are the proposed model results; the main diagonal shows the fraction of frames correctly classified for each class, and each row represents the probabilities of that class being confused with all the other classes. The results show that the diagonal average of the confusion matrix is 0.955, demonstrating that the method achieves high accuracy in recognizing abnormal breathing activities and body movements.

The method misses apnoea episodes occurring right after an other body movement episode, as shown in video clip 8 it treats the two episodes as one, since action recognition of abnormal events is computed based on the first activity, and after an abnormal action is recognized, the method searches for the next normal breathing episode instead of conducting further action recognition. On the other hand, if minor movements happen right after an apnoea episode (e.g. in video clip 11 and frequently occurring in the clinical data), the proposed method recognizes the entire session as an apnoea episode. In such cases, the human observer also defines the entire session as an apnoea episode.

According to Matusiewicz, MD [107], body movement, which indicates waking-up, is unlikely to happen just before an apnoea event because apnoea does not occur when the patient is awake. On the contrary, apnoea makes patients wake up unconsciously and may trigger a body movement directly after the apnoea event. In other words, the weakness of the system illustrated in clip 8 should not occur in clinical practice.

Figure 3.13 and Figure 3.14 illustrate the analysis results of three video samples. In the upper area, the spectrum of spatiotemporal motion quantization values  $e_t$  is presented with the episode number marked; the lower area shows the action detected and overall working flows. There are 7 episodes in Figure 3.13. Episode 1 is recognized as a normal breathing period, and during this period a spatiotemporal normal breathing template is continu-

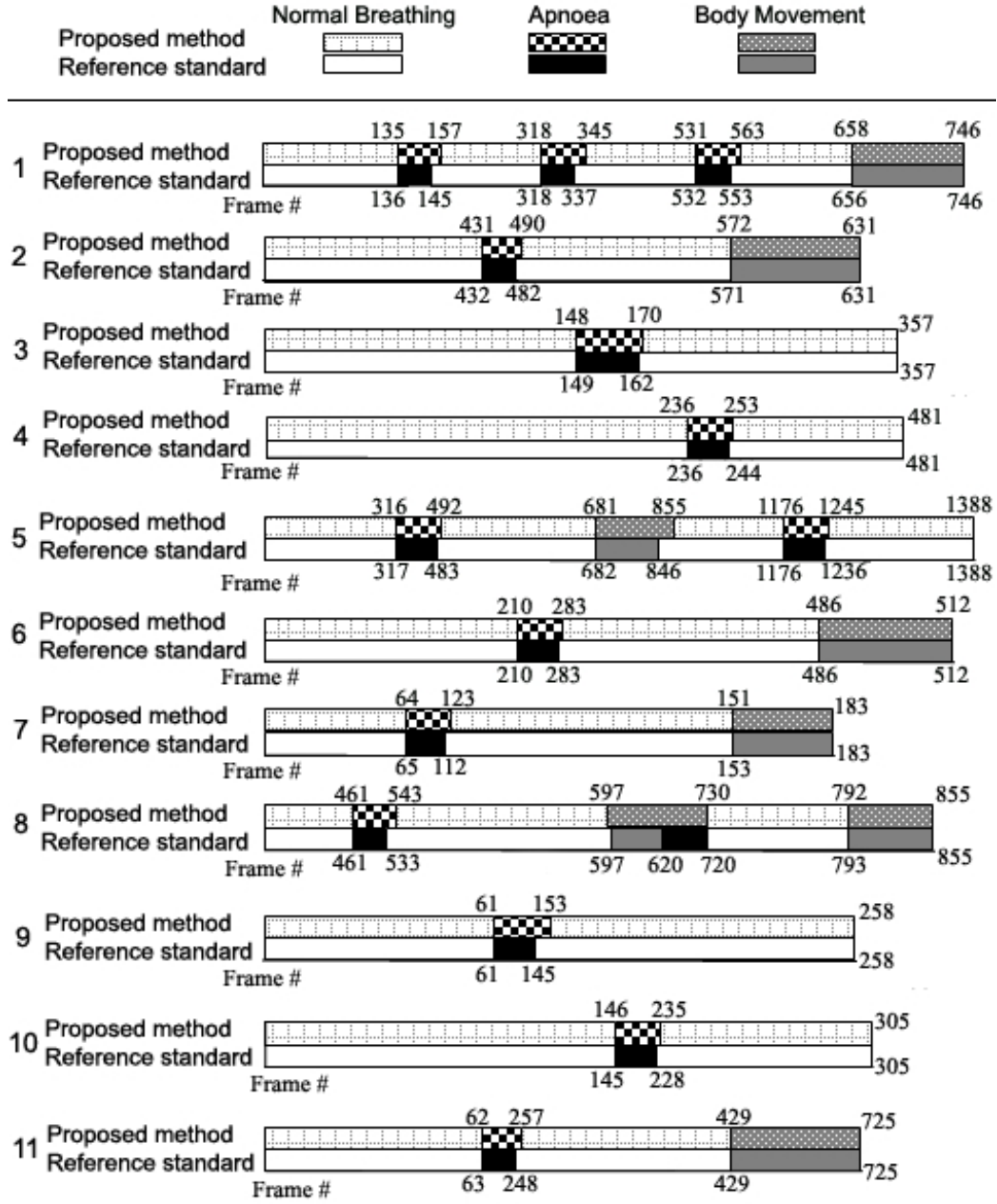


Figure 3.10: Experimental Results of 11 simulated video clips: constant differences occur at the end points of individual abnormal events between the proposed method and the reference standard because of the stabilizing factor  $n$  designed for switching status from “not normal breathing ”to “normal breathing ”.

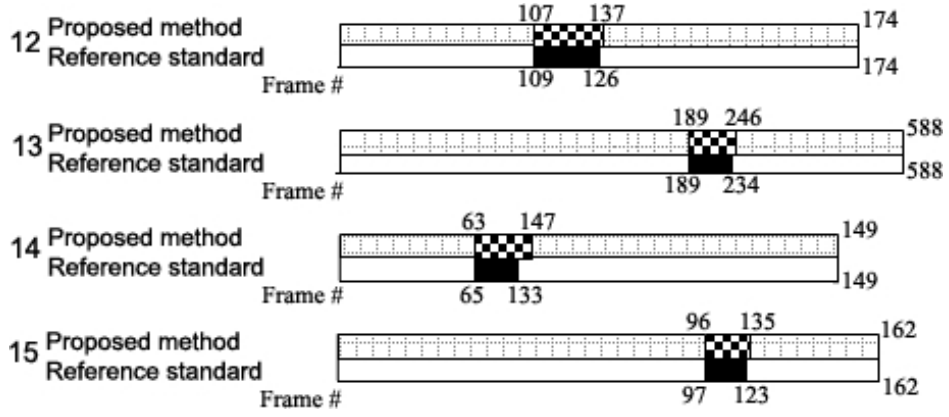


Figure 3.11: Experimental Results of four simulated video clips: constant differences occur at the end points of individual abnormal events between the proposed method and the reference standard because of the stabilizing factor  $n$  designed for switching status from “not normal breathing” to “normal breathing”.

Diagonal Average: 0.955

Reference standard	Normal Breathing	.964	.032	.004
	Apnoea	.001	.914	.085
	Body Movement	.012	0	.988
		Normal	Apnoea	Body
Estimation of Proposed Method				

Figure 3.12: Confusion Matrix of Action Classification on Simulated Data: the rows are ground truth; the columns are the proposed model results; the main diagonal shows the fraction of frames correctly classified for each class, and each row represents the probabilities of that class being confused with all the other classes.

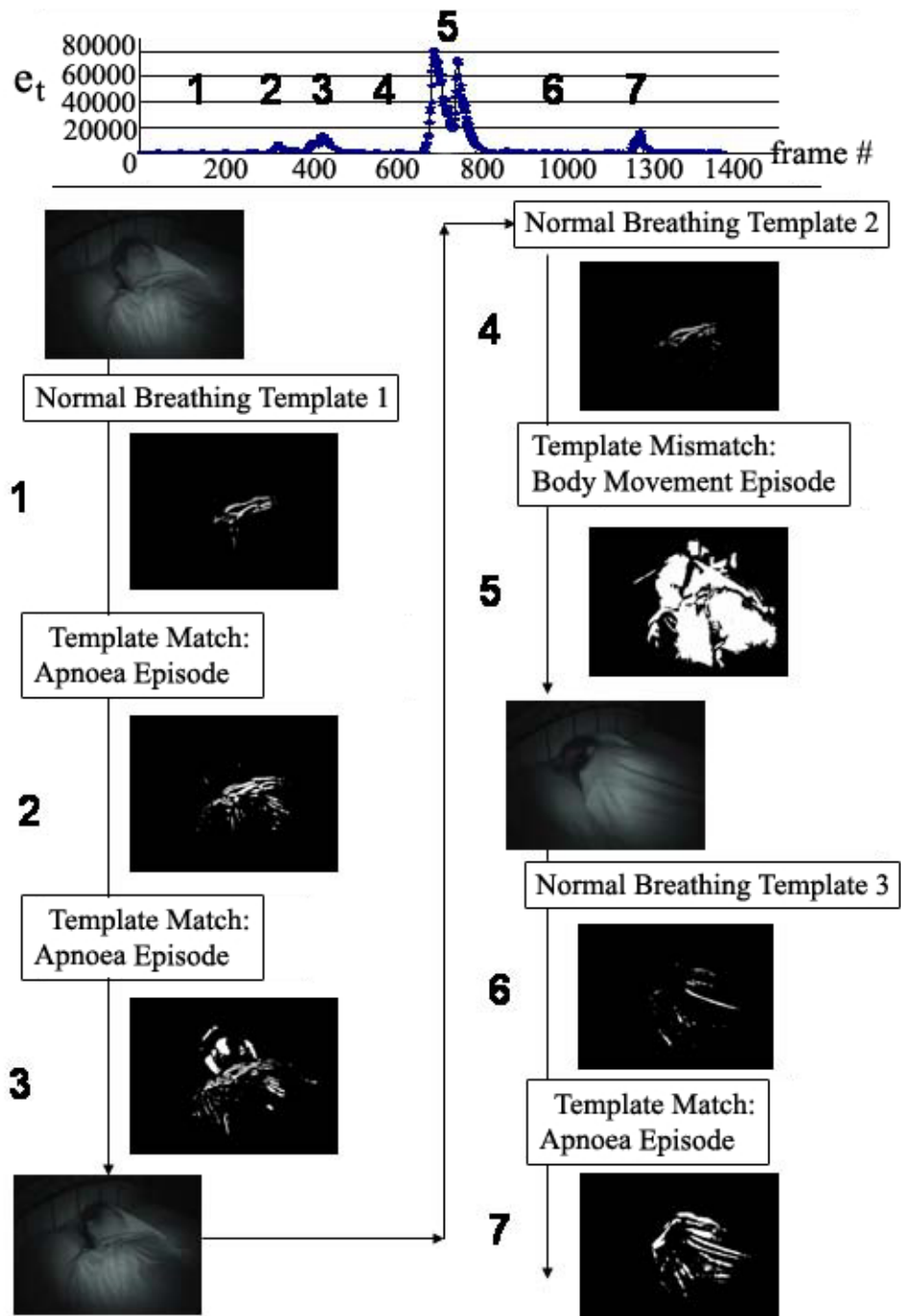


Figure 3.13: Illustration of one experimental analysis output.

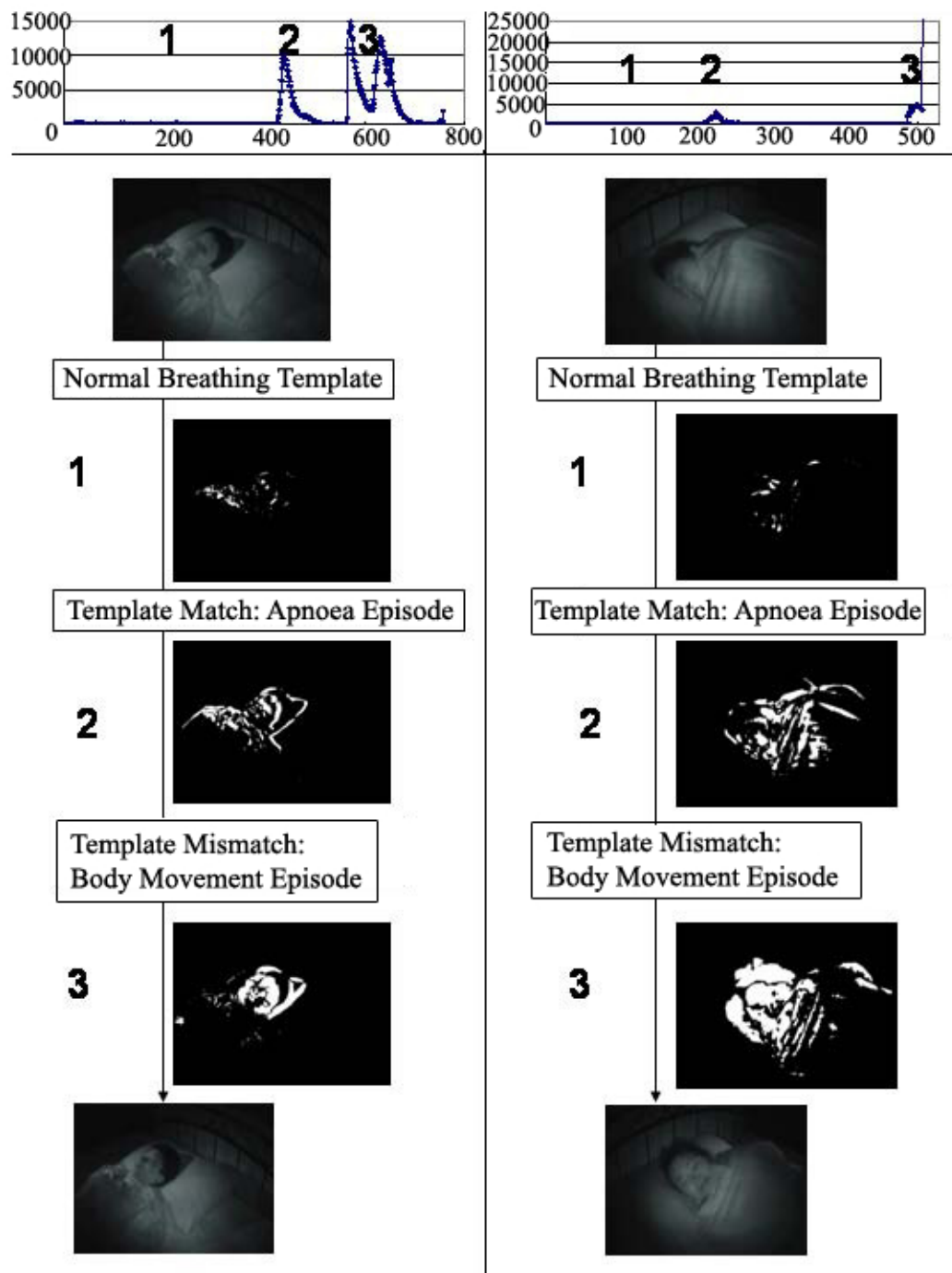


Figure 3.14: Illustration of two experimental analysis outputs.

ously developed. Episode 2 and 3 are abnormal events, both recognized as apnoea episodes by template matching using the template built in episode 1. Episode 4 is identified as a normal breathing episode and a new breathing template is built in this period. The template is then utilized for template matching in the subsequent abnormal event, episode 5, which is identified as a body movement episode. Afterwards, a normal breathing status is detected and the system restarts building a new template of normal breathing activity in episode 6. Later, the template from episode 6 is utilized to define episode 7 as an apnoea episode.

### 3.4.2 Experiments on Clinical Data

**Experimental Setup.** A video system is installed in the sleep lab of Lincoln County Hospital. The video system contains three infrared cameras: two on each side of wall targeting on the upper body of the patient from different angles and one on the ceiling capturing full body view. The design is intended to both support the development of intelligent video approaches, and to ensure that clinical data is soundly captured. In the experiments, the ceiling mounted camera is used for monitoring full covered body activity, and the other two for monitoring breathing activity.

Three symptomatic subjects (one severe and two moderate) and twenty non-symptomatic subjects were recruited to spend one night sleeping in the sleep lab for eight hours video recording. For the symptomatic data, five video clips are randomly sampled from the eight hour recordings of the severe OSA patient; four video clips are randomly sampled from the moderate and minor OSA sufferers (two from each). Each clip lasting 15 minutes, containing 22500 frames. On the other hand, six video clips are randomly sampled from non-symptomatic data of six different subjects. To produce a reference standard, the data were manually marked by the author, who was trained by three Medical Experts from Lincoln County Hospital – Matusiewicz, S.(Medical Doctor), Gravill, N.(Consultant Clinical Scientist) and Barnes, R.(Chief Clinical Physiologist), to identify apnoea (and hypopnoea) episodes using the over-breathing events. As the obtained ethical application does not include PSG, PSG can not be used in our experiments. Hence, it is suggested that an embletta (a portable PSG) is included for future research, in order to produce a more reliable reference standard.

According to the reference standard, we define an event to be correctly recognized if the majority of frames covered by the estimated event have the correct labelling, an admittedly generous test, and a confusion matrix is generated, of which the main diagonal shows the proportion of events correctly classified for each class.

		Diagonal Average: 0.94		
Reference standard	Normal Breathing	.965	.025	.0
	Apnoea	.019	.924	.056
	Body Movement	.0	.058	.941
		Normal	Apnoea	Body
		Estimation of Proposed Method		

Figure 3.15: Confusion Matrix on Clinical Data.

**Experimental Results.** In simulated data, the evaluation focuses on classification on events. There are two aspects to evaluate the performance of the proposed method, i.e. accuracy on classification of events versus classification of symptomatic and non-symptomatic subjects. Regarding the accuracy on classification of events, Figure 3.15 presents the experimental results in the form of confusion matrix. The diagonal average of the confusion matrix is 0.94, which demonstrates that the proposed vision analysis model achieves high accuracy in recognizing individual events for real clinical data. Some template matching outputs are shown in Figure 3.16.

However, the results show that there are slight confusions between body movement events and apnoea episodes, which points out a limitation of the proposed technique for future improvements. The simple action recognition model is less robust than the template matching model, but it is utilized when the template contains insufficient data, which occurs in two situations. The first situation is a combination of considerably low illumination and a subject with shallow breathing pattern, which can be solved by adjusting illumination level. The second situation is that if the timing for action recognition occurs right after the non-repetitious data thrown away by the adaptive template construction algorithm and the speed to discard the non-repetitious data is faster than the individual breathing repetitious cycle, the simple action recognition tends to be activated as well. As a result, more complex noise reduction techniques could be investigated for future research.

**Classification of Symptomatic and Non-symptomatic Subjects.** The Apnoea-Hypopnoea index (AHI) is generally used for evaluation of the severity of OSA in PSG studies, and is calculated as the average number of apnoeas plus hypopnoeas, per hour of sleep. According to Matusiewicz

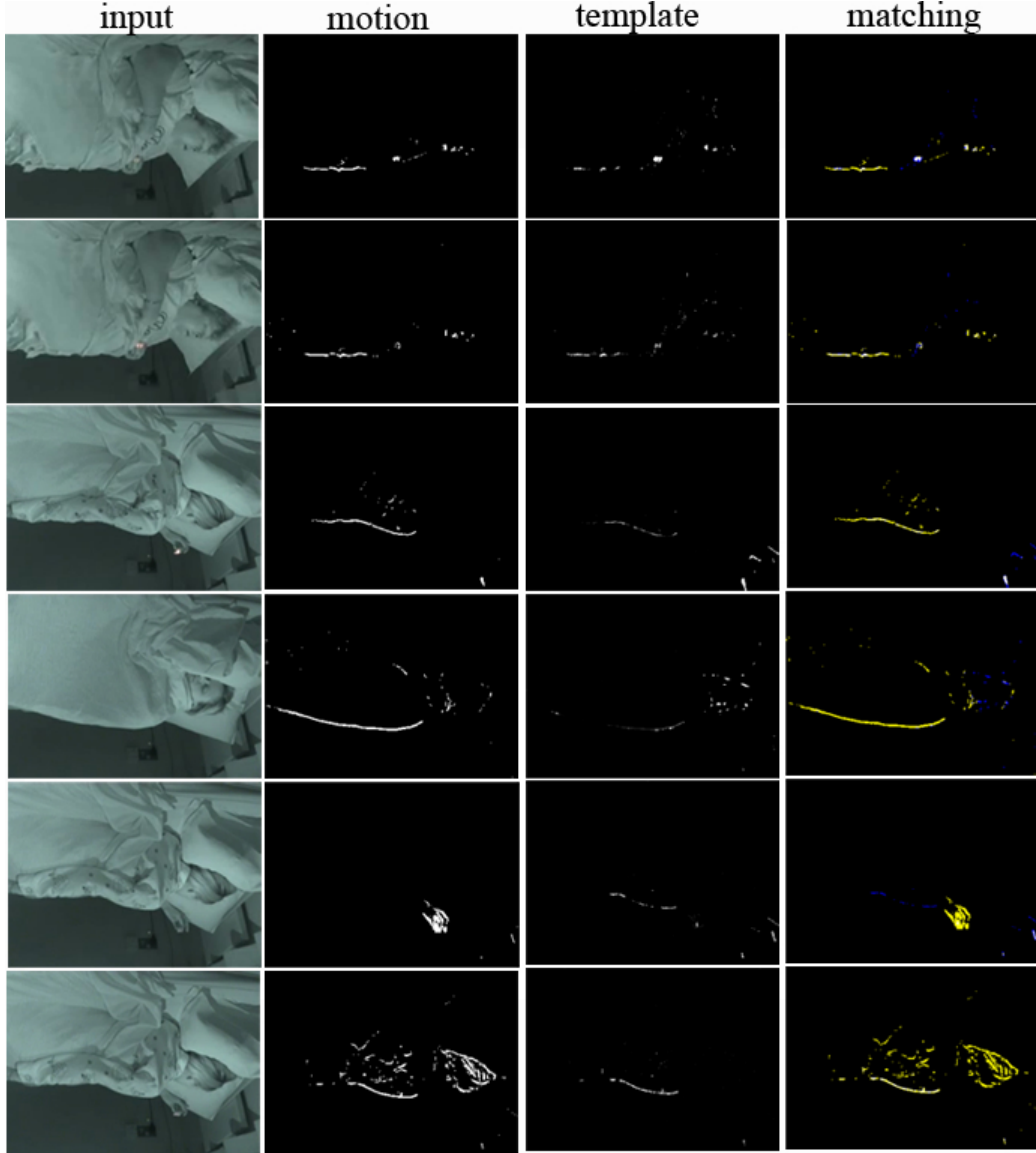


Figure 3.16: Template Matching Screenshots: given the timing  $t$  of template matching, each row contains raw image  $I$ , current motion vector  $M$ , online constructed action template  $T$ , matching result at time  $t$  (yellow:  $\sim T \wedge M$ ; blue:  $T \wedge \sim M$ ; white:  $T \wedge M$ ); the upper 3 rows are apnoea episodes and the lower 3 rows are body movement episodes.

and Gravill [107], it is normal for non-symptomatic subjects to have a few apnoea episodes during sleep, and generally the pulse oximetry traces of non-symptomatic subjects also show a small number of oxygen desaturation episodes ( $ODI < 5/\text{hour}$ ). The distinction between symptomatic subjects and non-symptomatic subjects is that the number of apnoea episodes is considerably higher for the former (the greater the number is, the more severe OSA patients suffer).

We report the number of abnormal episodes detected in individual clinical video clips, to show that the proposed algorithm is able to calculate an index (VAHI) which reflects the severity of the subject OSA: see Table 3.2. We treat detected deep breathing episodes as potential hypopnoea events, and sum up the number of apnoea episodes and  $0.5 \times$  the number of deep breathing episodes, and divide the total by the ratio of the length of the video clip to an hour; we formulate the VAHI as follows.

$$VAHI = (\#Apnoea + 0.5\#DeepBreathing)/(VideoLength) \quad (3.33)$$

Table 3.2 shows that the VAHI values of the symptomatic video clips are distinct from the non-symptomatic ones. In one clip, a non-symptomatic subject had a disturbed sleep and showed a number of body movement episodes and nine apnoea episodes (of which five are minor body movements but misclassified as apnoea episodes, and the other four are over-breathing episodes; this is normal as noted above). Due to limited time and the number of the symptomatic subjects, a model to classify symptomatic and non-symptomatic subjects is not built in this research, and we therefore suggest further investigation on classifying symptomatic and non-symptomatic subjects and stratifying symptomatic subjects for future work.

### 3.4.3 Model Parametrization and Sensitivity

Model parametrization is a common issue in computer vision, and the choice of the parameter values may significantly affect the performance of the algorithms. This section discusses the sensitivity of the proposed algorithm with regard to the control parameters, including  $\alpha$  used in motion detection,  $\lambda$  as the template quality threshold,  $\gamma_1, \gamma_2$  as the action classification threshold,  $\nu$  as the model switch control parameter,  $n$  as the stabilizing factor, and  $\kappa, \beta$  for simple action recognition model.

The proposed method was tested over a range of different parameter values; combinations tested are listed in Table 3.3. In the demonstration system, the combination ( $\alpha = 10, \lambda = 0.00117WH, \gamma_1 = 0.03, \gamma_2 = 0.004, \nu =$

Table 3.2: VAHI Values

	OSA	Apnoea	DB	Body	Video Length	VAHI
Symptomatic Vid1	Severe	11	47	2	15:00	138
Symptomatic Vid2	Severe	32	74	12	15:00	276
Symptomatic Vid3	Severe	20	79	11	15:00	238
Symptomatic Vid4	Severe	32	40	8	15:00	208
Symptomatic Vid5	Severe	33	68	33	15:00	268
Symptomatic Vid6	Moderate	81	59	12	15:00	442
Symptomatic Vid7	Moderate	1	37	2	15:00	78
Symptomatic Vid8	Moderate	67	67	16	15:00	402
Symptomatic Vid9	Moderate	27	60	6	15:00	228
Non-symptomatic Vid1	N/A	0	17	0	15:00	34
Non-symptomatic Vid2	N/A	0	3	0	15:00	6
Non-symptomatic Vid3	N/A	9	14	17	15:00	64
Non-symptomatic Vid4	N/A	1	0	1	15:00	4
Non-symptomatic Vid5	N/A	0	10	0	15:00	20
Non-symptomatic Vid6	N/A	0	13	1	15:00	26

OSA severity obtained from the ODI value from pulse oximetry; DB: Deep breathing; Body: Body Movement; Video Length (mm:ss).

Table 3.3: Values of Model Parameters

$\alpha$	$\lambda$	$\gamma_1$	$\gamma_2$	$\nu$	$n$	$\kappa$	$\beta$
8 ~ 10	11.7 $\Delta$	.03	.004	1.3	10	.26	4.6
10	10.5 ~ 16.5 $\Delta$	.03	.004	1.3	10	.26	4.6
10	11.7 $\Delta$	.05 ~ .025	.004	1.3	10	.26	4.6
10	11.7 $\Delta$	.03	.003 ~ .005	1.3	10	.26	4.6
10	11.7 $\Delta$	.03	.004	1.3	8 ~ 12	0.26	4.6
10	11.7 $\Delta$	.03	.004	1.3	10	.1 ~ .3	4 ~ 5

$\Delta = 0.0001WH$ , W:Width of a frame; H: Height of a frame.

1.3,  $n = 10$ ,  $\kappa = 0.26$ ,  $\beta = 4.6$ ) is adopted, and the same model parameter values were used on both simulated data and clinical data with different environmental settings (e.g. illumination, camera view and camera distance to the subject). Overall, the proposed method is not particularly sensitive to the parameter values. We discuss the important parameters below.

The front end motion detector parameter  $\alpha$  influences the motion detection results. When  $\alpha$  is small (e.g.  $\alpha = 6$ ), more motion is captured, as is noise; when  $\alpha$  is too high, all motion is filtered out. As a result, the selection of  $\alpha$  is important and can influence the settings of other parameters such as  $\lambda$ . A range of values were tested (8 ~ 10), and a large value ( $\alpha = 10$ ) is chosen to filter out high infrared noise.

Another important parameter,  $\lambda$ , determines whether to use cwMatch instead of the simple action recognition model. Compared to the simple model, cwMatch is more robust and accurate, but cannot be used if the online template contains insufficient information. A range of  $\lambda$  values were tested (0.00105 ~ 0.00165), and the lowest effective value ( $\lambda = 0.00117$ ) was selected in order to activate cwMatch as often as possible. As when  $\lambda$  is too low, the template quality cannot be guaranteed, and the algorithm becomes susceptible to “shallow breathing in a high noise environment”, and the action template tends to collect more noise than breathing actions. Other parameters ( $\gamma_1, \gamma_2, \nu, n, \kappa, \beta$ ) are set using the mean of the valid range, to obtain the best results on three short simulated video clips.

To summarize, an heuristic selection of parameter values was used, and more automatic determination of parameter values should be considered for future study.

### 3.5 Conclusion

This chapter presented novel approaches for detecting breathing signals and recognizing abnormal breathing activity from video, to assist in diagnosis of OSA. The proposed methods utilize infrared video information and avoid imposing positional limitations. The technique is real time and robust to heavy occlusion, variances of human breathing behavior and subject appearance, and substantial changes of camera view with respect to the subjects. Furthermore, shallow and abdominal breathing patterns do not adversely affect the performance of the proposed approach, and this technique is not susceptible to illumination changes.

The contributions of this chapter include a novel front end motion detection method to capture real time signals of breathing movements of covered human from video, and a model-free approach for extracting global patterns of motion and quantization of human breathing movement; these patterns are captured as spatiotemporal templates. The method contains an online breathing pattern template construction algorithm, an abnormal event detector, which detects when a quantifiable deviation from the latest normal breathing template occurs, and a novel action recognition model to classify action events, including apnoea episodes, deep breathing episodes, and body movement episodes. Importantly, the normal breathing activity template, which is utilized for action recognition, adapts over time to accommodate to changes in the shapes of normal breathing activities.

The method is shown to have good performance on a limited set of clinical data, and a larger set of simulated data. For future work, more clinical data collection and experimental analysis on symptomatic subjects will be conducted. Furthermore, investigation of automated methods to obtain key parameter values should be included.

Monitoring of breathing has broad applications such as polygraph, sleep studies, sport training, early detection of sudden infant death syndrome in neonates, and patient monitoring. Although the presented approach is mainly targeted at diagnosis of OSA, it could be utilized in other applications that require the analysis of breathing behavior, monitoring subtle and cyclical activity, or capturing adaptive patterns.

## Chapter 4

# Pose Recognition of Covered Human Body

This chapter presents two automated noninvasive video monitoring approaches to recover the human pose in conditions with *persistent heavy obscuration*, allowing for further analysis of *covered* human activity. There is some clinical evidence that the analysis of body pose or movement may be relevant to the diagnosis of OSA. For example, periodic limb movements during sleep are a common finding in patients with OSA [68, 82]. Although the current work does not extend to detection of OSA episodes, it is of interest as a background for future work in that area.

The structure of the chapter is as follows. Section 4.2 describes an existing stylized pose matching model, and presents experimental results on covered human body data. The limitations of this model motivate section 4.4–4.5, which introduce a weak human model to accommodate large variances in appearance and to efficiently produce upper body pose candidates, a new robust pose matching method, and a fast real time simple pose estimation method. In evaluation, section 4.6 presents the experimental results, and section 4.6.3 shows the statistical test results. Section 4.7 concludes the chapter.

### 4.1 Introduction

Estimating human body posture is important for automatic recognition of human activities, with broad applications ranging from human computer interfaces, video data mining, automated surveillance, sport training to medical diagnosis. There has been considerable work in pose recognition in recent

years. The posture of subjects with well-represented appearance or silhouette can be estimated reliably in some systems. However, recognizing the pose of a person who is persistently under cover remains challenging. Many existing approaches to pose estimation make simplifications to the measurement problem, either using motion data (e.g. [1, 45, 67, 146]) to extract silhouettes, or assuming knowledge of appearance or color (e.g. [40, 95, 128, 136]), and the subjects tend to wear close-fitting clothing (or even to be unclothed [40]) in order to extract such information more easily. Such methods work well given clear image cues (for simple detection models) and clean full body motion data (for dynamical models in tracking).

However, in real world applications, partial occlusion often occurs and these assumptions often fail. Apart from partial occlusion, it is likely that limited motion information is available from partial and irregular movements, which seriously affects the usability of the aforementioned methods. Although there is some published research investigating the monitoring of partially occluded humans [69, 129, 153, 170], the methods examined do not deal with pose estimation of consistently and almost wholly occluded subjects. The aforementioned methods are therefore too restrictive for our field of study. Traditional computer vision methods such as correlation, template matching, background subtraction, contour models and related techniques for object tracking are proven ineffective in [22, 75]. The goal of this work is to estimate human poses in conditions of persistent heavy occlusion with irregular movements; see Figure 4.1. The level of occlusion may vary between partially covered, near fully covered and uncovered. In addition, we do not require the subject to be uncovered when (s)he first appears in the scene nor do we require manual initialization.

The principal sources of difficulty in performing this task include: (a) large variances of the image features/appearance of the subject according to the occlusion level, (b) appearance data such as skin color, head-shoulder contour, body outline and ridges of the legs being inaccessible, (c) motion data being partial, irregular and obscured by the cover, (d) obscuration of the bodies' edges by the cover, and (e) strong wrinkle noises from the cover.

#### 4.1.1 Relevant Work

In section 2.3, related computer vision approaches are discussed. In this chapter, a stylized pose matching model by Ramanan *et al.* [128], which is effective in detecting and tracking human poses both indoor and outdoor, is reviewed and tested on the covered human video data. The experimental results are poor. Due to persistent occlusion and heavily obscured image features, the technique tends to be deceived by wrinkle noise, and completely

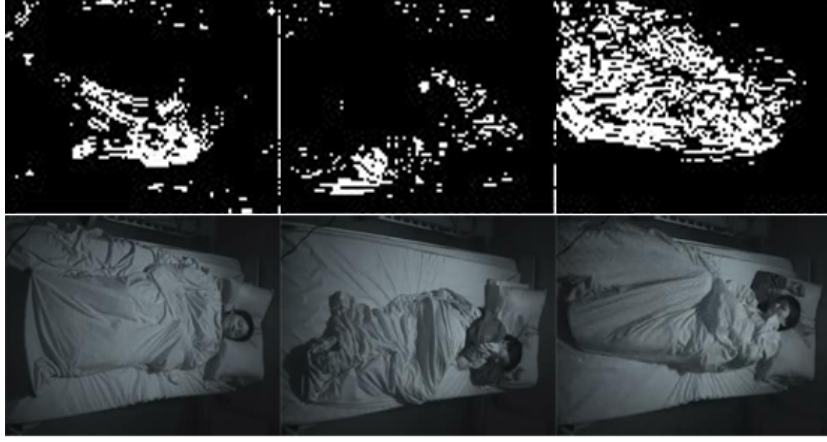


Figure 4.1: Limited Motion and Raw Image: the motion data of a covered sleeping subject is irregular, partial, fragmented and noisy.

fails to identify obscured human poses. Additional experiments were conducted using a later model by Ramanan [131]—the iterative parsing method, which iteratively learns better image features to improve pose estimation, but the pose estimation outcomes are worse, showing that the two existing methods are completely unsuitable to this problem domain. Details of Ramanan’s approaches [128, 131] are given in section 4.2.

#### 4.1.2 Proposed Methods

A weak human model (WHM) is introduced to efficiently produce hypotheses of the upper body parts from obscured human subjects. A new pose matching model, (cwPose), is integrated with WHM to identify the human pose. As the pose matching model, named MatchPose (=WHM+cwPose), takes 0.4 second to process a frame, a real time simple pose model, named RTPose, is also proposed. This integrates WHM with an upper leg pose estimator, a new representation extracting latent features from obscured legs, and a method to reinforce both model parameter and feature space using linking hypotheses.

In evaluation, the proposed methods are compared with Ramanan’s approaches [128, 131]. Regarding computational speed, without code optimization, the system runs near to real time (it takes 0.1 second to process a  $320 \times 240$  frame with a P4 2.4GHz CPU on average for the RTPose model and 0.4 second for the MatchPose model).

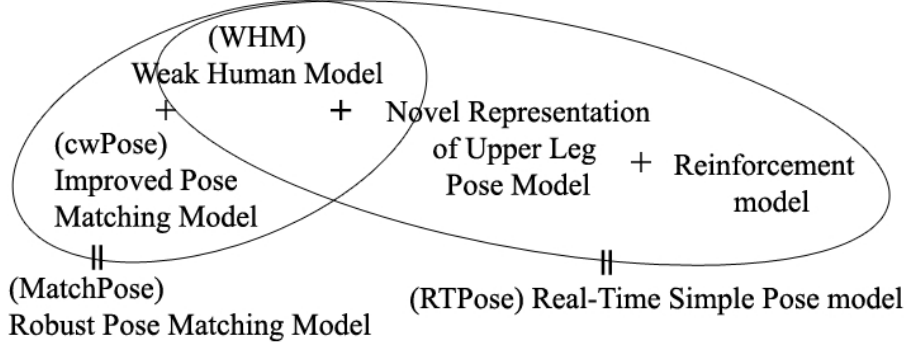


Figure 4.2: Diagram of the two proposed pose recognition methods.

## 4.2 Ramanan Approaches

Ramanan *et al.* [128] developed an automatic system to detect and track the poses of humans from a video sequence by learning their appearance. The human body is modelled as a “puppet” of rectangles connected at joint points, which is often called a pictorial structure [47, 48]. As model-based tracking is easier with a better model, Ramanan refines the generic appearance model by learning the appearance template for each person.

Ramanan’s approach has two stages: first it builds a puppet model of each person’s appearance, and then tracks by detecting those models in each frame. Two approaches are used to learn appearance templates: a bottom-up method and a top-down method. The bottom-up approach first detects candidate parts in each frame with an edge-based part detector, then clusters the resulting image patches to identify body parts that look similar across time, prunes clusters that never move or move too fast in some frames, and groups together candidate body parts found throughout a sequence.

The top-down approach detects stylized lateral walking poses with legs in the form of a distinctive scissor pattern using the tree pictorial structure and the chamfer template matching technique [152] with specified global constraints, which enforce separation of the left and right legs, and also similar appearance based on the  $\mathcal{L}_2$  distance between color histograms. The sample with the lowest cost is kept to build a discriminative appearance model for tracking, by building quadratic logistic regression classifiers in RGB space for each limb using all pixels inside the estimated limb rectangle as positives and all non-person pixels as negatives.

In our problem domain, as the appearance of the subject’s body is occluded by the cover and available body motion is partial and irregular, the



capturing the position  $(x, y)$  and orientation  $\theta$  of part  $i$ .  $P(X^i|X^{\pi(i)})$  is the standard geometric likelihood in a pictorial structure;  $P(I|X^i, C^i)$  is the local image likelihood.  $C^i$  is the part appearance template.

**Modify the Geometric and Image Likelihood Terms.** Importantly, Ramanan *et al.* modify the geometric and image likelihood terms to look for stylized poses. For  $P(X^i|X^{\pi(i)})$ , they manually set the kinematic shape potentials to be uniform within a bounded range consistent with walking laterally, as in Figure 4.3(b), and force the upper legs to be between 45 and 15 degrees with respect to the torso axis. Also, they measure the local image likelihood  $P(I|X^i, C^i)$  using shape matching cost functions by chamfer matching the rectangular edge template. They convolve the distance-transformed edge image [18, 152] with the stylized pose edge templates, and to exploit edge orientation cues, they quantize edge pixels into one of 12 orientations, compute the chamfer cost separately for each orientation with the manually set rotated edge templates, and add the costs together.

They look only at the configurations where all the limbs have high likelihoods (low matching costs) and generate 2000 samples per image from the posterior using the pictorial structure method [47]. Then, the samples are re-scored under two global constraints: (1) forcing similar appearance of left and right legs by computing the disparity in leg appearances using the  $\mathcal{L}_2$  distance between the color histograms and (2) discarding samples where the leg endpoints are within a distance of each other. Ultimately, the sample with the lowest cost is kept to build a discriminative appearance model for tracking using the temporal pictorial structure.

$$P(X_{1:T}^{1:N}, I_{1:T} | C^{1:N}) = \prod_t^T \prod_i^N P(X_t^i | X_{t-1}^i) P(X_t^i | X_t^{\pi(i)}) P(I_t | X_t^i, C^i) \quad (4.2)$$

where the subscripts are used to denote frames  $t \in \{1 \dots T\}$ , and  $P(X_t^i | X_{t-1}^i)$  is a motion model for an individual part.

#### 4.2.2 Shape Matching: Chamfer Matching

Object recognition can be achieved based on the object's color, texture and shape; in the absence of color and texture information as in this problem domain, we must rely on shape alone. Shape matching is a key problem in digital image analysis, and computing the distances between shapes is in principle a global operation. However, global operations are prohibitively costly. Therefore, algorithms that consider only small neighborhoods are necessary.

Thayananthan *et al.* [152] reviewed shape matching methods and compared two shape context approaches and the chamfer matching technique, showing that chamfer matching is more reliable given significant background clutter or variations in scale and shape.

Chamfer matching was first proposed in [13] and later refined in [18, 121] to detect objects based on global shape features. Given a set of trained shape templates, chamfer matching searches the image for locations where these templates are best matched to the image content. To search for a certain object, Borgefors [18] suggests that the template should be an ideal outline of that object, rather than edges extracted from an image.

Object shapes are compared using a distance transform (DT), which approximates global distances by propagating local distances. i.e. distances between neighboring pixels. Matches of a template to the distance-transformed image are found by shifting the template over the image and computing, at each location, the average distance value of all pixels that are covered by the template.

**Chamfer Distance Function.** Two binary images, consisting of feature and non-feature points  $\mathcal{U} = \{u_i\}_{i=1}^n$  and  $\mathcal{V} = \{v_j\}_{j=1}^m$ , are to be matched, and the feature can be any feature visible in both images, e.g. edges, corners, bright spots, or areas with a certain texture. A distance transformation is first applied to the no-feature image, also called the pre-distance image, to assign each non-feature pixel a value that is a measure of the distance to the nearest feature pixel and set the feature pixels to zero. The chamfer distance can then be computed as the average of DT values at the template point coordinates:

$$d_{cham}(\mathcal{U}, \mathcal{V}) = \frac{1}{n} \sum_{u_i \in \mathcal{U}} \min_{v_j \in \mathcal{V}} |u_i - v_j| \quad (4.3)$$

To reduce the effect of outliers and missing edges, the cost can be computed by using the mean of thresholded distances.

$$d_{cham}(\mathcal{U}, \mathcal{V}) = \frac{1}{n} \sum_{u_i \in \mathcal{U}} \max(\min_{v_j \in \mathcal{V}} |u_i - v_j|, \tau) \quad (4.4)$$

As the edge points are influenced by noise, Borgefors [20] indicated that it is a waste of effort to compute exact distances from inexact edges and showed the 3–4 DT to be good enough compared to the Euclidean distance. The propagation can be done either in parallel or sequentially. 3–4 DT is adopted in both [128] and this work; the algorithms are described in Appendix B.

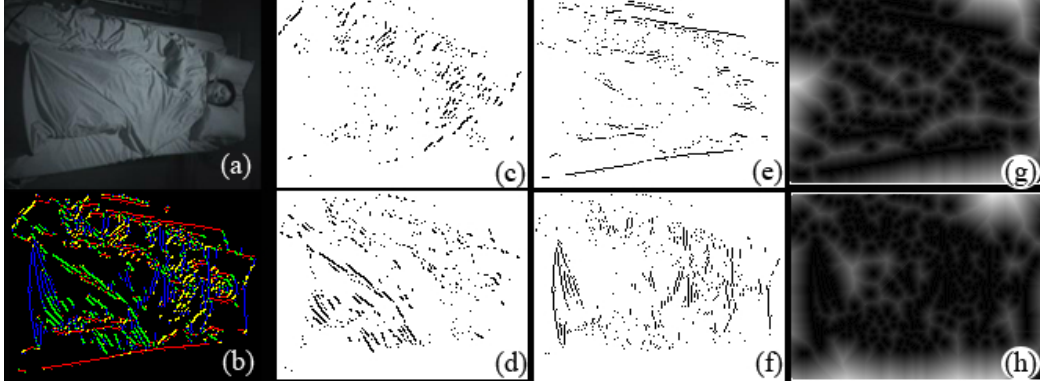


Figure 4.4: Edge Orientation Cues and Distance-Transformed Images: (a) raw image, (b) combined edge orientation, (c–f) individual edge orientation vectors, (g) DT vector based on vector(e), (h) DT vector based on vector(f)

**Using Distance Transformed Images.** The advantage of matching a template with the distance transformed image rather than with the original edge image is that the resulting similarity measure is smoothed, and the influence of outliers can be reduced by using a truncated distance for matching.

**Using Edge Orientation.** The original chamfer matching method and some extensions [13, 18, 61] do not make use of edge orientation information. Thayananthan *et al.* [152] pointed out that the discriminative power of the cost function can be enhanced by exploiting edge orientation cues of the pre-distance image. In their work, the edge orientation of each pixel in the pre-distance image is discretized into 8 regions, producing 4 DT vectors. The chamfer matching cost is then computed as the sum of DT values in individual DT images at the template point coordinates.

Given the number of edge orientations,  $J$ , to exploit and the template point coordinates,  $\mathcal{U}$ , the modified chamfer cost function can be formulated as follows.

$$d_{cham} = \sum_{x,y \in \mathcal{U}} \sum_{j=1}^{J/2} DT_{x,y}(j) \quad (4.5)$$

Ramanan *et al.* [128, 131] used Thayananthan *et al.*'s approach, and quantized edge pixels into one of 12 orientations. Figure 4.4 shows a raw image captured in the sleep lab with an edge image highlighting four different edge orientations, four separate edge orientation vectors, and two DT vectors.

### 4.2.3 Experimental Results on Covered Human Data

**Results of Ramanan’s Method.** The MATLAB code was downloaded from Ramanan’s homepage [132] to test the original stylized pose detector on covered human video data. Experimental results are displayed in Figure 4.5, showing the pose estimation outputs (the mode of the posterior), the posterior and the estimated pixels of the torso, lower arms and lower legs. As the technique is designed for un-occluded human detection, the results are extremely poor in application to obscured covered human subjects. In addition, the method segments hardly any pixels for the lower legs; see Figure 4.5(d).

To illustrate the challenge of obscuration and occlusion by the cover, figure 4.6 compares an un-occluded edge vector from an image used in [128] and an occluded edge image used in this research, showing that the edges of the covered body outline are not only limited but also very noisy. As a result, Ramanan’s detector is seriously influenced by strong wrinkle noise generated by the cover and heavily obscured image features. In addition, the computational cost of Ramanan’s detector is high – around 3 minutes to process one  $320 \times 240$  frame using PC with P4 2.4GHz CPU and 1G memory.

**Results of a Ramanan’s Improved Approach.** In [131], Ramanan introduced an iterative parsing approach to improve pose estimation by iteratively learning better and better features using local features. To explore if this enhanced version is suitable for our problem domain and if boosting local features helps, additional experiments were conducted. The technique is briefly described below.

Regarding pose estimation as inference in a probabilistic model, an iterative parsing process is developed for sequentially learning better features tuned to a particular image in order to improve pose estimation. The technique matches an edge-based deformable model to the image to obtain (soft) estimates of body part positions. Then, the algorithm uses the estimated body part positions to build a rough region model for each body part and the background, i.e. a color model for each part and the background. Afterwards, the algorithm builds a region-based deformable model that looks for possible torsos. Soft estimates of body position from the new model are then used to build new region models, and the process is repeated.

The method is tested with covered body images, and Figure 4.7 displays the pose estimation results, which are even worse than the original Ramanan’s approach.

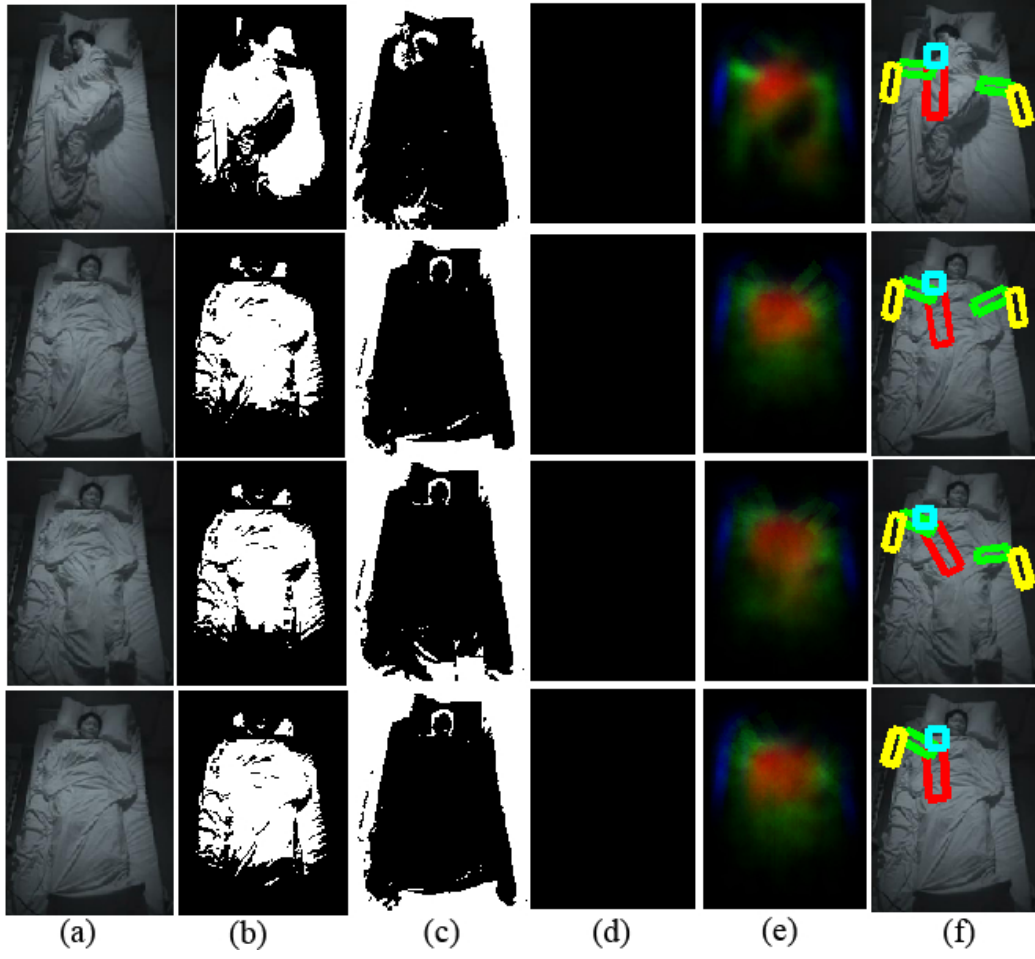


Figure 4.5: Results of Ramanan *et al.*'s Method [128]: (a) input images, (b) torso pixels, (c) lower arm pixels, (d) lower leg pixels, (e) posterior, (f) mode of posterior.

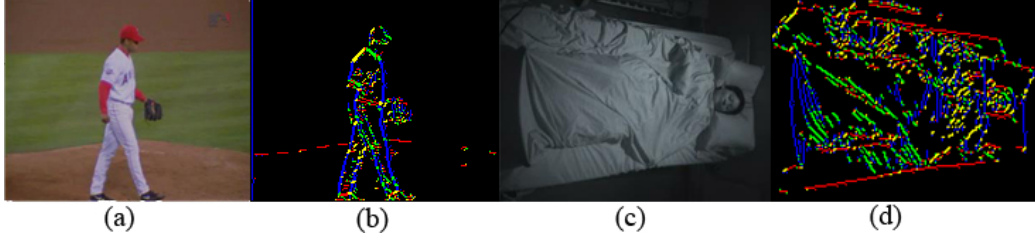


Figure 4.6: Comparison of occluded and un-occluded image feature: (a) Un-occluded human used in [128], (b) Un-occluded edge cue, (c) Covered human, (d) Noisy and obscured edges

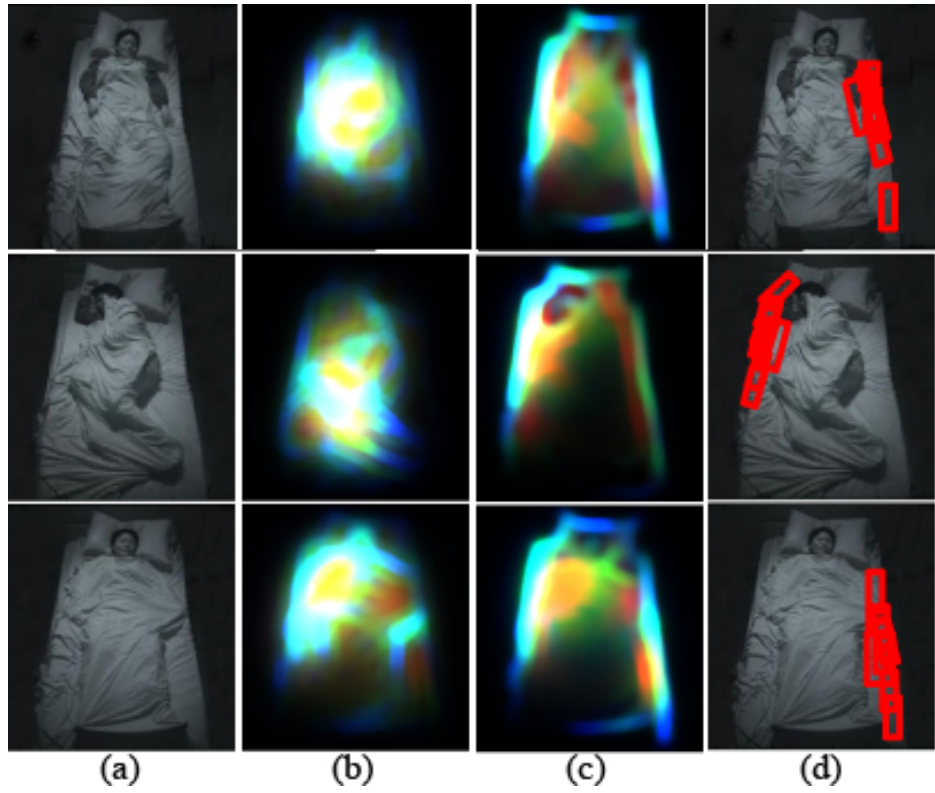


Figure 4.7: Results of Iterative Parsing Technique [131]: (a) inputs (b) the first edge-based parse (c) the second parse using region features from first parse (d) the best-scoring pose

**Analysis.** Appearance modelling is inappropriate in identifying the covered body pose since the appearance of the covered parts are identical due to the cover; the appearance of an individual body part may vary if partially covered, and the appearance model changes over time according to the occlusion status (covered, partially covered or un-covered).

In addition, chamfer matching is susceptible to cluttered schemes; Gavrilu [61] indicates that in cluttered scenes, the chamfer cost function typically has several local minima, and it is necessary to use a subsequent verification stage. As the covered body scene is heavily cluttered, containing strong wrinkle noise from the bedding and a limited number of weakly represented edges of the body, chamfer matching technique cannot be applied directly to the problem domain. In summary, novel methods are required.

## 4.3 The Weak Human Model

The initial set of hypotheses of the head and torso are proposed by a weak human model (WHM), which comprises a novel obscured head model, a novel obscured shoulder detector, a novel obscured torso model and a novel hip joint detector, and adopts the pictorial structure as the basic human representation. This section first introduces three image processing methods for feature extraction and then presents the four novel body part detectors.

### 4.3.1 Feature Extraction

As the subject is occluded and the image features are obscured by the cover, extracting informative features and excluding noise are key to pose recognition. An edge box map technique is introduced to abstract the scene and detect the torso; two edge detectors are built to effectively extract geometric information.

#### 1 Coarse Horizontal Oriented Edge Detector

We first attempt to extract important edges from the outline of the human body while discounting the wrinkles in the sheet. General edge detectors such as Sobel, Prewitt, Kirsch Compass and Laplacian [161] inevitably produce noisy information from wrinkles. Due to the horizontal layout of the bed, an oriented horizontal edge detector is used here to effectively detect object edges aligned with the body, and to remove noise.

In addition, in order to improve edge quality, a gaussian blur filter is applied and the image is down-scaled before processing. This discounts subtle

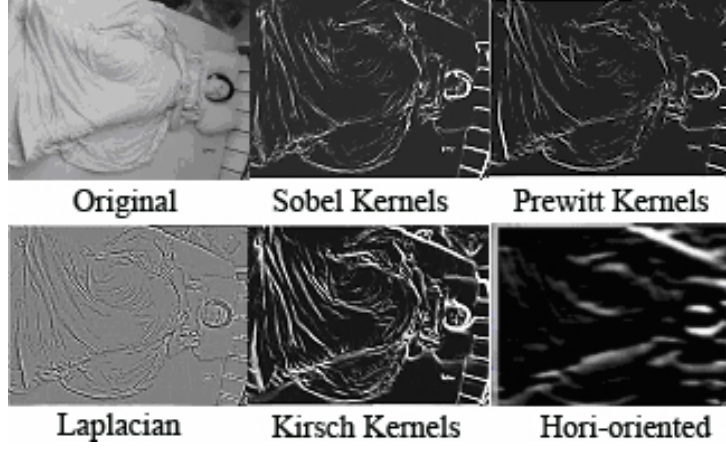


Figure 4.8: Comparison of general edge detectors and the proposed coarse horizontal oriented edge detector

noise and accommodates variance in the local feature representation due to obscuration, and increases computational speed. Figure 4.8 compares different edge detectors over a sample image, showing that the proposed approach outperforms others in both producing the outline of the human body and removing noisy edge information.

The horizontal oriented edge detector is a 2D convolution filter. Given an input image  $I(x, y)$ , the horizontal edge detector produces an edge vector  $I_1(x, y)$ , which is used for head detection and as an input for an edge box map method (see the next section).

$$\begin{aligned}
 I_1(x, y) &= I(x, y) \otimes G_1(w, v) \\
 &= \sum_{w=-m}^m \sum_{v=-n}^n I(x+w, y+v) G_1(w+m+1, v+n+1) \quad (4.6)
 \end{aligned}$$

where  $\otimes$  is the convolution operator; the convolution mask  $G_1(w, v)$  is a  $(2n+1) \times (2m+1)$  matrix and set as:

$$G_1(w, v) = \begin{bmatrix} -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \quad (4.7)$$

## 2 Edge Box Map

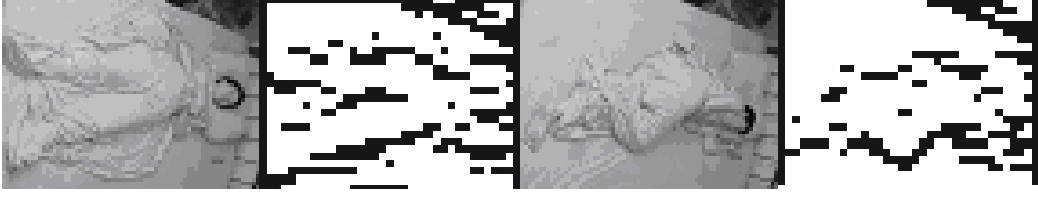


Figure 4.9: Edge Box Maps for Shape Abstraction: the torso can be detected under occlusion from the subject's arms, hands and the cover by utilizing the edge box maps.

As the targeted human body is covered, a portion of the geometric information is lost or even distorted. In order to deal with the issue of discontinuous and scattered edges, an edge box map approach is developed for further shape abstraction. Figure 4.9 displays two edge box maps calculated from edge images, which extract edge information from the original images in the left column. Each edge box thresholds a count of edges within the box. The resulting edge box maps are then utilized by the torso detector to find the covered torso part.

Given a  $(m \times n)$  horizontal oriented edge image  $I_1(x, y)$  and the size of an edge box  $(s \times s)$ , the edge box map  $B(a, b)$  is formulated below.

$$f(x, y) = \begin{cases} 1 & \text{if } I_1(x, y) \geq \nu \\ 0 & \text{otherwise} \end{cases} \quad (4.8)$$

$$B(a, b) = \begin{cases} 1 & \text{if } \sum_{x=(a-1)s}^{as-1} \sum_{y=(b-1)s}^{bs-1} f(x, y) / (ss) \geq \delta \\ 0 & \text{otherwise} \end{cases} \quad (4.9)$$

### 3 Coarse Vertical Oriented Edge Detector

The horizontal edge detector is augmented by a coarse vertical oriented edge detector, which provides auxiliary cues for detection of the shoulders and head; see Figure 4.11(c).

$$\begin{aligned} I_3(x, y) &= I(x, y) \otimes G_2(w, v) \\ &= \sum_{w=-m}^m \sum_{v=-n}^n I(x+w, y+v) G_2(w+m+1, v+n+1) \end{aligned} \quad (4.10)$$

where  $\otimes$  is the convolution operator; the convolution mask  $G_2(w, v)$  is a  $(2n + 1) \times (2m + 1)$  matrix and set as:

$$G_2(w, v) = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix} \quad (4.11)$$

### 4.3.2 Obscured Head Detection

There has been considerable work on face detection in computer vision research over the past ten years; most face detection systems impose postural constraints, requiring frontal view faces or some part of the face like the eyes and nose to be visible. However, patients sleeping with unconstrained poses and under the cover may have the face occluded, or may present half or less of the face when they sleep on the side. Therefore, a head detector robust to occlusion and various body postures is necessary.

Here, a head detector invariant to facial direction and partial occlusion by hands, shoulders or the cover is introduced. In order to accommodate large variances in the head posture and appearance, the detector constitutes four sub-models, utilizing *different coarse features* with *different rules*. The structure is somewhat similar to the cascade classifiers introduced by Viola and Jones [157] for real time face detection. Below, we first give a brief description of Viola and Jones's approach, and explain the differences of the new classifier structure.

#### A Cascade Classifiers by Viola and Jones

In general, classifiers with more features achieve higher detection rates and lower false positive rates, but require more time to compute. To achieve rapid face detection, Viola and Jones introduced an attentional cascade of classifiers, which achieves increased detection performance while radically reducing computational time, by using boosted simple classifiers to reject many of the negative sub-windows while detecting almost all positive instances; see Figure 4.10 (a). Each stage in the cascade reduces the false positive rate and is trained by adding features until the target detection and false positive rates are met (these rates are determined by testing the detector on a validation set). A one feature classifier achieves 100% detection rate and about 50% false positive rate; a five feature classifier achieves 100% detection rate

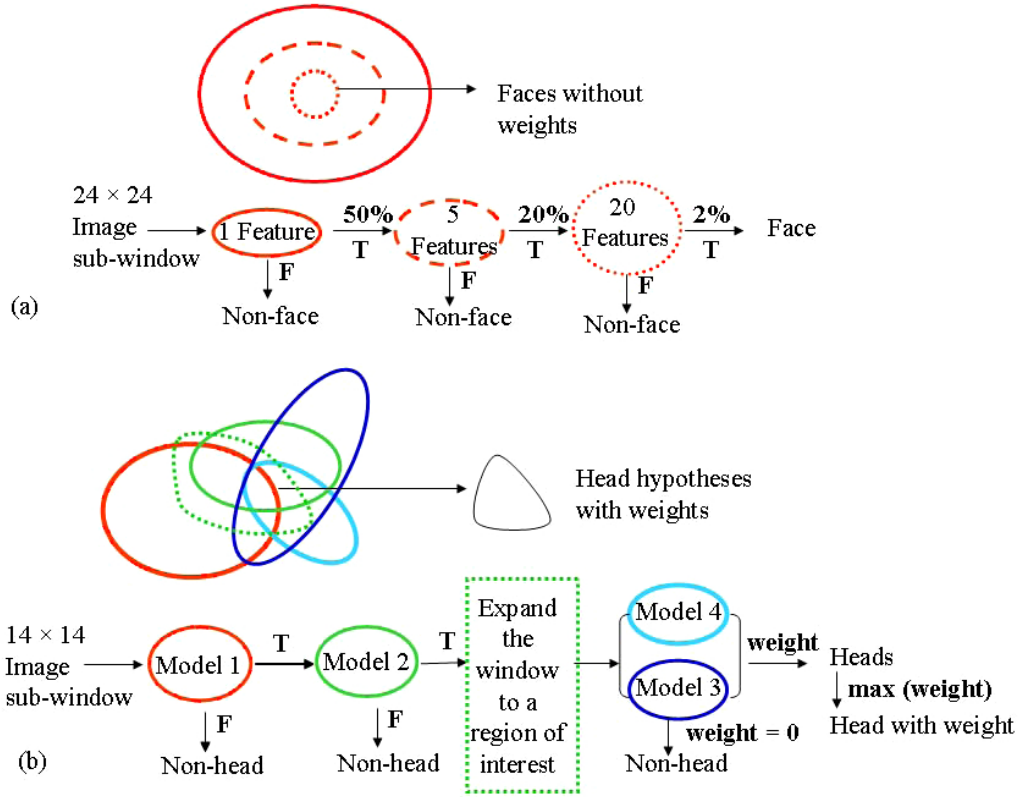


Figure 4.10: Cascade Structure: (a) Boosted Cascade with Single Model [157], (b) Proposed Cascade with Diverse Models

and 40% false positive rate; a 20 feature classifier achieve 100% detection rate with 10% false positive rate. Such a process is similar to a degenerate decision tree.

Stages are added until the overall target for false positive and detection rate is met. To detect *frontal upright faces*, they trained a face detection cascade, containing 38 stages with 6,061 features, and the number of features in the first five layers of the detector is 1, 10, 25, 25 and 50 features respectively. The remaining layers have increasingly more features. Each classifier was trained with the 4,916 hand labelled faces scaled and aligned to a base resolution of  $24 \times 24$  pixels (plus their vertical mirror images for a total 9,832 training faces) and 10,000 non-faces (also of size  $24 \times 24$  pixels using a variant of Adaboost [51]).

However, there are some disadvantages of this approach. Firstly, the time and manual effort on the training process are enormous, e.g. for building a

frontal face detector, 38 stages are trained. More importantly, there is a generalization issue to accommodate large variances in head postures and image features. As a result, a cascade with diverse models is proposed in the next section.

## B Structure of the Proposed Head Detector

The proposed cascade contains four simple head detectors, which evaluate various types of inputs in different models, and can quickly filter out non-head areas while accommodating variations in input data. The cascade allows training data in a low  $14 \times 14$  base resolution, contains significantly less layers in the cascade, and accommodates large variations in input features. Moreover, the training process is robust, without much human intervention, and is computationally efficient. Without code optimization, on a 2.8Ghz Pentium 4 processor, the head detector can process a  $320 \times 240$  image in  $\leq 0.01$  seconds (using greedy search and jumping, described below).

The scheme is illustrated in Figure 4.10 (b). *Model*<sup>1</sup> is an iterative boosting head detector, containing only 3 stages; *Model*<sup>2</sup> is a simple edge clustering model for head detection; *Model*<sup>3</sup> is a simple detector of the top of the head; *Model*<sup>4</sup> is an obscured shoulder detector. Each model is described in the following sections.

Following Ramanan's method, the four head detection models are combined together using weighting functions, and the overall weighting function of the head detector can be formulated as follows.

$$w(X^h) = \begin{cases} w(I_1, X^h)w(I_2, X^h)w(I_3, X^h) & \text{if } D(X^h, X^{sho}) \neq 0 \\ w(I_1, X^h)w(I_2, X^h)(w(I_3, X^h) + w(I_3, X^{sho})) & \text{otherwise} \end{cases} \quad (4.12)$$

where  $w(I_1, X^h)$ ,  $w(I_2, X^h)$ ,  $w(I_3, X^h)$ ,  $w(I_3, X^{sho})$  are the weighting functions of individual models,  $D(X^h, X^{sho})$  is the Euclidean distance between the joint points of  $X^h$  and  $X^{sho}$ ;  $I_1$  is an image observation by the coarse horizontal oriented edge detector (see Equation 4.6);  $I_2$  is an image observation by the Prewitt kernels [119], a contrast enhancement filter and a binary filter;  $I_3$  is an image observation by the vertical oriented edge detector (see Equation 4.10) and a binary filter. Figure 4.11 displays the three different image observations,  $I_1, I_2, I_3$ , for head detection.

### 1 *Model*<sup>1</sup>: Iterative Boosting Head Detector

The first model developed is a head detector, reflecting the fact that the head is generally the most easily detected body part of a sleeping subject.



Figure 4.11: Image observations for head detection: (a) Horizontal oriented edge image  $I_1$  (b) image by Prewitt kernels and a contrast enhancement filter  $I_2$  (c) Vertical oriented edge image  $I_3$  (d) Definition of potential areas of shoulders  $S_1, S_2, S_3, S_4$ , and the top of the head  $R_t$  to the head region  $R_h$

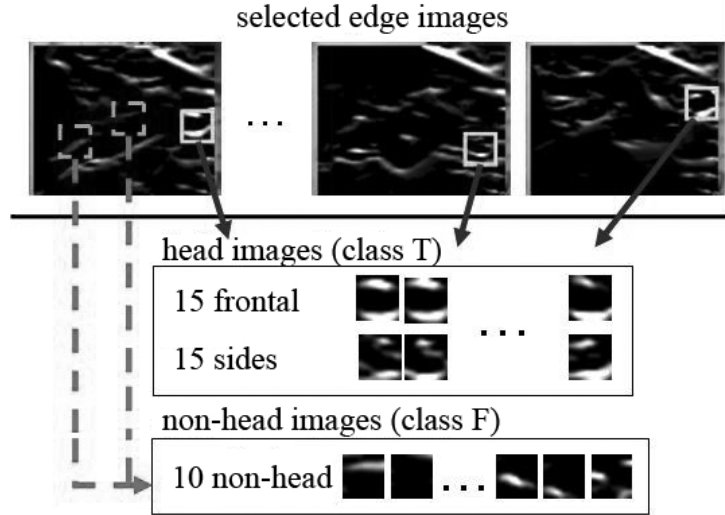


Figure 4.12: Initial training data for  $\mathcal{M}_1$ : the data format is a  $14 \times 14$  matrix using the intensity values of edge images  $I_1$

Following Viola and Jones’s approach [157], we used machine learning algorithms to train a head template invariant to head poses and partial occlusion. We introduce a low variance error boosting algorithm and an iterative boosting cascade. This method reduces human intervention and manual effort on training process. Without building an enormous number of stages, a simple 3-layer cascade classifier is iteratively constructed. Each boosting model consists of ten C4.5 decision trees [127], with the features automatically selected. To keep this chapter concise, the low variance error boosting algorithm and associated experiments are presented in Appendix C.

The initial training set consists of 30 hand labelled heads (class T), including 15 frontal and 15 side views, scaled and aligned to a base resolution of  $14 \times 14$  pixels, together with 10 randomly selected non-head images (class F); see Figure 4.12. All data was collected from a single video clip.

The first layer classifier  $\mathcal{M}_1$  was trained using the 40 training samples, and tested using the same video clip. Next, we randomly selected 10 false positive instances generated by  $\mathcal{M}_1$  into the training data of class F. The second layer classifier  $\mathcal{M}_2$  was then trained using the obtained 50 images. Another randomly selected 10 false positive instances generated by  $\mathcal{M}_1 \wedge \mathcal{M}_2$  were added into training data of class F. Using the collected 60 instances, the third layer classifier  $\mathcal{M}_3$  is trained. Figure 4.13 illustrates the iterative construction scheme. The weighting function of the iterative boosting head detector can be formulated as follows.

$$w(I_1, X_h) = \begin{cases} 1 & \text{if } \mathcal{M}_1 \rightarrow T \wedge \mathcal{M}_2 \rightarrow T \wedge \mathcal{M}_3 \rightarrow T \\ 0 & \text{otherwise} \end{cases} \quad (4.13)$$

Importantly, we employ coarse edge information to avoid the influence of different facial appearance, expression and direction. This helps to identify the most important patterns. The benefits of this are: (1) the base resolution of the data is reduced ( $14 \times 14$  versus  $24 \times 24$  in [157]); (2) the training data is significantly condensed (40 images versus 19,832 images in [157]); and (3) the number of layers in the cascade is greatly decreased (3 versus 38 in [157]). The ultimate result is that the computational cost is reduced from 0.06 to 0.01 seconds per frame.

Furthermore, the iterative construction of the boosting classifiers allows patterns and rules to be continuously refined by focusing on the false positive instances from the previous learning experience. This method utilizes a relatively small number of instances to build the machine learning models, and moreover all training data is selected from one single video clip with one subject. In evaluation, the experimental results show that the head detector works robustly in all 32 test video sequences from eight different people,

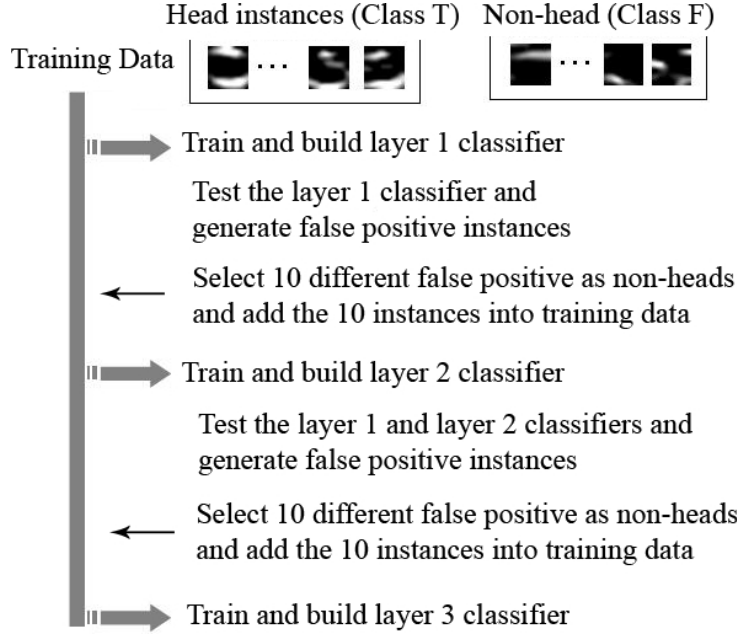


Figure 4.13: Iterative Construction of Boosting Cascade

which were recorded with four different filming angles in two different environments under two different illumination settings.

## 2 *Model*<sup>2</sup>: Simple Edge Clustering Model

The second measurement model filters out regions with low levels of edge response. The function  $w(I_2, X^h)$  indicates the number of edges in the region exceeding a threshold  $\alpha$ .

$$w(I_2, X^h) = \begin{cases} 1 & \text{if } \sum_{(x,y) \in h_2} I_2(x, y) > \alpha \\ 0 & \text{otherwise} \end{cases} \quad (4.14)$$

where  $\alpha = \varsigma \times \text{area of } X_h$  ( $\varsigma=0.1$ , which is determined empirically using training data);  $I_2$  is an image observation by the Prewitt kernels [119], a contrast enhancement filter and a binary filter (see Figure 4.11(b)).

## 3 *Model*<sup>3</sup>: Simple Detector of the Top of the Head

A simple head detector evaluates a region according to the edge response of the top of the head. Given a potential head region  $R_h$ , we evaluate the

top of the head  $R_t$  as illustrated in Figure 4.11(d), and define the confidence weight  $w(I_3, X^h)$  of the head:

$$w(I_3, X^h) = \sum_{(x,y) \in R_t} I_3(x, y) \quad (4.15)$$

where  $I_3$  is an image observation by the vertical oriented edge detector (see Equation 4.10) and a binary filter.

#### 4 *Model*<sup>4</sup>: Obscured Shoulder Detection

A commonly adopted method for shoulder detection is a head-shoulders contour matching model [95, 96, 170], which chamfer matches pre-trained head-shoulder contour shapes. However, the shape matching methods are susceptible to cluttered scenes and not suitable for the obscured body, as we showed in section 4.2.3. A novel obscured shoulder detector is presented here. There are two distinctive shoulder postures – the frontal posture and the side posture. The frontal posture is detectable by the roughly symmetric (obscured) two-side shoulders, but the appearance of the shoulder from the side view is not conspicuous. Hence, we create a weighting function  $w(I_3, X^{sho})$  for potentially heavily obscured frontal shoulders. We define four regions  $(S_1, S_2, S_3, S_4)$  with potentially two shoulder areas; see Figure 4.11(d).

Given a head hypothesis  $X_h$  with its topleft coordinate  $(x, y)_{X_h}$ , height  $h_{X_h}$  and weight  $w_{X_h}$ , we define  $(S_1, S_2, S_3, S_4)$  as follows.

$$w_{S_i} = w_{X_h}/2, h_{S_i} = h_{X_h} \quad (4.16)$$

$$(x, y)_{S_1} = (x_{X_h} - w_{X_h}/4, y_{X_h} - h_{X_h}) \quad (4.17)$$

$$(x, y)_{S_2} = (x_{X_h}, y_{X_h} - h_{X_h}) \quad (4.18)$$

$$(x, y)_{S_3} = (x_{X_h} - w_{X_h}/4, y_{X_h} + h_{X_h}) \quad (4.19)$$

$$(x, y)_{S_4} = (x_{X_h}, y_{X_h} + h_{X_h}) \quad (4.20)$$

To characterize roughly symmetric shoulders, a pair of shoulder features  $(a_1, a_2)$  is obtained from the four edge representation indices  $(c_1, c_2, c_3, c_4)$  of the four potential areas  $(S_1, S_2, S_3, S_4)$ .

$$c_i = \sum_{(x,y) \in S_i} I_3(x, y) \quad (4.21)$$

The shoulder weighting function is designed to characterize the symmetry of the shoulders (using the symmetric orientation index  $\Delta = |a_1 - a_2|$ ) and the strength of feature response levels  $(a_1, a_2)$ . The shoulder weighting function  $w(I_3, X^{sho})$  is formulated as follows.

$$w(I_3, X^{sho}) = \begin{cases} \frac{a_1 + a_2}{\Delta} & \text{if } a_1 \geq 0 \wedge a_2 \geq 0 \wedge \Delta > 0 \\ 1 & \text{if } a_1 \geq 0 \wedge a_2 \geq 0 \wedge \Delta = 0 \\ 0 & \text{otherwise} \end{cases} \quad (4.22)$$

where  $\Delta$  is created to represent symmetric orientation level.

$$\Delta = |a_1 - a_2| \quad (4.23)$$

$(a_1, a_2)$  is formulated as follows.

if  $c_1 \geq \beta \wedge c_3 \geq \beta$ ,

$$(a_1, a_2) = (c_1 - \beta, c_3 - \beta) \quad (4.24)$$

otherwise, if  $c_2 \geq \beta \wedge c_4 \geq \beta$ ,

$$(a_1, a_2) = (c_2 - \beta, c_4 - \beta) \quad (4.25)$$

otherwise, if  $\max(c_1, c_2) \geq \beta \wedge \max(c_3, c_4) \geq \beta$ ,

$$(a_1, a_2) = (\max(c_1, c_2) - \beta, \max(c_3, c_4) - \beta) \quad (4.26)$$

otherwise,

$$(a_1, a_2) = (-1, -1) \quad (4.27)$$

where  $\beta = 18$  is determined empirically from the training data.

To illustrate the above function, where higher weight is given to instances with higher symmetrical orientation ( $\Delta \downarrow$ ) and stronger edges level ( $a_1 \uparrow, a_2 \uparrow$ ), some examples of experimental results are listed in Table 4.1. Importantly,  $w(I_3, X^{sho}) = 0$  does not necessarily mean that the shoulders are not in the frontal posture because they may be obscured. However, whenever the shoulders are clearly detectable, we utilize the shoulder information  $I_3$  to assist in the estimation of the neighboring nodes, i.e.  $X^h$  and  $X^{tor}$ .

### 4.3.3 Obscured Torso Detection

The appearance of the torso varies considerably according to the level of occlusion by the cover, the hands or the arms. Hence, to accommodate large variance in the appearance of the torso, we develop two measurement models, including a new obscured torso detector and a novel hip joint detection model. In addition, three classes are defined for the output of the torso detector: type 1 – 45 degree head to torso; type 2 – 90 degree head to torso; type 3 – 135 degree head to torso. The algorithm is described as follows.

Table 4.1: Experimental Results on Shoulder Detection

$c_1$	$c_2$	$c_3$	$c_4$	$a_1$	$a_2$	$\Delta$	Importance $w(I_3, X^{sho})$
<b>23</b>	29	<b>37</b>	10	5	19	14	1.71
10	<b>35</b>	<b>34</b>	17	16	17	1	33
18	<b>27</b>	9	<b>37</b>	9	19	10	2.8
17	7	26	27	-1	-1	0	0

\*Two figures  $(a_1, a_2)$  are selected from  $(c_1, c_2, c_3, c_4)$ , shown in **Bold**.

### Obscured Torso Detection Algorithm

If  $w(I_3, X^h, X^{sho}) > 0$ , indicating that clear shoulders are visible

$$w(X^{tor}) = w(I_3, X^h, X^{sho})$$

output Type 2.

Otherwise,  $X^{tor'} = \arg \max(w(I_{tor}, X^{tor}))$

If  $X^{tor'} \in \{\text{Type 1}, \text{Type 3}\}$ , then compute the weight of the hip joint to the torso

$$\text{If } w(I_3, X^{tor'}, X^{hip}) < \tau \wedge w(I_3, X^{tor''}, X^{hip}) > \varphi$$

(where  $X^{tor''} = \text{Type 1}$  if  $X^{tor'} = \text{Type 3}$ ; otherwise  $X^{tor''} = \text{Type 3}$

, and  $\tau, \varphi$  are defined in equation 4.31 and 4.32)

$$w(X^{tor}) = w(I_{tor}, X^{tor''})$$

output  $X^{tor''}$

Otherwise

$$w(X^{tor}) = w(I_{tor}, X^{tor'})$$

output  $X^{tor'}$

## 1 Obscured Torso Detector

The core idea of the torso recognition algorithm is to search for a relatively smooth region within a reasonable distance and angle from the head, i.e. an area near to the head with a low interior edge box count.  $I_{tor}$  is the edge box map introduced in section 4.3.1, and the detection is as defined below; see Figure 4.14.

$$w(I_{tor}, X^{tor}) = \left( \sum_{(a,b) \in X^{tor}} B(a,b) \right)^{-1} \quad (4.28)$$

## 2 Obscured Hip Joint Detector

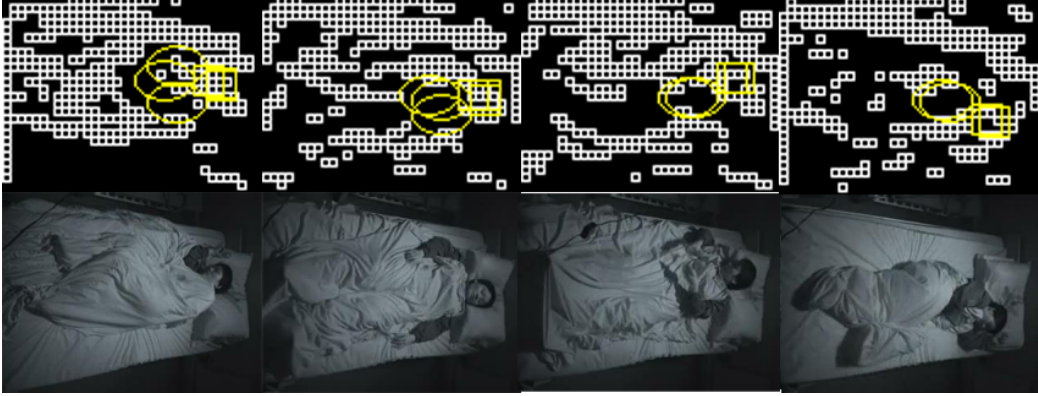


Figure 4.14: Edge box maps: regions with high  $w(I_{tor}, X^{tor})$  are highlighted

If the estimated torso from the previous section is Type 1 or Type 3, we assume both that the person is lying on his/her side and that the image observation of the hip joint to the torso is detectable. Similar to the design concept of the obscured shoulder detection model, three sub-regions  $\{J_{lp}\}_{p=1,2,3}$  are defined to search for the potential hip joint according to the estimated torso type  $l$ , in order to accommodate variances of the hip joint's position and obscuration by the cover. Figure 4.15 illustrates the image observation  $I_3$  and the corresponding hip joint sub-regions.

$$w(I_3, X^{tor}, J_{lp}) = w(J_{lp}) = \sum_{(x,y) \in J_{lp}} n(x, y) \quad (4.29)$$

$$w(I_3, X^{tor}, X^{hip}) = \begin{cases} \max(w(J_{11}), w(J_{12}), w(J_{13})) & \text{if } X^{tor} = Type_1 \\ \max(w(J_{31}), w(J_{32}), w(J_{33})) & \text{if } X^{tor} = Type_3 \end{cases} \quad (4.30)$$

We assume that  $w(I_3, X^{tor}, X^{hip}) \sim \mathcal{N}(m_j, \sigma_j^2)$ , and the two parameters in the Obscured Torso Detection Algorithm in section 4.3.3 are defined as follows.

$$\tau = m_j - 2 \times \sigma_j \quad (4.31)$$

$$\varphi = m_j \quad (4.32)$$

#### 4.3.4 Use Temporal Coherence

Spatio-temporal approaches have been shown to be useful in overcoming self-occlusion and image noise in recent research [94, 97, 143]. These methods exploit temporal coherency of feature points. In contrast, we exploit the



Figure 4.15: Estimated Pose and Image Observation for Hip Joint Detectors

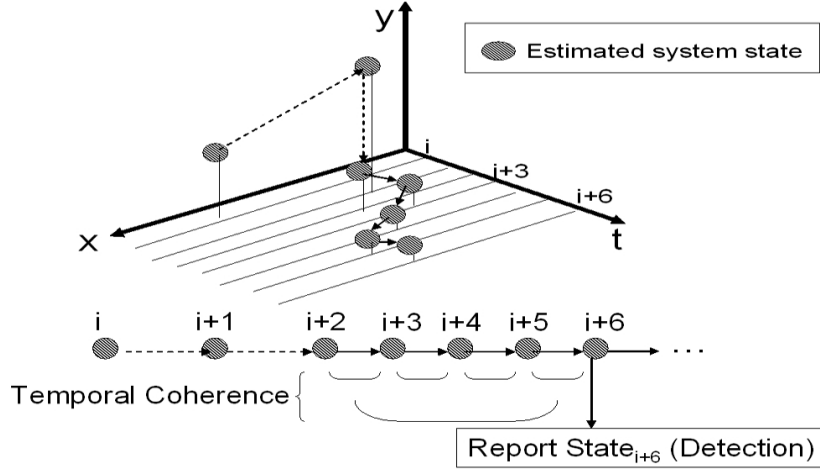


Figure 4.16: Temporal coherence on patterns.

property of temporal coherence on system states rather than features, constructing temporally coherent patterns. The advantage is that using temporal coherence in feature development risks transition errors from observed data to estimated states, whereas applying temporal coherence directly to system states avoids such risks.

On each time-step,  $t$ , the algorithm evaluates multiple hypotheses for the head position,  $\{X_i^h\}$ , and torso position,  $\{X_j^{tor}\}$ , and a single hypothesis for the upper legs position,  $g_t$ . The strongest hypothesis at time  $t$  is  $X_t = (X_t^h, X_t^{tor}, g_t)$ . If hypothesis  $X_t$  yields a reasonably consistent position over a sufficiently long time period, then detection with  $X_t$  is declared. We use a threshold for head displacement,  $\rho$ , and a minimum stable period,  $\tau$ .

Let  $k$  be a count of consecutive stable iterations; initialize  $k=0$ . On each iteration,  $t$ , if  $|X_t^h - X_{t-1}^h| < \rho$ , increment  $k$  by 1; otherwise, set  $k=0$ . Declare detection using  $X_t$  when  $k > \tau$ . An illustration is given in Figure 4.16 ( $\rho = 0.3, \tau = 3$  are used in our experiments).



Figure 4.17: Results of the search algorithm

#### 4.3.5 Search Method: Greedy Search with Jumping

As the availability of spatial features can be variable and the true hypothesis can be heavily occluded, the coarse-to-fine search strategy [50] sometimes misses detection. Hence, we conduct the search in an independent greedy manner every frame both to collect as much information as possible and to avoid filtering out the true hypothesis due to sampling (for head detection, the greedy search is on the right hand side,  $\frac{2}{5}$  of each frame).

In order to reduce the computational cost, we construct a jumping mechanism, which forces the greedy search to skip neighbor rows below a detected point. Importantly, the jumping function is combined with the computationally efficient head detector model (section 4.3.2), which finds an optimal position in the local area. Therefore, skipping can be conducted without missing targets and reduces computing time. Moreover, compared with post-processing hypotheses after search using the weighted mean, the mode or clustering [170], the jumping function avoids a merging process and saves effort on both searching and post-processing. Some results are presented in Figure 4.17.

Given an  $M \times N$  region of interest and the current position  $(x, y)$ , the next search position for the head  $(a, b)$  can be formulated as follows. ( $k=3$  is used based on our experiments).

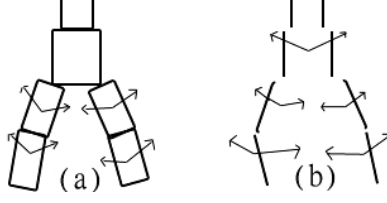


Figure 4.18: Pose Pictorial Structure: (a) Ramanan's Template Representation (b) Modified Template Representation

$$(a, b) = \begin{cases} (x + 1, y + k) & \text{if } (x, y) \in \{X^h\} \\ (x + 1, y) & \text{if } (x, y) \notin \{X^h\} \wedge (x + 1) < M \\ (0, y + 1) & \text{otherwise} \end{cases} \quad (4.33)$$

## 4.4 Robust Pose Matching Method (Match-Pose)

We introduce an enhanced pose matching algorithm, cwPose, for obscured human pose estimation, by improving the chamfer cost formula and Ramanan's template representation to overcome the issues of weakly represented image features and strong noise due to heavy occlusion. Although cwPose significantly improves pose estimation, it is still susceptible to strong noise. Hence, we combine it with WHM, which substantially reduces the search space by generating soft estimates of upper body parts, with cwPose as a subsequent tuning-up function to find the local minimum in a constrained space.

### 4.4.1 Modified Pose Matching Algorithm (cwPose)

The cwPose method is adapted from Ramanan's method to: two major modifications – the chamfer matching cost formula and the representation of the templates.

**Improved Chamfer Cost Function.** The first improvement is to exploit the edge orientation of the template and to compute matching cost on the correlated DT vectors. We improve the cost function by focusing on the specific strongest edge orientation information related to the template.

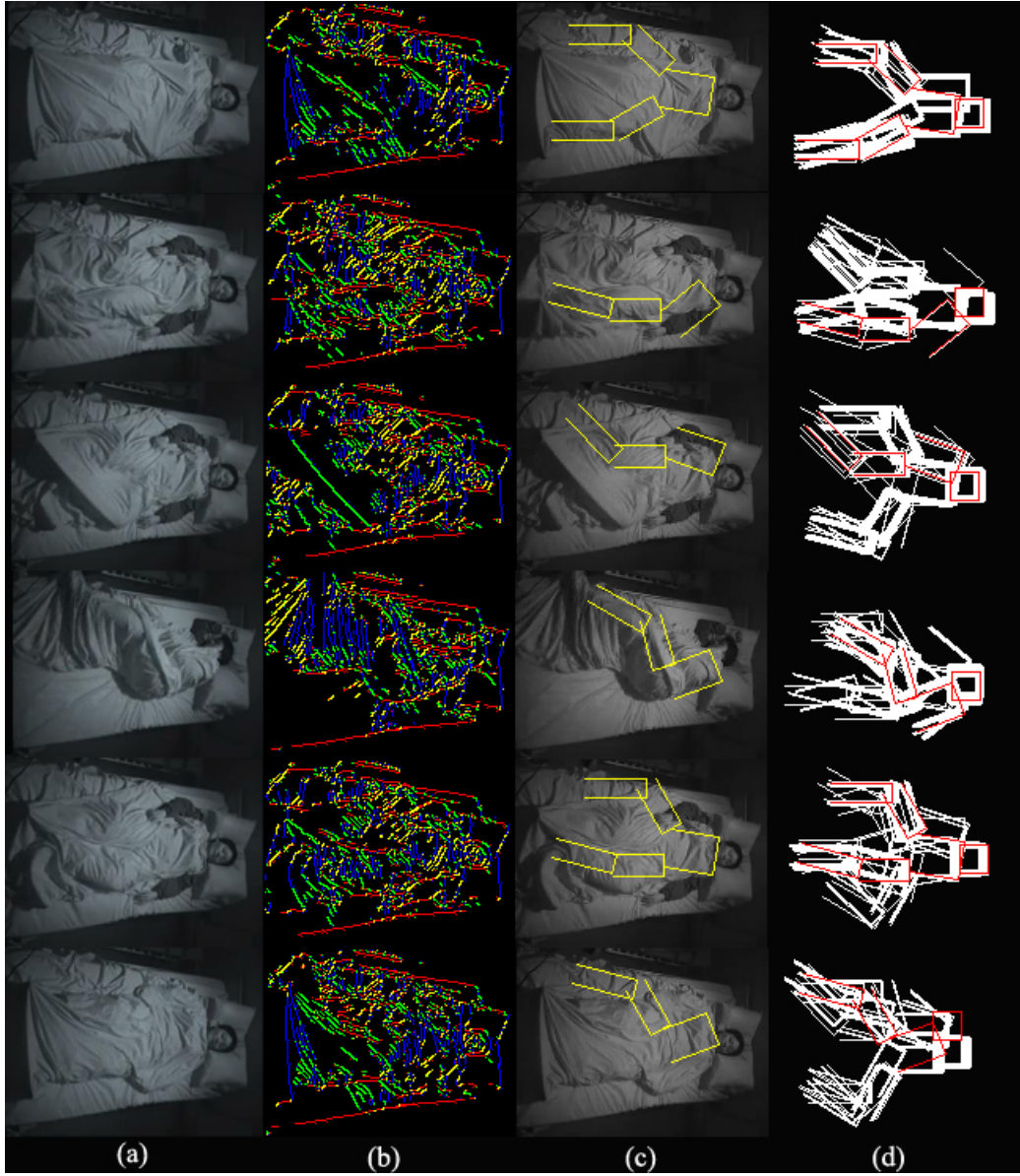


Figure 4.19: Results of Improved Pose Matching Model: (a) inputs (b) edge orientations (c) the best scoring pose (d) top 25 poses are highlighted in white, with the best scoring pose highlighted in red

Thayananthan *et al.* [152] improved chamfer matching by utilizing edge orientation information of pre-distanced images, and computed the cost as the sum of DT values for individual edge orientations at the template coordinates, but did not exploit edge orientation of feature points in the template; see Equation 4.5. Instead of summing across all orientations, we first identify the edge orientations of template features, and select the closest-correlated DT vector(s) for individual features. This focuses on closely-related information and filters out possible noise and distraction. For example, given four DT vectors,  $(DT(0^\circ), DT(45^\circ), DT(90^\circ), DT(135^\circ))$ , if the edge orientation of the template feature point is  $20^\circ$ , the correlated DT vectors,  $DT(0^\circ)$  and  $DT(45^\circ)$ , are selected to compute the chamfer matching cost.

If there are two DT vectors selected, instead of summing across these, the strongest representation (the lowest DT cost) is chosen. For example, if the template feature point is  $20^\circ$ , the chamfer matching cost is equal to  $\min(DT_{x,y}(0^\circ), DT_{x,y}(45^\circ))$ . This adopts the strongest possible performance of the matched point instead of its average performance.

The modified cost function is:

$$d_{cham} = \begin{cases} \sum_{x,y \in \mathcal{U}} DT_{x,y}(k^\circ) & \text{if } k^\circ = \theta^\circ \\ \sum_{x,y \in \mathcal{U}} \min(DT_{x,y}(k^\circ), DT_{x,y}(l^\circ)) & \text{if } k^\circ < \theta^\circ < j^\circ \end{cases} \quad (4.34)$$

where  $\mathcal{U}$  is the set of the template coordinates;  $\theta$  is the edge orientation of the feature point  $(x, y)$ .

**Modified Template Representation.** Only the outside borders of the body parts are used to match in order to capture relatively reliable edge features, and to avoid the large amount of noise generated by the cover. The modified representation of a person template is displayed in Figure 4.18.

**Improved Pose Estimation Outcomes.** Figure 4.19(c) shows the estimation results of cwPose. The experimental results show a small improvement in covered body pose estimation compared to Ramanan. Although the overall pose estimation performance is still poor and often misled by strong wrinkle noise, Figure 4.19(d) shows the correct pose is likely to be among the top 25.

## 4.4.2 The Integration Framework of MatchPose

Following Ramanan, MatchPose adopts the pictorial structure [47, 48]. It localizes the head first and then finds the remaining limbs in a *directed* search

scheme; see Equation 4.1. MatchPose integrates WHM and cwPose by using the cost function of cwPose and the weighting function of WHM. It first uses WHM to quickly identify soft estimates of weighted head and torso hypotheses, then uses cwMatch to match the best candidate (hard estimate) for each soft estimate, and re-evaluates each hard estimate by deducting its weight by WHM from its cost by cwPose. The output pose is the pose with the lowest total cost of the head, torso, upper right leg, upper left leg, lower right leg and lower left leg. The detailed integration algorithm of MatchPose is described in Algorithm 1.

As chamfer matching is computationally costly, the drawback of MatchPose is the processing speed. It takes 0.4 seconds to match a  $320 \times 240$  frame with a P4 2.4GHz CPU. Hence, we also propose a real time simple pose estimation algorithm in the next section.

---

**Algorithm 1** MatchPose Integration Algorithm

---

1. use WHM to conduct a directed search for the head and torso
    - 1.1 Obtains multiple pairs of weighted head and torso  $\{X^h, X^{tor}\}$
  2. Measure the chamfer cost of  $X^h$  and reset the costs  $c(X^h)$  as
    - 2.1  $c(X^h) = c_{cwPose}(X^h) - w_{WHM}(X^h)$
 where  $c_{cwPose}(X^h)$ : the matching cost by cwPose;  $w_{WHM}(X^h)$ : the weight by WHM.
  3. select the head(s)  $\{X^{h*}\}$  with the lowest cost
    - 3.1  $\{X^{h*}\} = \arg \min_{X^h} c(X^h)$
  4. use  $\{X^{h*}\}$  to sample torsos  $\{X^{tor}\}$  obtained by WHM
    - 4.1 select  $\{X^{tor*}\}$  within a distance tolerance  $d$  to  $X^{h*}$
  5. Measure the chamfer cost of  $\{X^{tor*}\}$  and reset the costs of sampled torso hypotheses as
    - 5.1  $c(X^{tor}) = c_{cwPose}(X^{tor}) - w_{WHM}(X^{tor})$
 where  $c_{cwPose}(X^{tor})$ : the cost by cwPose;  $w_{WHM}(X^{tor})$ : the weight by WHM.
  6. select the torso(s)  $\{X^{tor*}\}$  with the lowest cost
    - 6.1  $\{X^{tor*}\} = \arg \min_{X^{tor}} c(X^{tor})$
  7. use  $\{X^{tor*}\}$  and cwPose to search for Legs
  8. choose the pose(s) with the lowest cost.
- 

## 4.5 Real Time Simple Pose (RTPose)

As an alternative to MatchPose, a real-time simple pose estimation approach, RTPose, is proposed here. This combines WHM with an upper leg pose

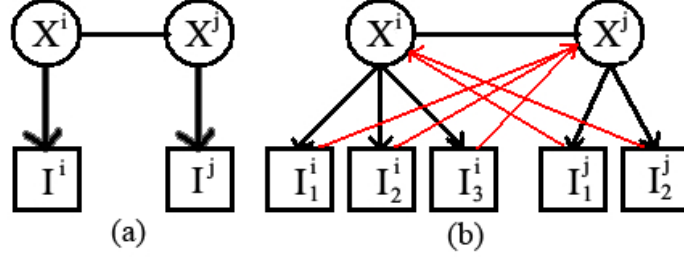


Figure 4.20: Relationship between two model parameters: (a) Markov Network (b) Reinforcement network

estimator, a new representation to extract latent features from obscured legs, and a reinforcement model. In this section, we first describe the reinforcement model, then the upper leg estimator and the integration framework.

#### 4.5.1 Reinforcement by the Linking Parameters

In order to deal with heavy occlusion, we propose a modified network that reinforces both the model parameter  $X^i$  and associated feature spaces  $\{I_i\}$  by its adjacent parameter  $X^j$  and a search framework that aggregates detections over time to produce a more reliable hypothesis.

In tracking human poses, a common representation of human configuration is a Markov Network [74, 171], which is similar to the temporal pictorial structure except that in the Markov Network the edges linking between body parts (model parameters) are undirected instead of directed; see Figure 4.20 (a). The joint posterior distribution of the Markov network is

$$P(X^{1:N}|I) \propto \prod_{(i,j) \in \mathcal{E}} P(X^i|X^j) \prod_i^N P(I|X^i) \quad (4.35)$$

where  $\mathcal{E}$  is the set of all undirected links;  $P(X^i|X^j)$  models the constraints between two adjacent body parts (the shape model); and  $P(I|X^i)$  is the local image likelihood.

**Reinforced feature space and model parameters.** To enforce the detection in obscured space, we propose a modified network by both adding auxiliary image observations  $\{I_k^i\}_{k=1:M}$  for each model parameter  $X^i$  and adding relationships  $\theta_{ji}(X^j, \{I_k^i\})$  between the adjacent model parameters and the image observations (details are given in the next section). As the

features are weakly represented in our problem domain, the reinforcement by linking hypotheses performs better than mere selection of hypothesis. Using weighting functions rather than posterior distributions, we not only reinforce the model parameter by the adjacent model parameter but also reinforce the obscured feature space in order to generate an accurate model parameter. The weighting function of the proposed model is formulated as follows, and Figure 4.20(b) illustrates the relationship between adjacent model parameters of the modified network.

$$w(X^{1:N}, I^{1:M}) = \prod_{(i,j) \in \mathcal{E}} w(X^i, X^j) \prod_i^N \left( \prod_k^M (w_i(\{I_k^i\}, X^i) - \theta_{ji}(X^j, \{I_k^i\})) \right) \quad (4.36)$$

## 1 Head Tracker: reinforced features and hypotheses

To estimate the model parameter  $X^h$  at time  $t$ , we reinforce the image observations  $\{I_k^h\}$  for  $X^h$  using the known adjacent model parameter  $X^{tor}$  at time  $t - 1$ . As in Ramanan, the image likelihood term is modified. The measurement of the local image likelihood  $P(I_k^h | X^h)$  uses the weighting function of the sub-head detector, and the new link  $\theta_{ji}(X^j, \{I_k^i\})$  zeros the confidence weights of a portion of features within the region  $\mathcal{A}$  derived from  $x_j$ , to decrease the likelihood of  $X^h$  occurring in the area derived from  $X^{tor}$ , that is  $(w_i(\{I_k^i\}, X^i) - \theta_{ji}(X^j, \{I_k^i\}))$ .

Three image observations  $I_1^h$ ,  $I_2^h$  and  $M_t$  are first extracted for head tracking.  $I_1^h$  is image observation by Prewitt edge detector,  $I_2^h$  is image observation by the coarse horizontal oriented edge detector, and  $M_t$  is the motion cue based on images processed by a convolution filter; sequences of processed images are used to compute Difference of Frames (DOF). The convolution filter for image preprocessing is formulated as follows. An example of DOF using raw images is illustrated in Figure 4.21(d), which contains comparatively less information than Figure 4.21(c). Given an input image  $I(j, k)$ ,

$$I(j, k)' = I(j, k) \otimes q(w, v) = \sum_{w=-K}^K \sum_{v=-K}^K I(j-w, k-v) q(w, v) \quad (4.37)$$

where  $2K + 1 = \text{size of } q(w, v)$  and  $q(w, v)$  is set to

$$q(w, v) = \begin{pmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix} \quad (4.38)$$

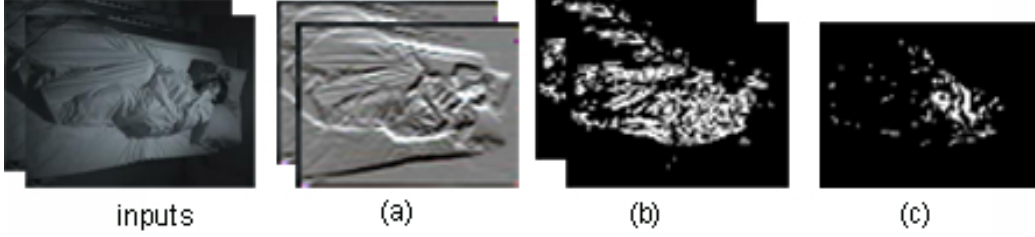


Figure 4.21: (a) input raw images (b) images by a convolution filter (c) DOF outputs using the processed images by convolution filter (d) DOF output using raw images.

To estimate  $X_t^h$ , we produce reinforced observations  $I_{1'}^h$ ,  $I_{2'}^h$  and  $M_t'$  from  $I_1^h$ ,  $I_2^h$  and  $M_t$  using the adjacent model parameter at time  $t - 1$ ,  $X_{t-1}^{tor}$ . We zero the confidence weights of features within the region  $\mathcal{A}$  derived from  $X_{t-1}^{tor}$ . For edge maps ( $I_{1'}^h$ ,  $I_{2'}^h$ ),  $\mathcal{A}$  is defined as a vertically expanded area of  $X_{t-1}^{tor}$  to reduce noise; for  $M_t'$ ,  $\mathcal{A}$  is the area of  $X_{t-1}^{tor}$  (see Figure 4.22 for the reinforced features). Next, we sample instances within the area with high likelihood at the previous frame. We then use the edge clustering model (see section 4.3.2) to search the area over  $I_{1'}^h$  and  $M_t'$ , producing two weight maps. We select two hypotheses  $X_1^h$  and  $X_2^h$  with the highest confidence weight from the two maps respectively. Importantly, we argue that appearance features like  $I_{1'}^h$  should be weighted much more than motion information like  $M_t'$ , because motion may be caused by the cover surface movement or hand movement. Hence, the hypothesis  $X_2^h$  derived from motion cannot be relied on to define the state, but can assist in improving the hypothesis and to activate inspection of different evidence. We measure the distance between  $X_1^h$  and  $X_2^h$  to confirm the precision of  $X_1^h$ . If  $|X_1^h - X_2^h| < \alpha$ , where  $\alpha$  is the tolerance, we define  $X_t^h = X_1^h$ ; otherwise, we produce an auxiliary hypothesis  $X_3^h$  using  $I_{2'}^h$  and  $X_{t-1}^h$  with the appearance model and define the state  $h_t^*$  as the average of  $X_3^h$ ,  $X_1^h$  and  $X_2^h$ . The initial set of hypotheses  $X^h$  and  $X^{tor}$  are proposed by WHM.

## 2 Torso Tracker

A motion event is a mixture of target movement and occluding object movements. Motion detected in the target will be used to update the hypothesis; motion by occluding objects will not. With regard to the torso, occluding object movements include arm movement and cover surface movement where the subject may pull or remove the cover. We estimate a new torso hypothesis based on motion, a latent image observation  $I_{tor}$ , previous

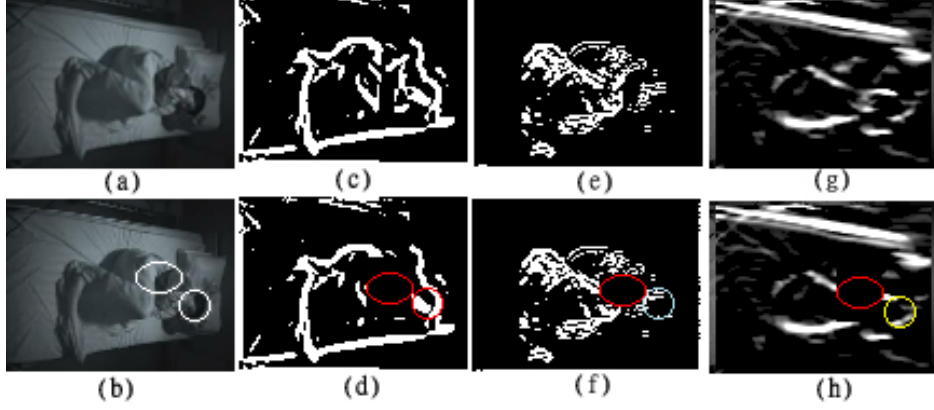


Figure 4.22: (a) raw image (b) system output (c) image observation  $I_1^h$  by Prewitt edge detector (d) reinforced feature space  $I_1^h$  with the associated hypothesis  $X_1^h$  (e) motion data  $M_t$  (f) reinforced motion data  $M_t'$  with the associated hypothesis  $X_2^h$  (g) auxiliary image observation  $I_2$  (h) reinforced auxiliary image observation  $I_2^h$  with its associated hypothesis  $X_3^h$ .



Figure 4.23: (a) edge box map with  $X_t^h$  (b) reinforced edge box map

torso hypothesis  $X_{t-1}^{tor}$ , the previous adjacent model parameter  $X_{t-1}^h$  and the current adjacent model parameter  $X_t^h$ :

When motion detected within the region of the previous torso hypothesis  $X_{t-1}^{tor}$  is over  $\beta$  percentage, it suggests the hypothesis may need updating. To confirm a torso activity occurred, we check if motion occurs within the region of  $X_{t-1}^h$  over  $\gamma$  percent. If true, we then adjust the torso hypothesis based on the two types of image observations. Firstly, we compute the intersection of the motion data and a vertically expanded area  $\mathcal{A}$  derived from  $X_{t-1}^{tor}$  using  $\zeta$ . Denoting the intersection as  $\mathcal{D}$ , we generate a temporary torso hypothesis using the center of  $\mathcal{D}$ :  $X_t^{tor'} = \overline{\mathcal{D}}$ . Secondly, we produce a latent image observation, edge box maps, as in Fig 4.23 (a) and reinforce the feature by zeroing the confidence weights of features within a region  $\mathcal{B}$ , where  $\mathcal{B}$  is a vertically enlarged area of  $X_t^h$ . Thirdly, we input the reinforced features and the

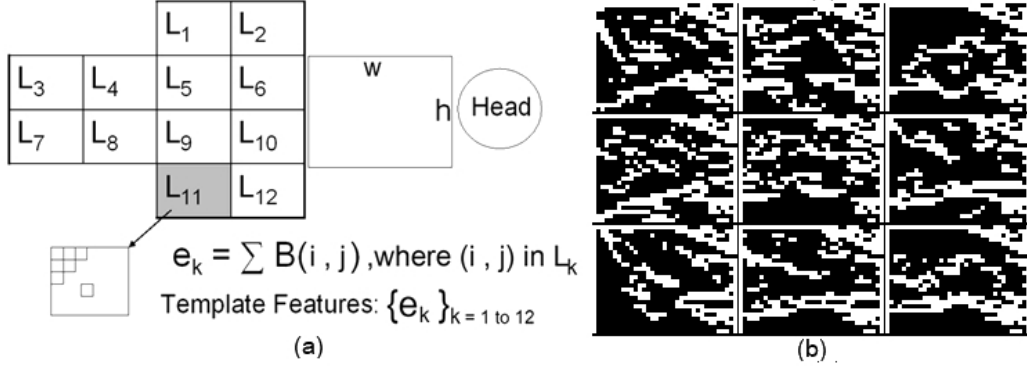


Figure 4.24: (a) Representation of upper-legs pose model (b) Some edge box maps.

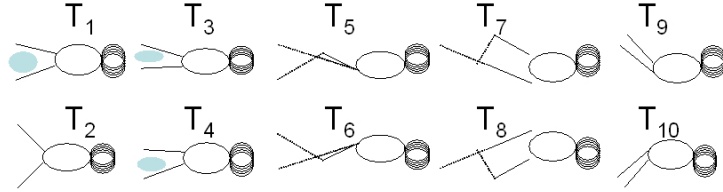


Figure 4.25: 10 upper-legs pose templates.

current head hypothesis  $X_t^h$  to an obscured torso measurement model [159] and generate another torso hypothesis  $X_t^{tor''}$ . Then, we compare the distance between  $X_t^{tor'}$  and  $X_t^{tor''}$  and define  $X_t^{tor}$  as follows ( $\varrho$  is a distance tolerance.)

$$X_t^{tor} = \begin{cases} \overline{X_t^{tor'} X_t^{tor''}} & , \text{ if } (|X_t^{tor''} - X_t^{tor'}| > \varrho) \\ X_t^{tor''} & , \text{ otherwise} \end{cases} \quad (4.39)$$

#### 4.5.2 Coarse Upper Leg Pose Recognition

Existing approaches for locating legs use cues like silhouettes, ridges, color blobs, parallel edges, or cone / rectangle shape edge pixels, and assume that these features are easily detectable. However, such an assumption is not applicable to our problem domain. Here, we introduce a novel representation for obscured upper leg pose recognition. The representation model contains 12 features  $\{e_k\}$  to represent the sum of edge boxes in individual subparts  $\{L_k\}$  in a given edge box map, where  $k = 1$  to 12.

To test the novel representation, we manually collect 26 images to produce

a training dataset to construct 10 different pose templates. Furthermore, instead of collecting images for all poses, we collect images for template  $T_3$ ,  $T_5$ ,  $T_8$  and  $T_9$ , and use mirror projection theory to produce training data for symmetric templates  $T_4$ ,  $T_6$ ,  $T_7$  and  $T_{10}$ . In training, we use the low variance error boosting introduced in Appendix C as the learning method for generating a number of classifiers, and the classifiers are built into a binary tree structure. In testing, we apply the model to a new dataset, which includes a number of poses and movements.

### 4.5.3 The Integration Framework of RTPose

RTPose integrates WHM with the novel upper leg pose estimator and the reinforcement tracker. It localizes the head and torso in a *undirected* search scheme (a Torso-to-Head Backward Selector is added here to aid WHM), and localizes the upper leg in a directed search scheme. RTPose first uses WHM to identify multiple weighted head and torso positions, then selects the strongest torso candidate, uses a Torso-to-Head selector to choose a head hypothesis, and uses the obtained torso and upper leg pose estimator to identify upper leg pose. The output is the obtained head, torso, upper leg pose; the head and torso are further refined over time by the reinforcement tracker. The Torso-to-Head Backward Selector is described below, and the integration algorithm of RTPose is presented in Algorithm 2.

---

**Algorithm 2** RTPose Integration Algorithm

---

1. use WHM to conduct a directed search for the head and torso
  - 1.1 Obtains multiple pairs of weighted head and torso  $\{X^h, X^{tor}\}$
2. Select the torso(s)  $\{X^{tor*}\}$  with the highest weight
  - 2.1  $\{X^{tor*}\}_{k=1:N} = \arg \max_{X^{tor}} w_{WHM}(X^{tor})$

where  $w_{WHM}(X^{tor})$  is the weight by WHM.

3. Set the torso hypothesis  $X^{tor*}$  as the average of the set
  - 3.1  $X^{tor*} = \overline{\{X^{tor*}\}_{k=1:N}}$
4. Conduct Torso to Head Backward Selector
  - 4.1 Select the head hypothesis  $X^{h*}$
5. torso to leg search: use  $X^{tor*}$  and the upper leg detector
  - 5.1 obtain  $X^{leg*}$

Output  $X^{h*}, X^{tor*}, X^{leg*}$  for detection

6. use the reinforcement model on  $X^{h*}, X^{tor*}$
-

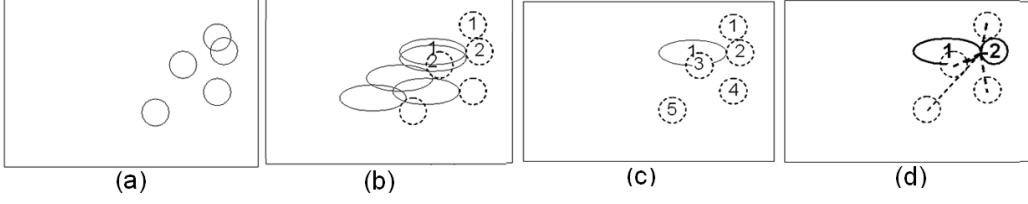


Figure 4.26: (a) Head hypotheses  $\{X^h\}$  by head detectors (b) Head to torso search:  $\{X^h, X^{tor}\}$  (c) Compare  $\{X^{tor}\}$  and choose the strongest one as  $X^{tor*}$ . (d) Torso to head search: output hypothesis  $(X^{h*}, X^{tor*})$ .

**Torso to Head Backward Selector.** Obtaining the strongest torso candidate  $X^{tor*}$ , we compute the distance between the joint location  $o$  of the torso  $X^{tor*}$  to the centers  $\{c_k\}$  of the head hypotheses  $\{X^h\}$ . We then choose the head hypothesis  $X^{h*}$  with the closest distance to  $o$  as the radius  $r_k$  of the head  $X_k^h$  as follows.

$$p = \arg \min_k (||o - c_k| - r_k|) \quad (4.40)$$

$$X^{h*} = X_p^h \quad (4.41)$$

Figure 4.26 illustrates an example of the undirected head and torso model, showing that the backward voting function chooses a head hypothesis in a related reasonable location, and hence  $X_3^h$  will not be selected. Also, the resulting head-torso pair  $(X_2^h, X_1^{tor})$  may not be identical to the original head-to-torso pairs, i.e.  $(X_1^h, X_1^{tor})$  or  $(X_2^h, X_2^{tor})$ .

## 4.6 Evaluation

This section describes the experimental setup and evaluation data (section 4.6.1) and compares the experimental results of the three different methods (Ramanan, RTPose, MatchPose), in section 4.6.2. Statistical significance test results are presented in section 4.6.3.

### 4.6.1 Experimental Setup and Data

Non-visible infrared is adopted, and the infrared video frames were acquired at 15 fps using a SONY infrared camcorder (DCR-HC-30E) at a resolution of  $320 \times 240$ . A short video clip for training the boosting templates of head and upper leg pose was captured using the environmental setting and the subject as illustrated in Figure 4.27(a).

Table 4.2: Evaluation Data Distribution

	Systematically Sampled	Randomly Sampled for Evaluation
Cover	7048	513
No Cover	433	42
High illumination	6837	461
Low illumination	644	94
Total	7481	555

The testing video clips were captured on eight subjects with various height, weight, gender and skin color. 32 video clips were filmed in three environments with different illumination and camera angle settings. The experimental data contains a number of unconstrained poses and various occlusion levels (i.e. fully covered, partially covered and without cover). From the 32 testing video clips, we randomly select 18 video clips, containing 22443 frames. For a quantitative evaluation, we first systematically sample frames at 0.3 second intervals, obtaining 7481 frames, and then randomly sample 555 frames for evaluation. The evaluation data can be categorized in various classes based on the illumination and occlusion; the distribution of each class is listed in Table 4.2.

To produce a reference standard, all evaluation frames were manually marked for individual body parts using Adobe Photoshop, and following Ramanan [128], we define a part to be correctly localized when the majority of pixels covered by the estimated part have the correct labelling.

As Ramanan’s approach is completely inapplicable in this problem domain (see section 4.2.3 for experimental results on covered human body), we re-implement the algorithm by removing its global constraints and appearance modelling to improve the usability in evaluation. Two articulated models were used in MatchPose and Ramanan to represent the human configuration: a two-leg human model and a one-leg human model. The two-leg human model has six parts, corresponding to the head, torso and two parts per leg; the one-leg human model has four parts, corresponding to the head, torso and two parts for the leg. The decision of which model to use is based on the chamfer matching cost of the torso hypotheses (by selecting the model with the lowest cost). In addition, to generate the part templates, we manually marked the location of each part in twenty images.

## 4.6.2 Experimental Results

We compare RTPose and MatchPose with Ramanan using the recognition rate, which is obtained by calculating how often individual parts are correctly localized. The recognition rates of the head, torso, upper legs and lower leg pose are presented in Table 4.3.

The experimental results show that RTPose and MatchPose achieve high recognition rates and are not sensitive to illumination changes. The two proposed methods outperform Ramanan’s method. Some randomly selected outputs of RTPose and MatchPose are displayed in Figure 4.27, 4.28, 4.29 and 4.30 with misdetection examples by Ramanan in Figure 4.31 and by both methods in Figure 4.32. Figure 4.27 shows RTPose outputs of body postures with different occlusion levels – near-complete occlusion, partial occlusion and no occlusion. Figure 4.28 presents the RTPose outputs of various subjects different from the training video clip. Some results of MatchPose are shown in Figure 4.29 (with a different environment setting from the training data) and Figure 4.30 (with subjects different from the training data).

To summarize, RTPose is computationally efficient (able to process 30 frames per second), provides coarse pose estimation and can be improved by adding more leg pose templates; MatchPose produces fine pose estimation but requires 0.4 seconds for every  $320 \times 240$  frame, which can be improved by adding tracking algorithms. The methods assume subjects lying horizontally. In cases with heavy obscuration with strong noise, we observe that RTPose performs better because it utilizes latent features without shape matching and is better able to deal with less numerous and weaker features; pose matching methods however misdetect strong noise. Figure 4.31 shows the edge orientations of a heavy obscuration case and the erroneous detection by Ramanan, and some misdetections by MatchPose and RTPose are displayed in Figure 4.32. Thorough statistical tests were conducted and discussed in the next section.

## 4.6.3 Statistical Significance Test

To analyze whether there are significant differences in the performance of the methods, statistical significance tests were conducted. We first describe the statistical test method, and then present the results.

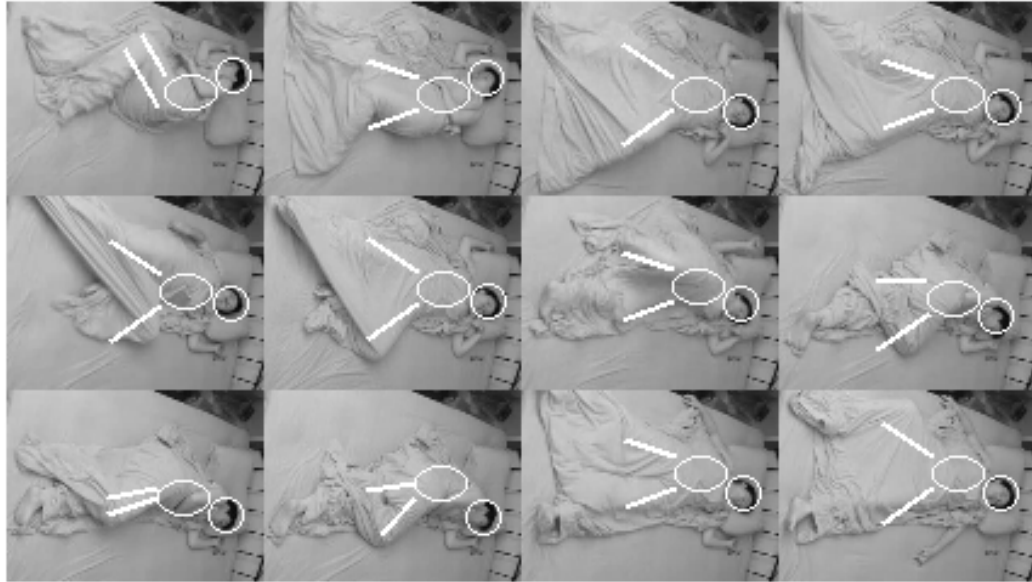
### 1 Statistical Test Method – McNemar’s test

To investigate what statistical test is more suitable to determine whether one method significantly performs better than another, Dietterich [38] eval-

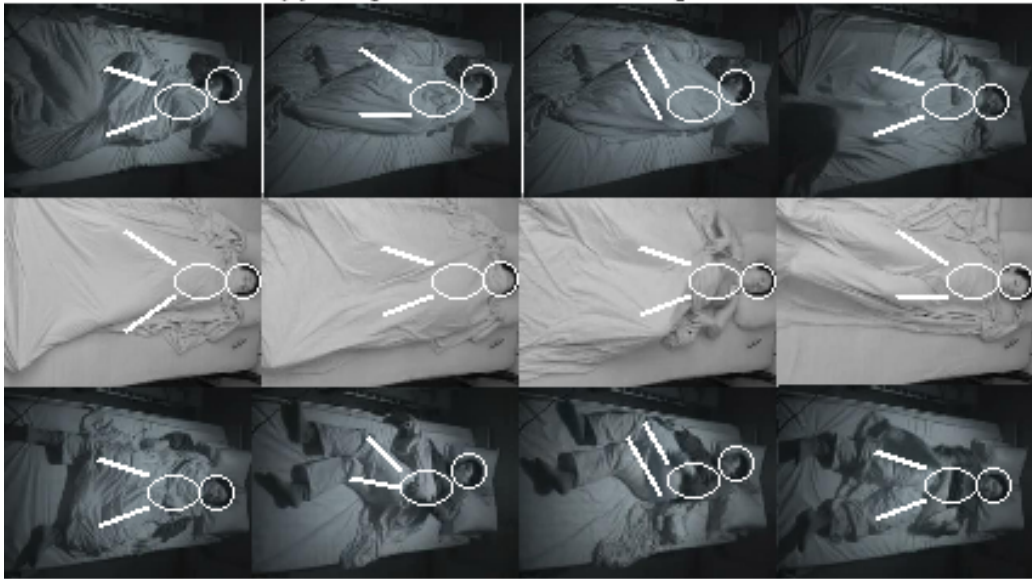
Table 4.3: Recognition Rates

All Images	Head	Torso	RUL	LUL	RLL	LLL
RTPose	0.89	0.95	0.74	0.71	N/A	N/A
MatchPose	0.90	0.90	0.72	0.77	0.61	0.72
Ramanan	0.66	0.59	0.5	0.38	0.26	0.52
High Illumination	Head	Torso	RUL	LUL	RLL	LLL
RTPose	0.88	0.95	0.72	0.7	N/A	N/A
MatchPose	0.89	0.87	0.73	0.76	0.65	0.72
Ramanan	0.57	0.31	0.31	0.21	0.07	0.41
Low Illumination	Head	Torso	RUL	LUL	RLL	LLL
RTPose	0.93	0.97	0.81	0.74	N/A	N/A
MatchPose	0.94	0.98	0.67	0.8	0.5	0.74
Ramanan	0.71	0.78	0.63	0.49	0.39	0.59
Cover	Head	Torso	RUL	LUL	RLL	LLL
RTPose	0.90	0.96	0.75	0.70	N/A	N/A
MatchPose	0.90	0.89	0.71	0.77	0.63	0.73
Ramanan	0.69	0.54	0.48	0.38	0.18	0.51
No Cover	Head	Torso	RUL	LUL	RLL	LLL
RTPose	0.8	0.93	0.7	0.8	N/A	N/A
MatchPose	0.94	0.97	0.75	0.8	0.45	0.69
Ramanan	0.53	0.8	0.6	0.37	0.6	0.53

RUL: right upper leg; LUL: left upper leg; RLL: right lower leg; LLL: left lower leg.



**(a) Body Rotation with various postures**



**(b) Various occlusion levels**

Figure 4.27: RTPose Outputs: (a) various poses with the same subject and environment setting as the training data (b) various occlusion levels with different environment setting from the training data

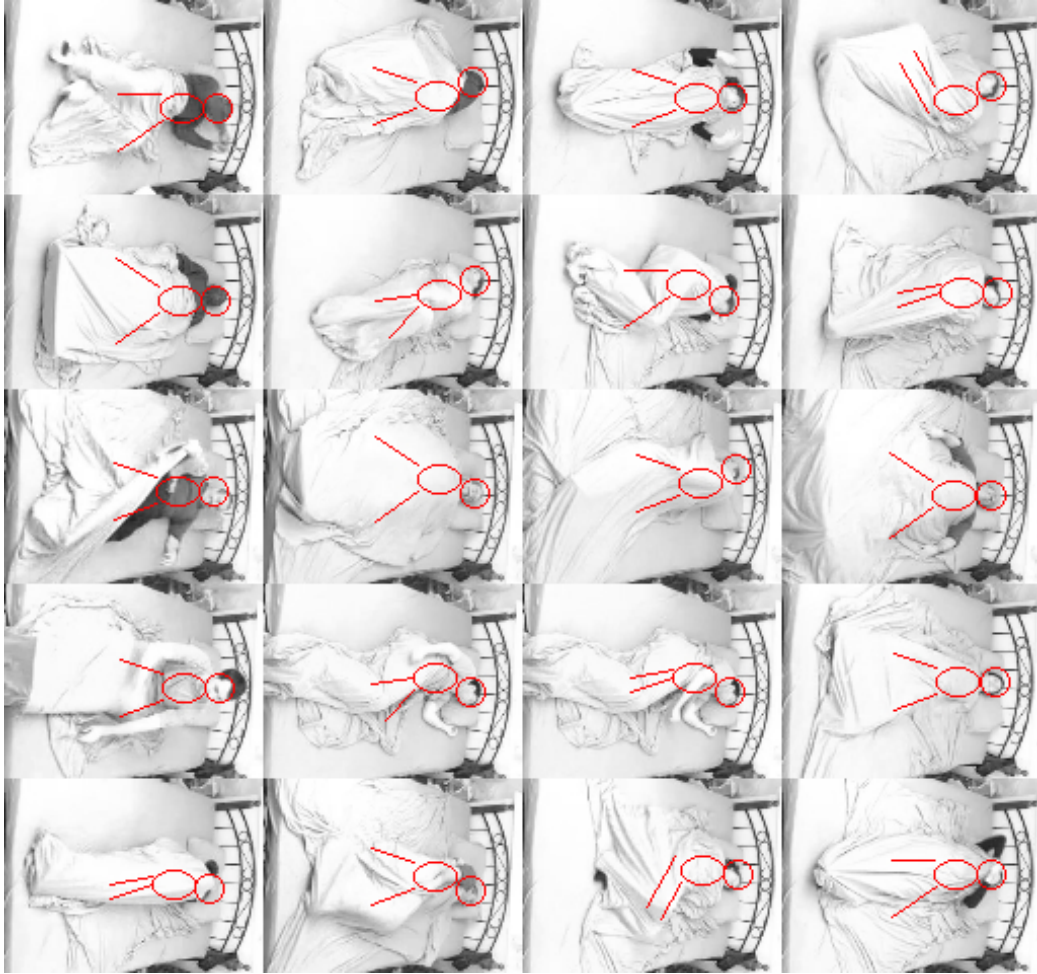


Figure 4.28: RTPose Outputs of eight different subjects, excluding the one used in the training video clip.

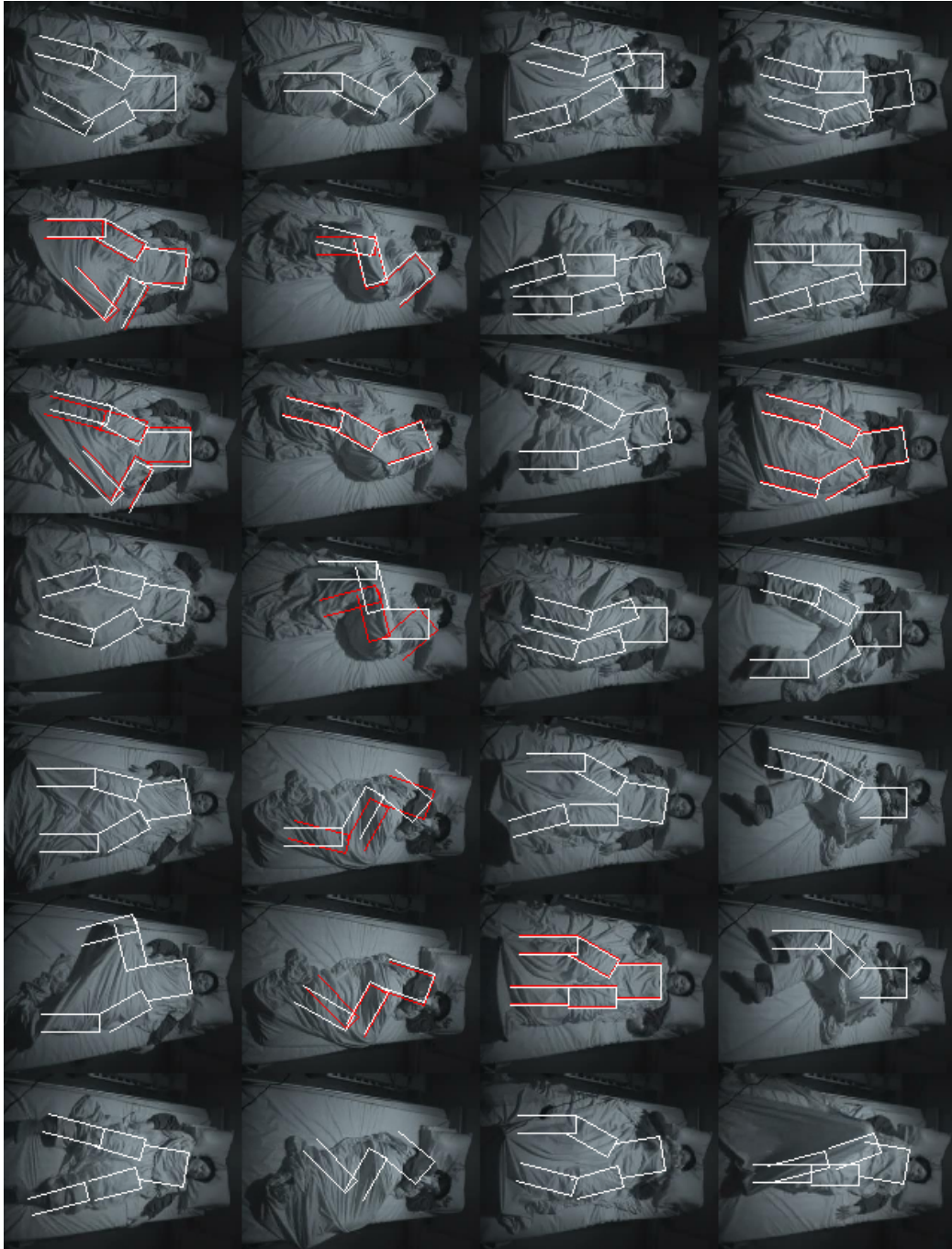


Figure 4.29: MatchPose Outputs of various poses are highlighted with white rectangles (and red rectangles if multiple configurations are obtained with the minimum chamfer matching cost), using the same subject as in the training video clip

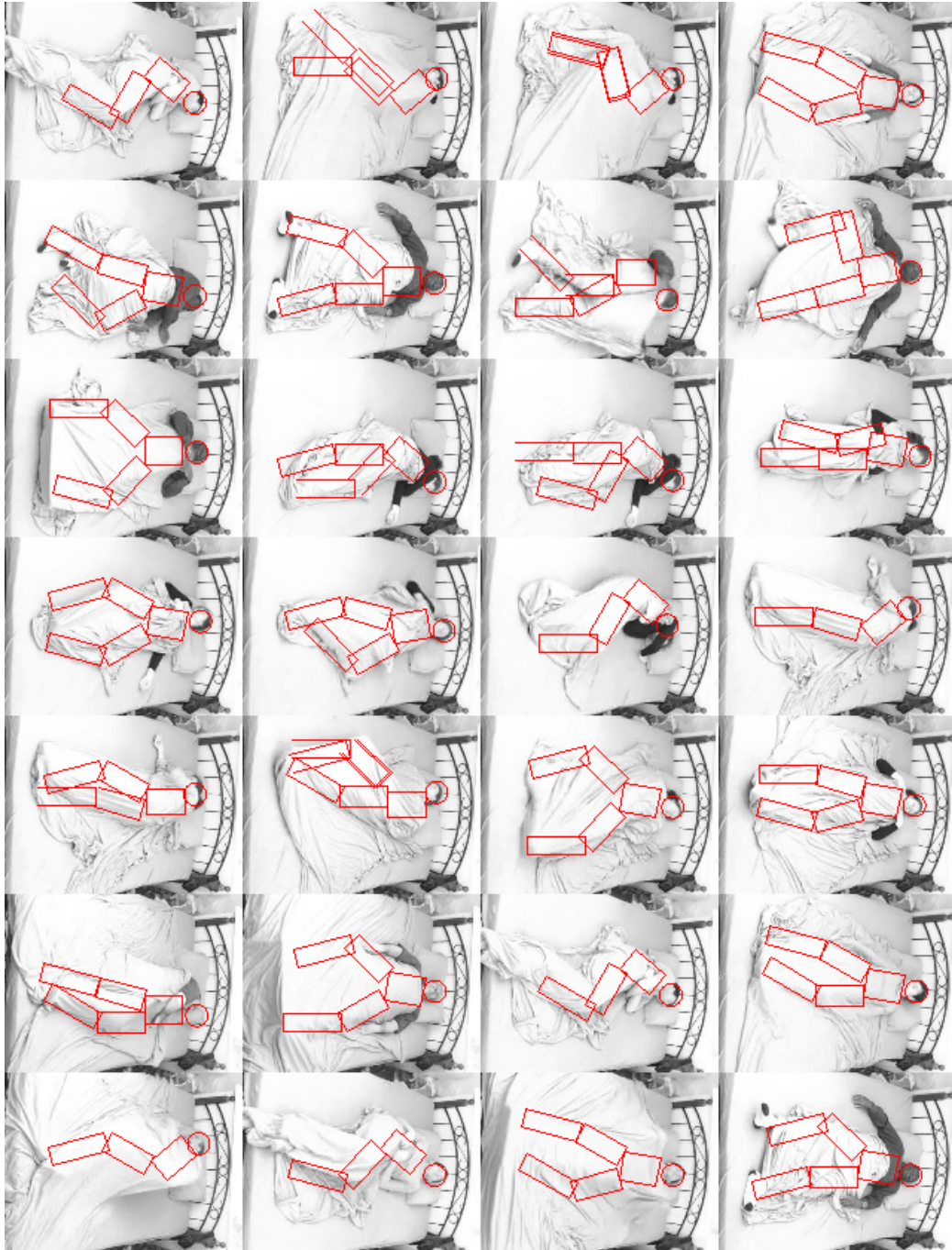


Figure 4.30: MatchPose Outputs on eight subjects, different from the one in the training video clip.

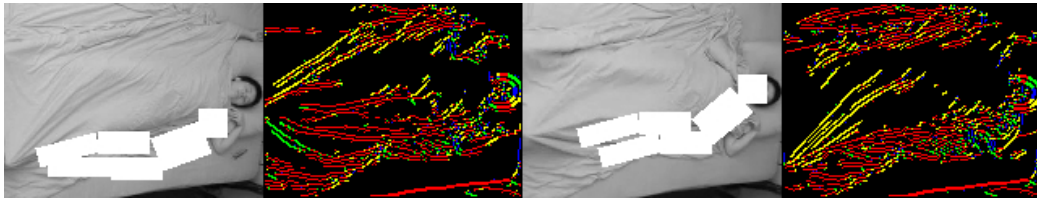


Figure 4.31: Edge Orientations of Heavily Obscured Data and Erroneous Detection by Ramanan.

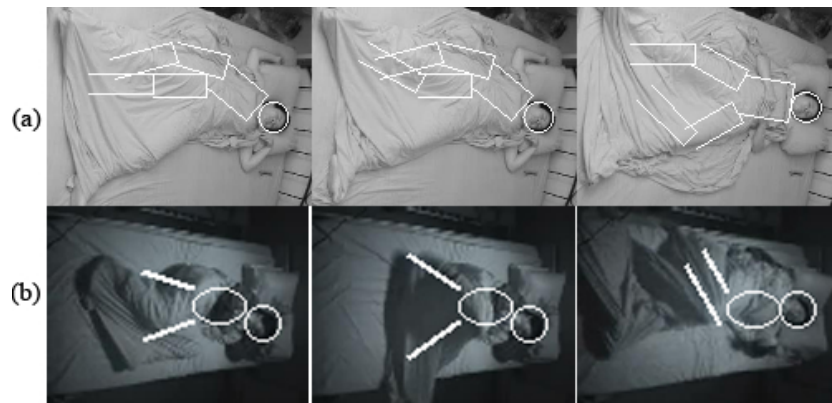


Figure 4.32: Misdetections by (a) MatchPose and (b) RTPose.

uated five statistical tests, including McNemar’s test, the resampled paired  $t$  test, the k-fold cross-validated paired  $t$  test, the 5x2cv paired  $t$  test and a simple test based on measuring the difference between the error rates of two algorithms. These statistical tests were compared experimentally to determine their probability of incorrectly detecting a difference when no difference exists (type **I** error). Dietterich suggested that McNemar’s test is the only test with acceptable Type **I** error for algorithms that can be executed only once. Hence, McNemar’s test is adopted in this work; a brief description is given below.

To compare two algorithms A and B, a contingency table is constructed using the number of their misclassifications.

#misclassifications by both	#misclassifications by A only
#misclassifications by B only	#misclassifications by neither

The following notation is used.

$n_{00}$	$n_{01}$
$n_{10}$	$n_{11}$

where  $n = n_{00} + n_{01} + n_{10} + n_{11}$  is the total number of testing examples.

McNemar’s test is based on a  $\chi^2$  test for goodness-of-fit that compares the distribution of counts expected under the null hypothesis to the observed counts. The null hypothesis is that the two algorithms have the same error rate, so that  $n_{01} = n_{10}$ ; the expected counts under the null hypothesis are:

$n_{00}$	$\frac{n_{01}+n_{10}}{2}$
$\frac{n_{01}+n_{10}}{2}$	$n_{11}$

Then, the statistic  $(\frac{(|n_{01}-n_{10}|-1)^2}{n_{01}+n_{10}})$  is approximately distributed as  $\chi^2$  with one degree of freedom, incorporating a continuity correction to account for the fact that the statistic is discrete while the  $\chi^2$  distribution is continuous. If the two algorithms have the same error rate, then the probability that this quantity is greater than  $\chi^2_{1,0.95} = 3.841459$  is less than 0.05. Hence, we can determine that the two algorithms have statistically significant different performances if  $(\frac{(|n_{01}-n_{10}|-1)^2}{n_{01}+n_{10}} > 3.841459)$ .

## 2 Statistical Test Results

The three paired methods were compared, including “RTPose vs Ramanan”, “MatchPose vs Ramanan”, and “RTPose vs MatchPose”, using different types of data (all images, high illumination, low illumination, with cover, without cover). The head, torso and a lower body part (LUL) were selected for statistical tests. Full statistical test results are presented in Appendix D, and the results are summarized in Table 4.4. The results show that RTPose and MatchPose outperform Ramanan.

There is a tradeoff between RTPose and MatchPose: RTPose runs real-time but does not provide lower leg pose; MatchPose provides fine pose estimation but costs 0.4 seconds to process a frame. Hence, if full body pose estimation is desirable, we recommend MatchPose. On the other hand, in the interests of computational speed, we recommend incorporating RTPose with motion information. Overall, for diagnosis of obstructive sleep apnoea, MatchPose is recommended to obtain fine pose estimation of obscured body for further activity recognition.

## 4.7 Conclusion

We have presented two monocular-video approaches for markerless pose estimation from a consistently fully or partially covered human without manual initialization: a robust pose matching model (MatchPose) that includes a novel weak human model (WHM) to accommodate the large variance of image features and a modified pose model (cwPose) adapted from a lateral walking pose detector [128] for people tracking; and a real time simple model (RTPose) containing WHM, a novel upper leg pose estimator, and a reinforcement tracker.

Experimental results demonstrate that the proposed two algorithms are able to identify the human configuration with various poses and occlusion levels, and they are recommended for different purposes. For diagnosis of OSA, we recommend MatchPose to obtain fine pose estimation of obscured human body for further activity recognition. In the future, we propose further investigation of obscured leg pose recognition, activity recognition using the obtained pose with motion, tracking, and the use of pose analysis in identifying OSA.

Table 4.4: Significance Performance Test Results

Head						
Rank	1	2	3	Torso		
All	$MA \approx RT$		RA	RT	MA	RA
HighI	$MA \approx RT$		RA	RT	MA	RA
LowI	$MA \approx RT$		RA	$MA \approx RT$		RA
Cover	$MA \approx RT$		RA	RT	MA	RA
NCov	MA	RT	RA	$MA \approx RT$		RA

LUL			
Rank	1	2	3
All	MA	RT	RA
HighI	MA	RT	RA
LowI	$MA \approx RT$		RA
Cover	$MA \approx RT$		RA
NCov	$MA = RT$		RA

The algorithms are ranked in this table. Rank 1–3: Best–Worst; HighI: High illumination; LowI: Low illumination; NCov: No cover; RT: RTPose; MA: MatchPose; RA: Ramanan.

# Chapter 5

## Conclusions

This thesis has investigated a new topic of video monitoring of breathing activity invariant to pose, camera view and occlusion, and an under-studied problem of pose estimation of covered human body. The thesis has made a number of significant contributions to the field of activity recognition and pose estimation. These contributions were required to build an automated video monitoring system in support of the diagnosis of OSA.

This chapter summarizes the main contributions of the thesis. In addition, we provide suggestions for areas that warrant future attention.

### 5.1 Summary of Contributions

#### 5.1.1 Monitoring of Breathing Activity

The literature review of Chapter 2 and 3 highlights the fact that to the author's best knowledge, there is no existing method suitable for video monitoring of breathing behavior of sleeping subjects.

The work presented in this thesis has shown that it is possible to analyze human breathing activity from video without special devices like thermal cameras, or constraints on posture or clothing. A new approach for recognizing abnormal breathing activity from video and assisting in diagnosis of obstructive sleep apnoea is presented. This approach avoids imposing positional constraints on the patient, and deals with fully or partially covered bodies. In addition, a novel motion detection model is built to capture subtle and cyclical breathing movements from video. An online spatial-temporal action template is introduced to capture the dynamic spatiotemporal shape of normal breathing activity, and adapts as the subject's pose changes. Furthermore, an action recognition approach is presented to detect abnormal

events and recognize abnormal breathing activities and limb movements.

This technique is real time and robust to heavy occlusion, variances of human breathing behavior and subject appearances, and substantial changes of camera view with respect to the subjects. Furthermore, shallow and abdominal breathing patterns do not affect the performance of the proposed approach, and this technique is not susceptible to illumination changes.

### **5.1.2 Pose Estimation of Covered Human Body**

Many existing approaches to pose estimation make simplifications to the measurement problem and work well given clear image cues (for simple detection models) or clean full body motion data (for dynamical models). Although there is some published research investigating the monitoring of partially occluded humans, the methods examined do not deal with pose estimation of consistently occluded subjects. To the author’s best knowledge, there is no previously published method to estimate pose from persistently covered human bodies.

This work introduces two novel monocular video approaches (MatchPose and RTPose) for full body pose recognition of the covered human body. Both methods are demonstrated to be able to recognize human poses with various postures and occlusion levels. They are recommended for different data types and purposes. For diagnosis of OSA, we recommend MatchPose to obtain fine pose estimation of obscured human body for further activity recognition. In the interests of computational speed, we recommend incorporating RTPose with motion information.

A robust Weak Human Model is introduced to effectively and efficiently identify the upper body poses from obscured human bodies, and a number of novel body part detectors are presented. Shape matching methods are reviewed, and a modification of the chamfer matching technique is presented to improve shape matching in cluttered scenes. A cascade of diverse models, an iterative boosting model, a low variance error boosting algorithm, and a reinforcement network are developed to improve pose estimation of obscured humans.

The pose estimation algorithms are not used in support of the diagnosis of OSA for now, but we propose to use the algorithms in conjunction with motion in future work, to develop a model-based action recognition approach in order to identify the human activities of medical interest, such as limb movements.

## 5.2 Future work

Model parametrization is a common issue in computer vision. Hence, it would be interesting to investigate automatic ways to obtain the parameter values. Further investigation of heavily obscured leg pose estimation and activity recognition using the obtained pose with motion can be fruitfully explored in the future. The analysis of postural changes and activities may prove to have diagnostic value in OSA and other conditions. It would also be interesting to extend the current work in this thesis to a broader domain. One intuitive extension would be to apply the newly developed methods to other breathing monitoring problems such as polygraph, sport training, early detection of sudden infant death syndrome in neonates, and patient monitoring, and to other obscured human monitoring domains such as surveillance.

# Bibliography

- [1] Agarwal, A., , Triggs, B.: 3D Human Pose from Silhouettes by Relevance Vector Regression. *Proceedings of Conference on Computer Vision and Pattern Recognition* **2** (2004) 882–888
- [2] Agarwal, A., , Triggs, B.: Tracking Articulated Motion using a Mixture of Autoregressive Models. *Proceedings of European Conference on Computer Vision* (2004).
- [3] Albu, A. B., , Beugeling, T.: A Three-Dimensional Spatiotemporal Template for Interactive Human Motion Analysis. *Journal of Multimedia* **2(4)** (2007) 45–54
- [4] Ambroise, C., , McLachlan, G. J.: Selection bias in gene extraction on the basis of microarray gene-expression data. *Proceedings of the National Academy of Sciences of the United States of America* **99(10)** (2002) 6562–6566.
- [5] Amin, H. M., , Cigada, M., , Fordyce, W. E., , Camporesi, E. M.: Experimental evaluation of a breath monitoring device. *Noninvasive monitoring of respiratory volume - Anaesthesia* **48(7)** (1993) 608–10
- [6] Amit, Y., , Blanchard G.: Multiple Randomized Classifiers. Technical report, University of Chicago (2001)
- [7] Ali, K. M., , Pazzani, M. J.: Error Reduction through Learning Multiple Descriptions. *International Journal of Machine Learning* **24** (1996) 173–202
- [8] Alon, U., , Barkai, N., , Notterman, D. A., , Gish, K., , Ybarra, S., , Mack, D., , Levine, A. J.: Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *National Academy of Science, Cell Biology* **96** (1999) 6745–6750

- [9] Armstrong, S. A., , Staunton, J. E. , , Silverman, L. B., , Pieters, R. et al.: MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. *Nature Genetics* **30** (2002) 41–47
- [10] Ash, A. A., , Michael, B. E., , Davis, R. E. , , Ma, C., , Izidore, S. L., , Andreas, R. et al.: Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403** (2000) 503–511
- [11] Appiah, K., , Hunter, A.: A Single-Chip FPGA Implementation of Real-time Adaptive Background Model. *Proceedings of IEEE Conference on Field-Programmable Technology* (2005) 95–102
- [12] Ballard, D. H., , Swain, M. J.: Color indexing. *International Journal of Computer Vision* **7(1)** (1991) 11–32
- [13] Barrow, H. G., , Tenenbaum, J. M., , Bolles, R. C., , Wolf, H. C.: Parametric correspondence and chamfer matching: Two new techniques for image matching. *Proceedings of International Joint Conference Artificial Intelligence* (1977) 659–663
- [14] Baudrier, E., , Millon, G., , Nicolier, F., , Ruan, S.: A New Similarity Measure Using Hausdorff Distance Map. *Proceedings of International Conference on Image Processing* (2004) 669–672
- [15] Bennett, L. S., , Langford, B. A., , Stradling, J. R., , Davies, R. J. O.: Sleep fragmentation indices as predictors of daytime sleepiness and nCPAP response in obstructive sleep apnea. *Am J Respir Crit Care Med* **158** (1998) 778–786
- [16] Blei, D. M., , Ng, A. Y., , Jordan, M. I.: Latent Dirichlet Allocation. *Journal of Machine Learning Research* **3** (2003) 993–1022
- [17] Bobick, A. F., , Davis, J. W.: The recognition of human movement using temporal templates. *IEEE Trans on Pattern Analysis and Machine Intelligence* **23(3)** (2001) 257–267
- [18] Borgefors, G: Hierarchical Chamfer Matching: A Parametric Edge Matching Algorithm. *IEEE Trans on Pattern Analysis and Machine Intelligence* **10(6)** (1988) 849–865
- [19] Borgefors, G: Distance Transformations in Digital Images. *Computer Vision, Graphics, and Image Processing* **34** (1986) 344–371
- [20] Borgefors, G: An improved version of the chamfer matching algorithm. *International conference Pattern Recognition* (1984) 1175–1177

- [21] Bouchrika, I., , Nixon, M. S.: Gait Recognition by Dynamic Cues. International conference Pattern Recognition (2008)
- [22] Boulton, T. E., , Micheals, R., , Gao, X., , Lewis, P., , Power, C., , Yin, W., , Erkan, A.: Frame-Rate Omnidirectional Surveillance and Tracking of Camouflaged and Occluded Targets. Proceedings of the Second IEEE Workshop on Visual Surveillance (1999) 48–55
- [23] Bauer, E., , Kohavi, R.: An empirical comparison of voting classification algorithms: Bagging, boosting, and variants. International Journal of Machine Learning **36** (1999) 105–139
- [24] Breiman, L.: Bias, Variance, and Arcing Classifiers. Technical report 460, Statistics Department, UC Berkeley (1996)
- [25] Breiman, L.: Bagging predictors. International Journal of Machine Learning **24** (1996) 134–140
- [26] Carr, G.: Mechanics of sport: a practitioner’s guide. Human Kinetics (1997).
- [27] Chalmond, B., , Francesconi, B., , Herbin, S.: Using Hidden Scale for Salient Object Detection. IEEE Transactions on Image Processing (2004) 2644–2656
- [28] Chekmenev, S. Y., , Rara, H., , Farag, A. A.: Non-contact, Wavelet-based Measurement of Vital Signs using Thermal Imaging. International Journal of Graphics, Vision and Image Processing **6** (2005) 25–30
- [29] Cheng, C. -M., , Hsu, Y. -L., , Young, C. -M., , Wu, C. -H.: Development of a portable device for tele-monitoring of snoring and OSAS symptoms. Telemedicine and e-Health **14**(1) (2008) 55–68
- [30] Chest: Chest Medicine. <http://www.priory.com/chest.htm> (2007)
- [31] Collins, R. T., , Liu, Y., , Leordeanu, M.: Online Selection of Discriminative Tracking Features. IEEE Transactions on Pattern Analysis and Machine Intelligence **27** (2005) 1631–1643
- [32] Coughlin, S.R., , Mawdsley, L., , Mugarza, J. A. et al.: Obstructive sleep apnoea is independently associated with an increased prevalence of metabolic syndrome. Eur Heart J **25** (2004) 735–741

- [33] Cutler, R., , Davis, L.S.: Robust real-time periodic motion detection, analysis, and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2** (2000) 781–796
- [34] Catherine, L. N., , Mani, D. R. , , Rebecca, A. B., , Pablo, T. et al.: Gene expression-based classification of malignant gliomas correlates better with survival than histological classification. *Cancer Research* **63** (2003) 1602–1607
- [35] Dasgupta, S., , Long, P. M.: Boosting with diverse base classifiers. *Proceedings of the Conference on Computational Learning Theory* (2003) 273–287
- [36] Dettling, M.: BagBoosting for tumor classification with gene expression data. *Bioinformatics* **20(18)** (2004) 3583–3593
- [37] Dinesh, S., , Phillip, G. F., , Kenneth, R., , Donald, G. J., , Judith, M. et al.: Gene expression correlates of clinical prostate cancer behavior. *Cancer Cell* **1** (2002) 203–209
- [38] Dietterich, T. G.: Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms. *Neural Computation* **10** (1998) 1895–1923
- [39] Domingo, C. , , Watanabe, O.: MadaBoost: A modification of AdaBoost. *Technical Reports on Mathematical and Computing Sciences TR-C138* (2000)
- [40] Deutscher, J., , Reid, I.: Articulated Body Motion Capture by Stochastic Search. *International Journal of Computer Vision* **2** (2005) 185–205
- [41] Deutscher, J., , Davison, A., , Reid, I.: Automatic Partitioning of High Dimensional Search Spaces Associated with Articulated Body Motion Capture. *Computer Vision and Pattern Recognition* **2** (2001) 669–676
- [42] Dollar, P., , Rabaud, V., , Cottrell, G., , Belongie, S.: Behavior recognition via sparse spatio-temporal features. *Proceedings of Visual Surveillance and Performance Evaluation of Tracking and Surveillance* (2005) 65–72
- [43] Efros, A. A., , Berg, A. C., , Mori, G., , Malik, J.: Recognizing Action at a Distance. *Proceedings of International Conference on Computer Vision* (2003) 726–733
- [44] Everitt, B. S.: The analysis of contingency tables. Chapman and Hall, London (1977)

- [45] Elgammal, A., , Lee, C.S.: Inferring 3D body pose from silhouettes using activity manifold learning. *Proceedings of the Conference on Computer Vision and Pattern Recognition* **2** (2004) 681–688
- [46] Eng, H. L., , Wang, J., , Kam, A. H., , Yau, W. Y.: A Bayesian framework for robust human detection and occlusion handling using human shape model. *Proceedings of the 17th International Conference on Pattern Recognition* (2004) 257–260
- [47] Felzenszwalb, P.F., , Huttenlocher, D.P.: Pictorial Structures for Object Recognition. *International Journal of Computer Vision* **61(1)** (2005) 55–79
- [48] Fischler, M.A., , Elschlager, R.A.: The Representation and Matching of Pictorial Structures. *IEEE Trans. Computers* **22(1)** (1973) 67–92
- [49] Flemons, W. W., , Littner, M. R., , Rowley, J. A., , Gay, P. et al.: Home Diagnosis of Sleep Apnea: A Systematic Review of the Literature An Evidence Review the American Thoracic Society *CHEST* **124(4)** (2003) 1543–1579
- [50] Fleuret, F., , Geman, D.: Coarse-to-Fine Face Detection *International Journal of Computer Vision* **41(1)** (2001) 85–107
- [51] Freund, Y., , Schapire, R. E.: A Decision-Theoretic Generalization of On-Line Learning and an Applicatin to Boosting *Computer and System Sciences* **55** (1997) 119–139
- [52] Freund, Y., , Schapire, R.: Experiments with a new boosting algorithm. *Proceedings of the Thirteenth International Conference on Machine Learning*, San Francisco (1996) 148–156
- [53] Freund, Y.: An Adaptive Version of the Boost by Majority Algorithm. *International Journal of Machine Learning* **43(3)** (2001) 293–318
- [54] Friedman, J.: Stochastic Gradient Boosting. *Computational Statistics and Data Analysis* **38** (2002) 367–368
- [55] Friedman, J. H., , Hastie, T. , , Tibshirani, R.: Additive logistic regression: A statistical view of boosting. *The Annals of Statistics* **28** (2000) 337–374
- [56] Gavin, J. G., , Roderick, V. J., , Li-Li, H., , Steven, R. G., , Joshua, E. B., , Sridhar, R., , William, G. R., , David, J. S., , Raphael, B.: Translation of Microarray Data into Clinically Relevant Cancer Diagnostic Tests

- Using Gene Expression Ratios in Lung Cancer and Mesothelioma. *Cancer Research* **62** (2002) 4963–4967
- [57] Golub, T. R., , Slonim, D. K., , Tamayo, P., , Huard, C., , Gaasenbeek, M. et al.: Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* **286** (1999) 531–537
- [58] Gao, J., , Collins, R. T., , Hauptmann, A. G., , Wactlar, H. D.: Articulated Motion Modeling for Activity Analysis. *Proceedings of the international Conference on Computer Vision and Pattern Recognition Workshop* **1** (2004) 20–28
- [59] Gastaut, H., , Tassinari, C.A., , Duron, B.: Polygraphic study of the episodic diurnal and nocturnal (hypnic and respiratory) manifestations of the Pickwick syndrome. *Brain Res* **1** (1966) 167–186
- [60] Gavrilu, D.: The visual analysis of human movement: A survey. *Comput. Vis. Image Understanding* **73(1)** (1999) 82–98
- [61] Gavrilu, D.: Pedestrian Detection from a Moving Vehicle. *Proceedings of the 6th European Conference on Computer Vision* **2** (2000) 37–49
- [62] Ghafoor, A., , Iqbal, R. N., , Khan, S.: Robust image matching algorithm. *Proceedings of Video Image Processing and Multimedia Communications 4th EURASIP Conference* **1** (2003) 155–160
- [63] Gibson, G. J.: Obstructive sleep apnoea syndrome: underestimated and undertreated. *British Medical Bulletin* **72** (2004) 49–64
- [64] Golpe, R., , Jimenez, A., , Carpizo, R. et al.: Utility of home oximetry as a screening test for patients with moderate to severe symptoms of obstructive sleep apnea. *Sleep* **22(7)** (1999) 932–937
- [65] Gonzalez, R. C., , Woods, R. E.: *Digital Image Processing*. Massachusetts Addison-Wesley (1992) 585
- [66] Gorelick, L., , Blank, M., , Shechtman, E., , Irani, M., , Basri, R.: Actions as Space-Time Shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **29(12)** (2007) 2247–2253
- [67] Guo, F., , Qian, G.: Learning and Inference of 3D Human Poses from Gaussian Mixture Modeled Silhouettes. *Proceedings of the international Conference on Pattern Recognition* **2** (2006) 43–47

- [68] Haba-Rubio, J., , Stane, L., , Krieger, J., , Macher, J. P.: Periodic limb movements and sleepiness in obstructive sleep apnea patients. *Sleep Medicine* **6** (2005) 225–229
- [69] Hoey, J.: Tracking using Flocks of Features, with Application to Assisted Handwashing. *Proceedings of British Machine Vision Conference* **1** (2006) 367–376
- [70] Hoffstein, V., , Mateika, S., , Nash, S.: Comparing perceptios and measurements of snoring. *Sleep* **19(10)** (1996) 783–789
- [71] Hofmann, T.: Probabilistic Latent Semantic Analysis. *Proceedings of international ACM SIGIR conference on research and development in information retrieval* (1999) 50–57
- [72] Hossain, J., , Shapiro, C.: The prevalence, cost implications, and management of sleep disorders: An overview. *Sleep Medicine Reviews* **6(2)** (2002) 85–99
- [73] Hu, M.: Visual Pattern Recognition by Moment Invariants. *IRE Trans. Information Theory* **8(2)** (1962) 179–187
- [74] Hua, G., , Yang, M.-H., , Wu, Y.: Learning to Estimate Human Pose with Data Driven Belief Propagation. *Proceedings of the International Conference on Computer Vision and Pattern Recognition* **2** (2005) 747–754
- [75] Huang, Z. Q., , Jiang, Z.: Tracking Camouflaged Objects with Weighted Region Consolidation. *Proceedings of the IEEE Digital Image Computing Technqiues and Applications* (2005) 161–168
- [76] Huang, C., , Ai, H., , Wu, B., , Lao, S.: Boosting nested cascade detector for multi-view face detection. *Proceedings of the International Conference on Pattern Recognition* **2** (2004) 415–418
- [77] Huang, C., , Ai, H., , Wu, B., , Lao, S.: Vector boosting for rotation invariant multi-view face detection. *Proceedings of the International Conference on Computer Vision* **1** (2005) 446–453
- [78] Hunsaker, D. H., , Riffenburgh, R. H.: Snoring significance in patients undergoing home sleep studies *OtolaryngologyHead and Neck Surgery* **134** (2006) 756–760

- [79] Huttenlocher, D. P., , Klanderman, G. A., , Rucklidge, W. J.: Comparing Images Using Hausdorff Distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **15**(9) (1993) 850–863
- [80] Hyun, Medics: Baby breath monitor. <http://hyun.en.ecplaza.net/14.asp> (2006)
- [81] Jaeggli, T., , Caenen, G., , Fransens, R., , Gool, L. V.: Analysis of Human Locomotion Based on Partial Measurements. *Proceedings of IEEE Motion* (2005) 248–253
- [82] Javaheri, S., , Abraham, W. T., , Brown, C. et al.: Prevalence of obstructive sleep apnoea and periodic limb movement in 45 subjects with heart transplantation. *European Heart Journal* **25** (2004) 260–266
- [83] Jenkins, O.C., , Gonzlez, G., , Loper, M.: Tracking human motion and actions for interactive robots. *Proceeding of Human-Robot interaction* **25** (2007) 365–372
- [84] Johnsson, G.: Visual motion perception. *Scientific American* (1975) 76–88
- [85] Kanagala, R., , Murali, N. S., , Friedman, P.A. et al.: Obstructive sleep apnea and the recurrence of atrial fibrillation. *Circulation* **107**(20) (2003) 2589–94
- [86] Kaneko, Y., , Floras, J. S., , Usui, K. et al.: Cardiovascular effects of continuous positive airway pressure in patients with heart failure and obstructive sleep apnea. *N Engl J Med* **348**(13) (2003) 1233–41
- [87] Kaneshiro, N. K.: Heart-respiratory monitor - infants. A.D.A.M. <http://www.nlm.nih.gov/medlineplus/ency/article/007236.htm> (2005)
- [88] Ke, Y., , Sukthankar, R., ,Hebert, M.: Efficient visual event detection using volumetric features *Proceeding of International Conference on Computer Vision* **1** (2005) 166–173
- [89] Kingshott, R. N., , Vennelle, M., , Hoy, C. J., , Engleman, H. M. et al.: Predictors of improvements in daytime function outcomes with CPAP therapy. *Am J Respir Crit Care Med* **161** (2000) 866–871
- [90] Kohavi, R., , Provost, F.: Special Issue on Applications of Machine Learning and the Knowledge Discovery Process. *Machine Learning* **30** (1998) 271–274

- [91] Kohavi, R., , Wolpert, D.: Bias plus variance decomposition for zero-one loss functions. *Proceedings of the Thirteenth International Machine Learning Conference*.
- [92] Kuncheva, L. I.: Diversity in multiple classifier systems. *Information Fusion* **6** (2005) 3–4
- [93] Kushida, C. A., , Kushida, M. R., , Morgenthaler, T. et al.: Practice Parameters for the Indications for Polysomnography and Related Procedures An Update for 2005. *SLEEP* **28(4)** (2005) 499–519
- [94] Lan, X., , Huttenlocher, D.: A unified spatio-temporal articulated model for tracking. *Proceedings of Computer Vision and Pattern Recognition* **1** (2004) 722–729
- [95] Lee, M.W., , Cohen, I.: A Model-Based Approach for Estimating Human 3D Poses in Static Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6** (2006) 905–916
- [96] Lee, M.W., , Nevatia, R.: Body Part Detection for Human Pose Estimation and Tracking. *IEEE Workshop on Motion and Video Computing* (2007)
- [97] Li, B., , Meng, Q., , Holstein, H.: Articulated motion reconstruction from feature points. *Pattern Recognition* **41** (2008) 418–431
- [98] Li, B., , Sezan, M. I.: Adaptive Video Background Replacement. *Proceedings of IEEE International Conference on Multimedia and Expo* (2001) 269–272
- [99] Liebe, B., , Seemann, E., , Schiele, B.: Pedestrian Detection in Crowded Scenes. *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition* **1** (2005) 878–885
- [100] Liu, J., , Shah, M.: Learning Human Actions via Information Maximization. *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition* (2008) 1–8
- [101] Lipton, A.: Local Application of Optic Flow to Analyse Rigid versus Non-Rigid Motion. *ICCV Workshop on Frame-Rate Vision* (1999)
- [102] Long, P.M., , Vega, V.B.: Boosting and Microarray Data. *International Journal of Machine Learning* **52** (2003) 31–44

- [103] Lv, F., , Nevatia, R.: Single View Human Action Recognition using Key Pose Matching and Viterbi Path Searching. *Proceedings of Computer Vision and Pattern Recognition* (2007) 1–8
- [104] Mack, D. C., , Kell, S. W., , Alwan, M., , Turner, B., , Felder, R. A.: Non-invasive analysis of physiological signals - a vibration sensor that passively detects heart and respiration rates as part of a sensor suite for medical monitoring. *Summer Bioengineering conference Florida* (2003)
- [105] Mack, D., , Alwan, M., , Turner, B., , Suratt, P., , Felder, R.: A Passive and Portable System for Monitoring Heart Rate and Detecting Sleep Apnea and Arousals - Preliminary Validation. *Proceedings of the Transdisciplinary Conference on Distributed Diagnosis and Home Healthcare* (2006)
- [106] Makarov, A.: Comparison of Background Extraction Based Intrusion Detection Algorithms. *Proceedings of International Conference on Image Processing* (1996) **1** 521–524
- [107] Matusiewicz, S., , Gravill, N.: Personal Interview. Consultant Physicians in the Medical Physics Department of Lincoln County Hospital in United Kingdom (2006)
- [108] McKenna, S.J., , Nait-Charif, H.: Tracking human motion using auxiliary particle filters and iterated likelihood weighting. *International Journal of Computer Vision* **25** (2007) 852–862
- [109] Moeslund, T.B., , Granum, E.: A Survey of Computer Vision-Based Human Motion Capture. *International Journal of Computer Vision and Image Understanding* **81** (2001) 231–268
- [110] Moody, G.B., , Mark, R.G., , Bump, M.A., , Weinstein, J.S. et al.: Clinical Validation of the ECG-Derived Respiration (EDR) Technique. *Computers in Cardiology* **13** (1986) 507–510
- [111] Murthy, R., , Pavlids I., , Tsiamyrtzis, P.: Touchless Monitoring of Breathing Function. *Proceeding of the 26th Annual International Conference of the IEEE EMBS* (2004) 1196–1199
- [112] Neven, A. K., , Middelkoop, H. A. M., , Kemp, B., , Kamphuisen, H.A.C., , Springer, M.P.: The prevalence of clinically significant sleep apnoea syndrome in the Netherlands. *Thorax* **53** (1998) 638–642

- [113] Newman, A. B., , Nieto, F.J., , Guidry, U. et al.: Relation of sleep-disordered breathing to cardiovascular disease risk factors: The Sleep Heart Health Study. *Am J Epidemiol* **154** (2001) 50–59
- [114] Nguyen, H. T., , Smeulders, A. W. M.: Fast Occluded Object Tracking by a Robust Appearance Filter. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **26** (2004) 1099–1104
- [115] Niebles, J.C., , Li, F.-F.: A Hierarchical Model of Shape and Appearance for Human Action Classification. *Proceeding of Computer Vision and Pattern Recognition* (2007) 1–8
- [116] Niebles, J.C., , Wang, H., , Li, F.-F.: Unsupervised Learning of Human Action Categories Using Spatial-Temporal Words. *International Journal of Computer Vision* **79** (2008) 299–318
- [117] Nillius, P., , Eklundh, J.-O.: Fast Block Matching with Normalized Cross-Correlation using Walsh Transforms. Report ISRN KTHNAP–0211–SE (2002)
- [118] Nishida, Y., , Hori, T., , Suehiro, T., , Hirai, S.: Monitoring of Breath Sound under Daily Environment by Ceiling Dome Microphone. *IEEE International Conference on System, Man and Cybernetics* (2000) 1822–1829
- [119] Nixon, M., , Aguado, A.: *Feature Extraction and Image Processing*. 105, Newnes, Oxford (2002)
- [120] Ng, A. K. , , Wong, K. Y., , Tan, C. H., , Koh, T. S.: Bispectral Analysis of Snore Signals for Obstructive Sleep Apnea Detection. *Conference of the IEEE EMBS* (2007) 6195–6198
- [121] Olson, C.F., , Huttenlocher, D.P.: Automatic target recognition by matching oriented edge pixels. *IEEE Transactions on Image Processing* **6(1)** (1997) 103–113
- [122] OSAOnline: OSAOnline. Retrieved November 17, 2008 from <http://www.osaonline.com>.
- [123] Pepperell, J. C., , Davies, R. J., , Stradling, J. R.: Sleep studies for sleep apnoea. *Physiological Measurement* **23** (2002) 39–74
- [124] Pomeroy, S. L., , Tamayo, P., , Gaasenbeek, M., , Sturla, L. M. et al.: Prediction of central nervous system embryonal tumour outcome based on gene expression. *Nature* **415** (2002) 436–442

- [125] Punjabi, N. M., , Sorkin, J. D., , Katzel, L. I. et al.: Sleep disordered breathing and insulin resistance in middle-aged and overweight men. *Am J Respir Crit Care Med* **165** (2002) 677–682
- [126] Puvanendran K, Goh KL. From snoring to sleep apnea in a Singapore population. *Sleep Res Online* **2** (1999) 11-14
- [127] Quinlan, J. R.: Bagging, boosting and c4.5. *Proceedings of the thirteenth national conference on artificial intelligence* (1996) 725–730.
- [128] Ramanan, D., , Forsyth, D.A., ,Zisserman, A.: Tracking People by Learning their Appearance. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **29** (2007) 65–81
- [129] Ramanan, D., , Forsyth, D.A.: Finding and tracking people from the bottom up. *Proceedings of Computer Vision and Pattern Recognition* **2** (2003) 467–474
- [130] Ramanan, D., , Forsyth, D.A.: Using temporal coherence to build models of animals. *Proceedings of international conference on Computer Vision* (2003) 338–345
- [131] Ramanan, D.: Learning to parse images of articulated bodies. *Advances in Neural Information Processing Systems* **19** (2007) 1129–1136
- [132] Ramanan, D.: Web Homepage of Deva Ramanan. <http://www.ics.uci.edu/~dramanan>
- [133] Ran, Y., , Weiss, I., , Zheng, Q., , Davis, L. S.: Pedestrian Detection via Periodic Motion Analysis. *International Journal of Computer Vision* **71(2)** (2007) 143-160
- [134] Randall, D. P.: Remote Respiratory Monitor. *Proceedings of the 8th Annual IEEE Symposium on Computer-Based Medical Systems* (1995) 204–211
- [135] Raytek, Corp.: Fluke ThermoView Ti30 thermal imager. Available: <http://www.radir.com/thermalimager.htm>
- [136] Ren, X., ,Berg, A.C., ,Malik, J.: Recovering Human Body Configurations Using Pairwise Constraints between Parts. *Proceedings of the international Conference on Computer Vision* **1** (2003) 824–831

- [137] Roberts, R.: Sleep labs add beds in awakened market. *Kansas City Business Journal* (2007) Retrieved December 10, 2008, from <http://kansascity.bizjournals.com/kansascity/stories/2007/01/22/story5.html>
- [138] Schapire, R.E., , Singer, Y.: Improved boosting algorithms using confidence-rated predictions. *Machine Learning* **37** (1999) 297–336
- [139] ScienceDaily: Sleep Apnea Increases Risk Of Heart Attack Or Death By 30 Percent. *American Thoracic Society* (2007) Retrieved November 17, 2008, from <http://www.sciencedaily.com/releases/2007/05/070520183533.htm>
- [140] ScienceDaily: Obstructive Sleep Apnea Causes Earlier Death In Stroke Patients, Study Finds. *American Thoracic Society* (2008) Retrieved November 17, 2008, from <http://www.sciencedaily.com/releases/2008/05/080518182655.htm>
- [141] ScienceDaily: Eye Conditions Linked With Obstructive Sleep Apnea. *Mayo Clinic* (2008) Retrieved November 17, 2008, from <http://www.sciencedaily.com/releases/2008/11/081110154040.htm>
- [142] Sibel, N. T., , Maybank, S. J.: Fusion of multiple tracking algorithms for robust people tracking. *Proceedings of European Conference on Computer Vision* **4** (2002) 373–387
- [143] Sigal, L., , Bhatia, S., , Roth, S., , Black, M., , Isard, M.: Tracking loose-limbed people. *Proceedings of the Computer Vision and Pattern Recognition* **1** (2004) 421–428
- [144] Sivan, Y., , Kornecki, A., , Schonfeld, T.: Screening obstructive sleep apnoea syndrome by home videotape recording in children. *European Respiratory Journal* **9** (1996) 2127–2131
- [145] Sjostrom, C., , Lindberg, E., , Elmasry, A., , Hagg, A., , Svardsudd, K., , Janson, C.: Prevalence of SA and snoring in hypertensive men: a population based study. *Thorax* **57** (2002) 602–607
- [146] Sminchisescu, C., , Kanaujia, A., , Li, Z., , Metaxas, D.: Discriminative Density Propagation for 3D Human Motion Estimation. *Proceedings of Computer Vision and Pattern Recognition* **1** (2005) 390–397
- [147] Stauffer, C., , Grimson, W.E.L.: Adaptive background mixture models for real-time tracking. *Proceedings of Computer Vision and Pattern Recognition* **2** (1999) 2246–2252

- [148] Stone, L.D., , Corwin, T.L., , Barlow, C.A.: Bayesian Multiple Target Tracking. Artech House (1999)
- [149] Storck, K., , Karlsson, M., , Ask P., , Loyd, D.: Heat Transfer Evaluation of the Nasal Thermistor Technique. IEEE Transactions on Biomedical Engineering **43(12)** (1996) 1187–1191
- [150] Svetlana, I., , Mammo, H. Y., , John, W. A., , Michael, E. H. et al.: A gated deep inspiration breath-hold radiation therapy technique using a linear position transducer. Applied Clinical Medical Physics **6(1)** (2005) 61–70
- [151] Tan, A.C. , , Gilbert, D.: Ensemble machine learning on gene expression data for cancer classification. Applied Bioinformatics **2** (2003) S75–S83
- [152] Thayananthan, A., , Stenger, B., , Torr, P., , Cipolla, R.: Shape Context and Chamfer Matching in Cluttered Scenes. Computer Vision and Pattern Recognition **1** (2003) 1–1
- [153] Tobias, J., , Geert, C., , Rik, F., , Van, G.L.: Analysis of Human Locomotion based on Partial Measurements. Proceedings of IEEE Workshop on Motion and Video Computing **2** (2005) 248–253
- [154] Udwadia, Z. F., , Doshi, A. M., , Lonkar, S. G., , Singh, C. I.: Prevalence of Sleep-disordered Breathing and Sleep Apnea in Middle-aged Urban Indian Men. American Journal of Respiratory and Critical Care Medicine **169** (2004) 168–173
- [155] Valstar, M., , Pantic, M., , Patras, I.: Motion History for Facial Action Detection in Video. Proceedings of IEEE International Conference on Systems, Man and Cybernetics (2004) 635–640
- [156] Veer, L. J., , Dai, H., , van de Vijver, M. J., , He, Y. D., , Hart, A. A., , Mao, M., Peterse, H. L. et al.: Gene expression profiling predicts clinical outcome of breast cancer. Nature **415(6871)** (2002) 484–5
- [157] Viola, P., , Jones, M.: Rapid object detection using a boosted cascade of simple features. Proceedings of IEEE CVPR Conference (2001) 511–518
- [158] Visi: Visi-3 Digital Video System.  
<http://www.stowood.co.uk/page26.html>

- [159] Wang, C.-W., , Hunter, A.: A Novel Approach to Detect the Obscured Upper Body in application to Diagnosis of Obstructive Sleep Apnea. *IAENG International Journal of Computer Science* **35** (2008) 110–118
- [160] Wang, C.-W., , Ahmed, A., , Hunter, A.: Locating the Upper Body of Covered Humans in application to Diagnosis of Obstructive Sleep Apnea. *Proceedings of World Congress on Engineering* **2** (2007) 662–667
- [161] Wang, C.-W.: Real Time Sobel Square Edge Detector for Night Vision Analysis. *Proceedings of International Conference on Image Analysis and Recognition, Lecture Notes in Computer Science* (2006) 404–413
- [162] Wang, C.-W., , Hunter, A.: The Detection of Abnormal Breathing Activity by Vision Analysis in application to Diagnosis of Obstructive Sleep Apnea. *Encyclopedia of Healthcare Information Systems* **1** (2008) 416–424
- [163] Wang, C.-W., , Ahmed, A., , Hunter, A.: Vision Analysis in Detecting Abnormal Breathing Activity in application to Diagnosis of Obstructive Sleep Apnoea. *Proceedings of the 28th international conference of the IEEE Engineering in Medicine and Biology Society* (2006) 4469–4473
- [164] Wang, H., , Parker, J. D., , Newton, G. E., , Floras, J. S., , Mak, S., , Chiu, K. L. et al.: Influence of obstructive sleep apnea on mortality in patients with heart failure. *J Am Coll Cardiol* **49** (2007) 1625–1631
- [165] Wang, C.-W., , Hunter, A.: A Low Variance Error Boosting Algorithm. *International Journal of Applied Intelligence*, Springer Netherlands, Published online: 21 Feb 2009
- [166] Warmuth, M. K., , Liao, J., , Ratsch, G: Totally corrective boosting algorithms that maximize the margin. In *Proceedings of the 23rd international Conference on Machine Learning* **148** (2006) 1001–1008
- [167] Webb, G. I.: MultiBoosting: A Technique for Combining Boosting and Wagging. *International Journal of Machine Learning* **40** (2000) 159–196
- [168] Winn, J., , Shotton, J.: The Layout Consistent Random Field for Recognizing and Segmenting Partially Occluded Objects. *Proceedings of IEEE Computer Vision and Pattern Recognition* (2006) 37–44
- [169] Wixson, L.: Detecting salient motion by accumulating directionary-consistent flow. *IEEE Trans. Pattern Anal. Machine Intell* **22** (2000) 774–780

- [170] Wu, B., , Nevatia, R.: Detection and Tracking of Multiple, Partially Occluded Humans by Bayesian Combination of Edgelet based Part Detectors. *International Journal of Computer Vision* **2** (2007) 247–266
- [171] Wu, Y., , Yu, T.: A Field Model for Human Detection and Tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **5** (2006) 753–765
- [172] Yilmaz, A., , Shah, M.: Actions sketch: a novel action representation. *Proceedings of IEEE Computer Vision and Pattern Recognition* **1** (2005) 984–989
- [173] Young, T., , Palta, M., , Dempsey, J. et al.: Estimation of the clinically diagnosed proportion of sleep apnea syndrome in middle aged men and women. *Sleep* **20** (1997) 705–706
- [174] Yeoh, E. J., , Ross, M. E., , Shurtleff, S. A., , Williams, W. K., , Patel, D., , Mahfouz, R., , Behm, F. G. et al.: Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. *Cancer Cell* **1(2)** (2002) 133–143
- [175] Zembutsu, H., , Ohnishi, Y., , Tsunoda, T., , Furukawa, Y., , Katagiri, T., , Ueyama, Y. et al.: Genome-wide cDNA microarray screening to correlate gene expression profiles with sensitivity of 85 human cancer xenografts to anticancer drugs. *Cancer Research* **62(2)** (2002) 518–27
- [176] Zhu, Z., , Fei, J., , Pavlidis, I.: Tracking Human Breath in Infrared Imaging. *Proceeding of the 5th IEEE Symposium on Bioinformatics and Bioengineering* (2005) 227–231

# Appendix A

## Terms and Definition

- Sleep Apnea: A sleep disorder that causes breathing to stop during sleep for anywhere from ten seconds up to several minutes.
- Oxygen Saturation: A measure of how much oxygen the blood is carrying as a percentage of the maximum it could carry. It can be obtained from pulse oximetry.
- CPAP (Continuous Positive Airway Pressure Therapy): A therapy delivers air into the patient's airway through a specially designed nasal mask or pillows. It is considered the most effective nonsurgical treatment for the alleviation of snoring and obstructive sleep apnea.
- Polysomnography (PSG): A diagnostic test, which a number of sensor leads are placed on the patient during sleep to record brain activity, eye, jaw muscle and leg muscle movement, airflow, respiratory effort, heart rhythm and oxygen saturation.
- Infrared: Infrared light lies between the visible and microwave portions of the electromagnetic spectrum. It is used in night-vision equipment when there is insufficient visible light available.
- Hypopneas: Reductions in airflow or respiratory effort during sleep.
- Thermal Imaging: An analogue pictorial representation or visualization of temperature differences.

# Appendix B

## 3–4 DT Algorithms

In the binary image, each feature pixel is first set to zero and each non-feature pixel is set to infinity.

### 1 Parallel DT

For iteration  $k$ , define the value  $v_{i,j}^k$  of the pixel in position  $(i, j)$ . (The iterations continue until no value changes and the number of iterations is proportional to the longest distance occurring in the image.)

$$\begin{aligned} v_{i,j}^k = \min(&v_{i-1,j-1}^{k-1}, v_{i-1,j}^{k-1} + 3, \\ &v_{i-1,j+1}^{k-1} + 4, v_{i,j-1}^{k-1} + 3, \\ &v_{i,j}^{k-1}, v_{i,j+1}^{k-1} + 3, \\ &v_{i+1,j-1}^{k-1} + 4, v_{i+1,j}^{k-1} + 3, \\ &v_{i+1,j+1}^{k-1} + 4) \end{aligned} \tag{B.1}$$

### 2 Sequential DT

Forward:

for  $i = 2 \dots rows$  do  
for  $j = 2 \dots columns$  do

$$\begin{aligned} v_{i,j} = \min(&v_{i-1,j-1} + 4, v_{i-1,j} + 3, \\ &v_{i-1,j+1} + 4, v_{i,j-1} + 3, v_{i,j}) \end{aligned} \tag{B.2}$$

Backward:

for  $i = rows - 1 \dots 1$  do  
for  $j = columns - 1 \dots 1$  do

$$\begin{aligned} v_{i,j} = \min(&v_{i,j}, v_{i,j+1} + 3, v_{i+1,j-1} + 4, \\ &v_{i+1,j} + 3, v_{i+1,j+1} + 4) \end{aligned} \tag{B.3}$$

# Appendix C

## A Low Variance Error Boosting Algorithm

We introduce a robust variant of AdaBoost, *cw-AdaBoost*, that uses weight perturbation to reduce variance error, and is particularly effective when dealing with data sets, such as microarray data, which have large numbers of features and small number of instances. The algorithm is compared with AdaBoost, Arcing and MultiBoost, using twelve gene expression datasets, using 10-fold cross validation. The new algorithm consistently achieves higher classification accuracy over all these datasets. In contrast to other AdaBoost variants, the algorithm is not susceptible to problems when a zero-error base classifier is encountered. The performance is analyzed by considering the bias/variance decomposition of the classification error rate.

### C.1 Introduction

A large number of studies have shown the effectiveness of ensemble learning algorithms in improving classifier performance. Breiman [25] introduced the Bagging algorithm, which forms an ensemble by aggregating multiple classifiers, each of which is trained using a bootstrapped training set (randomly sampled with replacement from the training set). This approach is very effective in reducing the variance of the ensemble classifier, and is particularly useful if using “unstable” base classifiers such as decision trees or neural networks [24], that can produce convoluted decision regions which vary heavily according to the selection of the training set. In contrast, Freund’s [52] Adaboost ensemble algorithm uses weighted training samples, and the weights are deterministically updated to emphasize misclassified instances from the

training set. This allows even a relatively simple base classifier algorithm to adjust for complex decision surfaces, allowing both the bias and the variance to be reduced.

However, Boosting does have some known limitations, including that the deterministic sampling does not necessarily optimize the rate of variance reduction, and issues that occur when zero-error or high error base classifiers are created. The algorithm introduced in this work addresses these limitations.

### C.1.1 Related Work

The Boosting algorithm is extremely powerful, and consequently has received a great deal of attention from the machine learning community, not least in addressing some of the known limitations. Several authors have attempted to integrate the stochastic element of bagging into a boosting framework. Friedman [54] proposed a stochastic gradient Boosting, which randomly draws sub-samples of the training data (without replacement) at each iteration to train individual base classifiers. The intention of this method is to use the bootstrap sampling approach of Bagging to improve the variance reduction of Boosting. However, it leads to a problem that using smaller sub-samples in training base models causes the variance of the individual base classifiers to increase. Webb [167] proposed a MultiBoost algorithm by combining a modified boosting algorithm with wagging. MultiBoost wraps Boosting inside Bagging, utilizing the continuous Poisson distribution to generate a number of randomly weighted (sampled) data from the original training dataset and then constructs bags of individual ensembles, each of which learns by Boosting from the weighted samples (which in effect provide a randomized weighting start-point for the Boosting algorithm). A detailed analysis is given in section C.2.3.

AdaBoost and its variants typically impose a stopping condition on the base classifier error rate. If this exceeds 0.5, they stop as the underlying theory only guarantees decreasing ensemble error performance for base-classifiers with better than random performance. This stopping criteria may be encountered due to the distortions introduced by the boosted weighting of some instances. However, they also stop if the error rate hits zero – this is surprisingly common in problems with low numbers of instances and large numbers of variables, where it is in fact still useful to form ensembles to counteract the high variance inherent in such a data set.

Early stopping of AdaBoost is a form of shrinkage, leading to low generalization and higher variance error. However, early stopping or low generation problem occur in original AdaBoost Algorithm and its successors. Variants of

AdaBoost that halt under these conditions include MadaBoost by Domingo and Watanabe [39], *LPBoost*, *TotalBoost<sub>v</sub>*, *TotalBoost<sub>v</sub><sup>g</sup>*, Brownboost [53], BagBoosting [36], Logitboost [55], *AdaBoost<sub>v</sub><sup>\*</sup>* and *AdaBoost<sub>v</sub><sup>g</sup>* by Warmuth et al [166]. Warmuth *et al* explicitly highlight early stopping as an important issue for future research.

In the original AdaBoostM1 paper [52], Freund and Schapire pointed out the main disadvantage of AdaBoostM1 that is unable to handle weak hypotheses with error greater than 0.5. It halts induction when error is greater than 0.5. To prevent early stopping, a variant of AdaBoostM1 is proposed by Bauer and Kohavi [23] to overcome this weakness. If the error is greater than 0.5, this variant of AdaBoostM1 throws away the base classifier and bootstraps a new sample set from the original input training set with identical weight 1 for every instance. It then re-builds a base classifier using the new sample. Although this allows ensemble building to continue, and may aid with variance reduction, it also discards the boosted weights which are largely responsible for the bias reduction of AdaBoost. The model has another weakness – if one of the base classifiers achieves zero error rate, boosting stops, and furthermore the error free base classifier gets infinite voting power and becomes the only voter, turning the ensemble to a single classifier model.

Webb [167] addressed this latter issue by assigning the voting power of the error free classifier a specific value,  $\log(10^{10})$ , and restarting the boosting process using a new bootstrap sample from the original training set, echoing Bauer and Kohavi’s approach to high errors. Webb then combines the modified algorithm with wagging and introduces another boosting algorithm, MultiBoost. However, these modified algorithms still suffer from low generalization; detailed analysis is given in section C.2.3.

The importance of diversity in the pool of base classifiers has been discussed in a number of papers [6, 7, 35, 92, 102], showing that ensembles that enforce diversity fare better than ones that do not. The motivation of this work is to investigate a technique to overcome the low generalization problem of boosting algorithms. We apply the proposed technique to three boosting algorithms: the original AdaBoostM1, MultiBoost (Boosting without stopping conditions) and Arcing (Another type of Boosting with stopping conditions), and recommend the variant with highest performance.

## C.1.2 Motivation

This research was motivated by the investigation of ensemble learning in the classification of gene expression data, which typically is high dimensional with a relatively low number of instances. We have observed that popular existing

ensemble methods, including Bagging [25], Boosting (AdaBoostM1) [52] and Arcing (ArcX4) [24] and MultiBoost [167], encounter specific problems in processing such data sets which are not necessarily encountered in data sets with lower dimensionality and more samples. In our experiments in the classification of twelve gene expression data, we found that one or two error free models often dominate the ensemble. A detailed analysis is presented in section C.2.

The main contribution of this research is to introduce a weight perturbation technique for boosting algorithms that increases the diversity of the base models, so reducing variance, without damaging the bias performance, and without allowing early stopping. The algorithm continues to perform, and to improve performance, when error free classifiers are encountered. The algorithm is also able to work with “unstable” (i.e. complex) classifiers such as decision trees, which inherently have relatively low bias, and are less likely to generate high error rate models than simpler base classifiers dealing with boosted samples.

The algorithm maintains instance weights, like boosting, but in addition to updating these to emphasize misclassified instances (thus reducing bias), it uses an efficient resampling technique to perturb the weights – in effect, bootstrapping from the weighted training set. This perturbation reduces variance, and allows the algorithm to continue successfully even if a zero error base classifier is encountered.

We have experimented with modified versions of Boosting, Arcing and MultiBoost, generating three modified algorithms (cw-AdaBoost, cw-Arcing and cw-MultiBoost). In evaluation, these algorithms were compared with Bagging, Boosting, Arcing and MultiBoost, in the classification of 12 gene expression datasets [8, 9, 10, 34, 37, 56, 57, 124, 156, 174, 175] utilizing the 10-fold cross validation technique. The experimental results show that the modified algorithms achieve significantly better performance than the original approaches. The cw-AdaBoost algorithm consistently achieves higher accuracy over 12 gene expression datasets than the existing algorithms.

The outline of this appendix is as follows. In section C.2, four benchmark algorithms (Bagging, Boosting, Arcing and MultiBoost) are described. Section C.3 describes the new algorithms and the proposed modification technique, and section C.4 presents the experimental results. We conclude in section C.5. A detailed presentation of experimental results is given in section C.6.

## C.2 Analyses of Benchmark Ensemble Learning Algorithms

We have benchmarked the algorithm against the original AdaBoostM1, MultiBoost (a variant of Boosting that also tries to integrate the advantages of bagging), and Arcing (Another type of boosting without stopping conditions). We have experimented with variants of each of these algorithms using the new resampling approach, and have also benchmarked against Bagging. The study shows that the proposed modification improves all of these boosting variants, but that the simple cw-AdaBoost (AdaBoostM1 integrated with the new sampling algorithm) is most effective.

### C.2.1 Bagging

Bagging [25] forms an ensemble by bootstrapping from the data set to build individual base classifiers. These are combined using an un-weighted voting mechanism, and the classification output is the most often predicted class label. It is characteristic of Bagging that base models are constructed independently. In other words, knowledge is not accumulated between iterations: previously learned experience does not affect the learning process afterwards.

#### 1 Bagging Algorithm:

Given a training set  $S : (x_1, y_1), \dots, (x_M, y_M)$  with labels  $y_j \in Y = \{1, \dots, N\}$ , a base learner  $I$  and the number of base models to build  $T$ , produce the Bagging classifier  $C^*(x)$  by the following steps.

1. for  $i = 1$  to  $T$ 
  - 1.1.  $\hat{S} =$  bootstrap sample from  $S$  (i.i.d. sample with replacement)
  - 1.2. build a base model  $C_i = I(\hat{S}_i)$
2.  $C^*(x) = \arg \max_{y \in Y} (\sum_{t: C_t(x)=y} 1)$

#### 2 Weakness Analysis:

Bagging cannot reduce bias below that of the base classifiers.

### C.2.2 AdaBoost (AdaBoostM1)

The breakthrough feature of boosting is the sequential development of base classifiers. The algorithm assigns weights to instances; in particular, the weights of misclassified instances are increased with each iteration, so that increased attention is paid to correcting mistakes made on previous iterations. The major difference between Bagging and Boosting is that individual

base models in Bagging are built independent to each other whereas base models in Boosting are adaptively built. Freund and Schapire [52] proposed several extensions of Boosting called adaptive Boosting, including AdaBoost, AdaBoostM1, AdaBoostM2 and AdaBoost.R. In this research, we adopt AdaBoostM1 as the benchmark Boosting method.

### 1 AdaBoostM1 Algorithm:

Given a training set  $S : (x_1, y_1), \dots, (x_M, y_M)$  with labels  $y_j \in Y = \{1, \dots, N\}$ , a base learner  $I$  and the number of base models to build  $T$ , produce the Boosting classifier  $C^*(x)$  by the following steps.

1. Create a new set  $S_1$  with instance weight  $w_k = 1$  where  $k = 1 \dots M$
2. for  $i = 1$  to  $T$ 
  - 2.1. build a base model  $C_i = I(S_i)$
  - 2.2.  $E_i = \frac{1}{M} (\sum_{x_k \in S_i: C_i(x_k) \neq y_k} w_k)$
  - 2.3. if  $(E_i > 0.5) \vee (E_i = 0)$ , deduct 1 from  $i$  and abort loop.
  - 2.4.  $B_i = \frac{E_i}{1-E_i}$
  - 2.5. for each  $x_k \in S_i$ , if  $C_i(x_k) \neq y_k$ , then multiply  $B_i$  to  $w_k$
  - 2.6. Normalize weights
3.  $C^*(x) = \arg \max_{y \in Y} (\sum_{t: C_t(x)=y} \log \frac{1}{B_t})$

### 2 Weakness Analysis:

AdaBoostM1 terminates when a base classifier with error greater than 0.5, or equal to 0, is obtained. That is, the Boosting algorithm stops learning when its performance on the training data is worse than by guessing, or it achieves perfect performance. In the most extreme case, if the error is zero on the first iteration then the algorithm constructs a single base classifier; this happens surprisingly frequently in gene expression data analysis, where the high input dimensionality and low number of instances often make it possible to achieve perfect performance on the *training* set. In such situations, the Boosting algorithm is unable to construct an effective ensemble, and its performance is drastically reduced; it has problems of low generalization and high variance. In addition, AdaBoostM1 does not have any stochastic element, and so although it achieves some variance reduction by virtue of the diverse ensembles generated, this effect is sometimes more limited than it might be.

## C.2.3 Modified AdaBoostM1 and MultiBoost

MultiBoost [167] wraps Boosting inside Bagging and generates each bagged ensemble by Boosting, in order to combine the advantages of Boosting in bias

reduction and Bagging in variance reduction. The adopted boosting algorithm is an AdaBoostM1 variant [23], which removes the stopping condition when the error rate is greater than 0.5. Step 2.3 of the original AdaBoostM1 algorithm is modified to:

### Modified AdaBoostM1 by Bauer and Kohavi [23]

- 2.3.1 If  $E_i > 0.5$ , set  $S_i$  to a bootstrap sample from original  $S$  with weight 1 for every instance and go back to step 2.1 to restart building a classifier (this step is limited to 25 times after which it exits the loop)
- 2.3.2 If  $E_i = 0$ , deduct 1 from  $i$  and abort loop

However, there is still early stopping issue in the modified AdaBoostM1 algorithm when an error free base classifier is obtained. Therefore, Webb further modifies the boosting algorithm to remove the stop conditions when  $E_i = 0$ . When  $E_i = 0$ , he assigns the voting power of the base classifier to  $\log(10^{10})$ , resets instance weights to random weights using the continuous Poisson distribution, and re-starts the training procedure.

### 1 MultiBoost Algorithm:

Given a training set  $S : (x_1, y_1), \dots, (x_M, y_M)$  with labels  $y_j \in Y = \{1, \dots, N\}$ , a base learner  $I$ , the number of base models to build  $T$ , and a vector of integers  $V_j$  specifying the iteration at which each subcommittee  $j > 1$  should terminate, produce the MultiBoost classifier  $C^*(x)$  by the following steps.

1. Create a new set  $S_1$  with instance weight  $w_k = 1$  where  $k = 1 \dots M$
2. set  $j = 1$
3. for  $i = 1$  to  $T$ 
  - 3.1. if  $V_j = i$ ,
    - 3.1.1. reset  $S_i$  to random weights drawn from continuous Poisson distribution.
    - 3.1.2. normalize weights
    - 3.1.3. increment  $j$  by 1
  - 3.2. build a base model  $C_i = I(S_i)$
  - 3.3.  $E_i = \frac{1}{M}(\sum_{x_k \in S_i: C_i(x_k) \neq y_k} w_k)$
  - 3.4. if  $E_i > 0.5$ ,
    - 3.4.1. set  $S_i$  to random weights drawn from the continuous Poisson distribution
    - 3.4.2. normalize weights
    - 3.4.3. increment  $j$  by 1
    - 3.4.4. go to step 3.2

- 3.5. if  $E_i = 0$ ,
  - 3.5.1. set  $B_i = 10^{-10}$
  - 3.5.2. set  $S_i$  to random weights drawn from the continuous Poisson distribution
  - 3.5.3. normalize weights
  - 3.5.4. increment  $j$  by 1
- 3.6. Otherwise,
  - 3.6.1.  $B_i = \frac{E_i}{1-E_i}$
  - 3.6.2. for each  $x_k \in S_i$ ,
    - 3.6.2.1. if  $C_i(x_k) \neq y_k$ , divide  $w_k$  by  $2E_i$
    - 3.6.2.2. otherwise, divide  $w_k$  by  $2(1 - E_i)$
    - 3.6.2.3. if  $w_k < 10^{-8}$ , set  $w_k$  to  $10^{-8}$
4.  $C^*(x) = \arg \max_{y \in Y} (\sum_{t: C_t(x)=y} \log \frac{1}{B_t})$

## 2 Weakness Analysis:

There are two problems with this design: first, the algorithm discards previously learned knowledge (in the form of Boosting weights) and restarts the training procedure from scratch even when it has obtained a zero error on the input training data; second, the algorithm sets  $B_i$  to  $10^{-10}$  when an error free base model is obtained. The latter seriously affects the performance of the ensemble model, damaging its generalization performance, as such base classifiers dominate the ensemble. This drawback is apparent in our experimental results, showing that the algorithm performs very poorly in some of the datasets, such as “colon tumor” and “prostate outcome.”

Furthermore, resetting the weights on each iteration of Bagging discards the knowledge on weight setting gained during Boosting. We can expect each run of Boosting to converge back towards approximately the same weights, but the procedure is time-consuming. The experimental results are consistent with our theory and show that MultiBoost improves more slowly than our new algorithms. Fig C.1 shows the results on one gene expression dataset, i.e. Breast Cancer, and illustrates the faster convergence of cw-AdaBoost and cw-Arcing. In addition, we evaluate the performance of MultiBoost, which sets  $B_i$  to a bigger value  $10^{-8}$  when  $E_i = 0$  to assign smaller decision power to error free classifiers. The aim is to investigate if the performance of MultiBoost can be improved. However, there is no clear improvement by changing  $B_i$  value when  $E_i = 0$ . The results are displayed in Table C.1.

Table C.1: MultiBoost Performance with different $B_i$				
iteration	10	20	30	40
<i>Prostate<sub>Outcome</sub></i>				
MultiBoost( $B_i = 10^{-10}$ )	57.14	76.19	71.43	76.19
MultiBoost( $B_i = 10^{-8}$ )	52.38	76.19	71.43	76.19
cw-AdaBoost	100.00	100.00	95.24	95.24
<i>Breast<sub>Cancer</sub></i>				
MultiBoost( $B_i = 10^{-10}$ )	83.51	86.60	87.63	91.75
MultiBoost( $B_i = 10^{-8}$ )	85.57	86.60	89.69	91.75
cw-AdaBoost	90.72	95.88	95.88	95.88
<i>Colon<sub>Tumor</sub></i>				
MultiBoost( $B_i = 10^{-10}$ )	80.65	79.03	79.03	79.03
MultiBoost( $B_i = 10^{-8}$ )	80.65	79.03	79.03	79.03
cw-AdaBoost	93.55	93.55	93.55	91.94

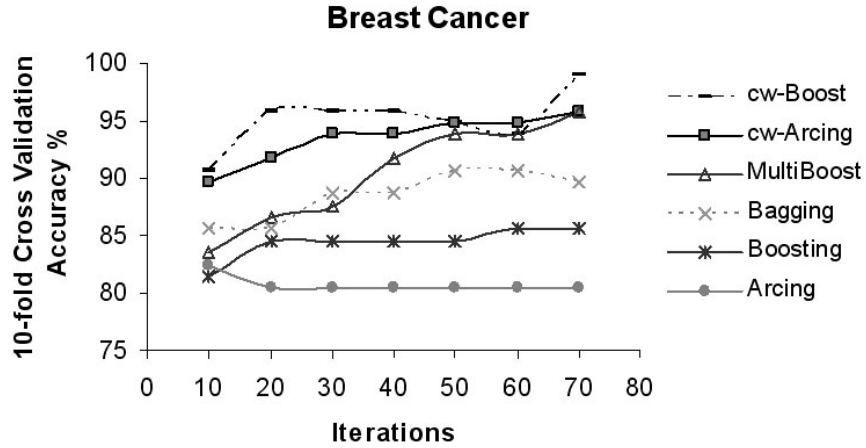


Figure C.1: Fast convergence of cw-AdaBoost and cw-Arcing

### C.2.4 Arcing

There are two types of Arcing, i.e. Arc-fs using weighted voting and arcX4 using un-weighted voting. In this work, we adopt arcX4 because the arcX4 algorithm is suggested to have a slight edge in test set error results [24] on smaller datasets, and the experimental datasets in this research tend to have fewer instances. The framework of Arcing is similar to the one employed in Boosting. They both proceed in sequentially self-adjusting steps. However, there are three major differences between Arcing and Boosting: (1) Arcing does not employ a stop condition; (2) Arcing adopts an un-weighted voting system; (3) Arcing adapts its behavior based on the accumulation  $\{E_k\}$  of its faults in history and examines all previous base classifiers' faults when constructing a new base classifier, whereas Boosting considers only the error of the previous iteration's base classifier.

#### 1 Arcing Algorithm:

Given a training set  $S : (x_1, y_1), \dots, (x_M, y_M)$  with labels  $y_j \in Y = \{1, \dots, N\}$ , a base learner  $I$  and the number of base models to build  $T$ , produce the Arcing classifier  $C^*(x)$  by the following steps.

1. Create a new set  $S_1$  with instance weight  $w_k = 1$  where  $k = 1 \dots M$
2. Create a vector  $\{E_k\}$  where  $E_k = 0$  and  $k = 1 \dots M$
3. for  $i = 1$  to  $T$ 
  - 3.1. build a base model  $C_i = I(S_i)$
  - 3.2. for each  $x_k \in S_i$ , if  $C_i(x_k) \neq y_k$ , add 1 to  $E_k$  and set  $w_k = 1 + E_k^4$
  - 3.3. Normalize weights
4.  $C^*(x) = \arg \max_{y \in Y} (\sum_{t: C_t(x)=y} 1)$

#### 2 Weakness Analysis:

A drawback of Arcing is its deteriorating performance and the decreasing diversity of base models as more base models are built. Once a base classifier exactly fits the training dataset, there will be no change in the accumulated misclassification values  $\{E_k\}$ , and hence all instances' weights remain the same in building the next base model, due to the re-weighting function ( $w_k = 1 + E_k^4$ ). Thus, Arcing will continuously produce identical base models once an error-free classifier is built. In the worst case, Arcing may generate a basket of identical base classifiers. In other cases, after an error free base classifier is trained, Arcing will continuously produce identical base models until the maximum number of base models are built. Consequently, the diversity of base models of Arcing gradually decreases after that point. Our experimental results show that the accuracy of the entire Arcing model deteriorates once this happens.

## C.3 Proposed Modification: cw-resampling

Boosting halts induction when the optimization problem becomes infeasible [39] [166]. Bauer and Kohavi [23] and Freund and Schapire [52] addressed the early stopping issue, but left open the question of iteration bounds for future research. Although Webb [167] and Breiman [24] introduced the modified boosting algorithms, MultiBoost and Arcing, to address stopping conditions issues, these modified boosting algorithms still suffer from generalization issues, as discussed above.

The proposed modification focuses on the optimization of boosting to reduce variance, and on preventing the algorithms from failing when error free classifiers occur. We therefore specify three key requirements of the proposed modification.

### 1 Key Requirements of the Proposed Modification

First, instead of stopping, the algorithm should be able to continue optimization and build more classifiers when an error free model is obtained. Second, the algorithm should be able to utilize the knowledge accumulated during sequential learning. In other words, when  $E_i = 0$ , the ensemble does not reset weights or restart from a bootstrapped set, which throws away the knowledge learned. Third, the decision of an ideal ensemble model should depend on a number of non-identical mature decision makers with low error rate rather than a few decision makers.

### 2 Design

The proposed model uses a weight perturbation approach to effectively resample around the weightings produced by boosting. This allows the algorithm to continue adding base classifiers even if a zero base classifier is discovered, and indeed to perform bias removal (by intermittent boosting steps) in this circumstance. Furthermore, for robustness, the decision power of base classifiers is based on their error rate rather than a fixed value. This allows boosting models to benefit from variance reduction and alleviates the overfitting problem. An illustration of the proposed design in comparison with the existing boosting algorithms is presented in Fig C.2.

### 2 Implementation

1. Early Stop (AdaBoostM1, MadaBoost, LPBoost, TotalBoost...)
 

$C1...C13 \longrightarrow \text{Stop, error rate } =0 \longrightarrow 13 \text{ base classifiers, but } C13 \text{ gains infinite decision power.}$
2. Reset weights (MultiBoost, Modified AdaBoostM1)
 

$C1...C13 \xrightarrow{\text{reset}} C1...C15 \xrightarrow{\text{reset}} C1.C2 \longrightarrow 30 \text{ classifiers}$   
 $\uparrow$  restart learning
 $\uparrow$ 
Decision dominated by C13, C15
3. No Stop and no resetting weights (Arcing)
 

$C1...C13, C13 \dots C13 \longrightarrow 30 \text{ classifiers but 18 are identical}$   
Decision dominated by C13
4. Proposed Structure
 

$C1...C13_1 \ C13_2 \dots C13_{18} \longrightarrow 30 \text{ diverse classifiers}$   
with 18 mature and different decision makers

Figure C.2: Illustration of the proposed design and low generalization issues of existing boosting methods: If  $C13$  is an error free classifier, boosting methods in the first group will produce only 13 classifiers no matter how large the number of base models originally specified, and as  $C13$  gains infinite decision power, the decision of these ensembles is dominated by one base classifier; boosting methods in the second group assign considerably high decision power to the two error free models,  $C13, C15$ , and thus the decision is dominated by these two base classifiers; boosting methods in the third group continuously produce identical classifiers  $C13$ , and the decision of such ensemble models is dominated by this one classifier; the proposed structure generate diverse classifiers and an effective ensemble with 30 different and 18 mature decision makers.

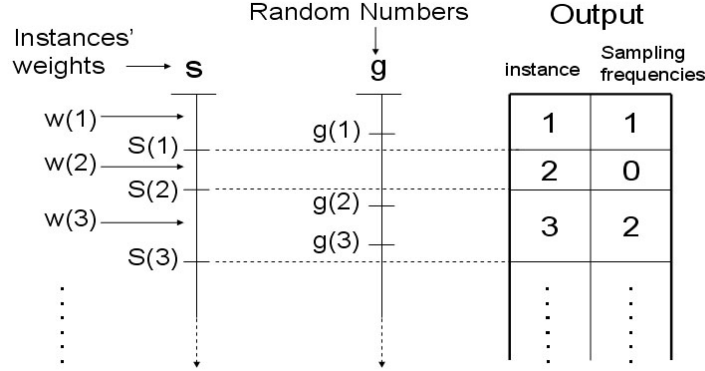


Figure C.3: Re-sampling Scheme

First, before training each base classifier we alter the instance weights using the random resampling approach for weighted instances described below; second, provided the base classifier performance is not zero we update the weights using a boosting approach. Thus, even if the base classifier has error zero, the algorithm continues to produce diverse classifiers. The weight perturbation algorithm injects some randomness into the learning behavior, without wholly discarding the knowledge built up in previous iterations of boosting. It effectively bootstraps a new training set by sampling from the weighted training set generated on the previous iteration (i.e. it uses the instance weights to influence the selection frequency in bootstrapping). It thus keeps the sequential adaptive learning strengths of boosting, while injecting randomness to generate diverse classifiers and improve performance where boosting would fail.

Importantly, the standard “fairly sampling” technique as used by AdaBoost [23] and Bagging is not suitable here. The standard sampling technique resets every instance’s weight to 1 and then samples with replacement. This throws away all knowledge learned, and re-starts learning from the beginning, which loses the virtue of boosting algorithms. An extreme example is MultiBoost, which resets weights both periodically and whenever  $\epsilon_i > 0.5 \vee \epsilon_i = 0.5$ . Under situations without error free base models obtained, MultiBoost has slower convergence and poorer performance than the new variant, with less than 40 base classifiers, as shown in Fig C.1.

The weight perturbing algorithm is computationally efficient, with time complexity  $O(n)$ . It uses an array of cumulative weight bins,  $s$ , and an array of cumulative random numbers,  $g$ , normalized to the same final sum.

The new weights are assigned according to the number of random values associated with the corresponding bin; see fig C.3. This is equivalent to producing the new weights by weighted resampling with replacement but has optimal time complexity. The new algorithms (cw-AdaBoost, cw-Arcing and cw-MultiBoost) and the weight perturbing algorithm (cw-Resample) are presented below.

### C.3.1 cw-AdaBoost Algorithm

Given a training set  $S : (x_1, y_1), \dots, (x_M, y_M)$  with labels  $y_j \in Y = \{1, \dots, N\}$ , a base learner  $I$ , the number of base models to build  $T$ , and an integer  $R$  (maximum number of times to perturb data; in experiments we use  $R=10$ ), produce the cw-AdaBoost classifier  $C^*(x)$  by the following steps.

1. Create a new set  $S_1$  with instance weight  $w_k = 1$  where  $k = 1 \dots M$
2. for  $i = 1$  to  $T$ 
  - 2.1. set  $r = 0$
  - 2.2. perturb  $S_i$  using cw-Resample
  - 2.3. build a base model  $C_i = I(S_i)$
  - 2.4. increment  $r$  by 1
  - 2.5.  $E_i = \frac{1}{M} (\sum_{x_k \in S_i: C_i(x_k) \neq y_k} w_k)$
  - 2.6. if  $(E_i = 0) \wedge (r \leq R)$ , go to step 2.2.
  - 2.7. if  $(E_i > 0.5) \vee (E_i = 0)$ , deduct 1 from  $i$  and abort loop.
  - 2.8.  $B_i = \frac{E_i}{1-E_i}$
  - 2.9. for each  $x_k \in S_i$ , if  $C_i(x_k) \neq y_k$ , then multiply  $B_i$  to  $w_k$
  - 2.10. Normalize weights
3.  $C^*(x) = \arg \max_{y \in Y} (\sum_{t: C_t(x)=y} \log \frac{1}{B_t})$

### C.3.2 cw-Arcing Algorithm

Given a training set  $S : (x_1, y_1), \dots, (x_M, y_M)$  with labels  $y_j \in Y = \{1, \dots, N\}$ , a base learner  $I$ , the number of base models to build  $T$ , and an integer  $R$  (maximum number of times to perturb data; in experiments we use  $R=10$ ), produce the cw-Arcing classifier  $C^*(x)$  by the following steps.

1. Create a new set  $S_1$  with instance weight  $w_k = 1$  where  $k = 1 \dots M$
2. Create a vector  $\{E_k\}$  where  $E_k = 0$  and  $k = 1 \dots M$
3. for  $i = 1$  to  $T$ 
  - 3.1. set  $r = 0$
  - 3.2. perturb  $S_i$  using cw-Resample
  - 3.3. build a base model  $C_i = I(S_i)$
  - 3.4. increment  $r$  by 1

- 3.5.  $E_i = \frac{1}{M}(\sum_{x_k \in S_i: C_i(x_k) \neq y_k} w_k)$
- 3.6. if  $(E_i = 0) \wedge (r \leq R)$ , go to step 3.2.
- 3.7. for each  $x_k \in S_i$ , if  $C_i(x_k) \neq y_k$ ,
  - 3.7.1 add 1 to  $E_k$
  - 3.7.2  $w_k = 1 + E_k^4$
- 3.8. Normalize weights
4.  $C^*(x) = \arg \max_{y \in Y} (\sum_{t: C_t(x)=y} 1)$

### C.3.3 cw-MultiBoost

Given a training set  $S : (x_1, y_1), \dots, (x_M, y_M)$  with labels  $y_j \in Y = \{1, \dots, N\}$ , a base learner  $I$ , the number of base models to build  $T$ , a vector of integers  $V_j$  specifying the iteration at which each subcommittee  $j > 1$  should terminate, and an integer  $R$  (maximum number of times to perturb data; in experiments we use  $R=10$ ), produce the MultiBoost classifier  $C^*(x)$  by the following steps.

1. Create a new set  $S_1$  with instance weight  $w_k = 1$  where  $k = 1 \dots M$
2. set  $j = 1$
3. for  $i = 1$  to  $T$ 
  - 3.1. set  $r = 0$
  - 3.2. perturb  $S_i$  using cw-Resample
  - 3.3. build a base model  $C_i = I(S_i)$
  - 3.4. increment  $r$  by 1
  - 3.5.  $E_i = \frac{1}{M}(\sum_{x_k \in S_i: C_i(x_k) \neq y_k} w_k)$
  - 3.6. if  $(E_i = 0) \wedge (r \leq R)$ , go to step 3.2
  - 3.7. if  $(E_i > 0.5) \vee (E_i = 0)$ , deduct 1 from  $i$  and abort loop
  - 3.8.  $B_i = \frac{E_i}{1-E_i}$
  - 3.9. if  $V_j = i$ ,
    - 3.9.1. reset  $S_i$  to random weights drawn from continuous Poisson distribution.
    - 3.9.2. normalize weights
    - 3.9.3. increment  $j$  by 1
  - 3.10. otherwise,
    - 3.10.1. for each  $x_k \in S_i$ , if  $C_i(x_k) \neq y_k$ , multiply  $B_i$  to  $w_k$
    - 3.10.2. normalize weights
4.  $C^*(x) = \arg \max_{y \in Y} (\sum_{t: C_t(x)=y} \log \frac{1}{B_t})$

### C.3.4 cw-Resample Algorithm

Given a dataset  $S$ : a sequence of instances with weights:  $(i_1, w_1), \dots, (i_M, w_M)$ , we produce new dataset with the following steps.

1. Generate  $M$  random number:  $r_1, \dots, r_M$
2.  $R = \sum_{j=1 \dots M} r_j$
3.  $W = \sum_{j=1 \dots M} w_j$
4. set  $a = 1$  and  $b = 1$
5. let  $g(a) = (\sum_{j=1 \dots a} \frac{r_j}{R}) \times M$
6. let  $s(b) = (\sum_{j=1 \dots b} w_j)$
7. if  $(a > M) \wedge (b > M)$ , then terminate.
8. if  $g(a) < s(b)$ ,
  - 8.1. select instance  $i_b$  into the output dataset
  - 8.2. increment  $a$  by 1
  - 8.3. go to step 5
9. otherwise,
  - 9.1. increment  $b$  by 1
  - 9.2. go to step 5

## C.4 Experiments

The experiments are conducted using 12 published gene expression datasets [8, 9, 10, 34, 37, 56, 57, 124, 156, 174, 175], which are obtained from [151]. Details of the data cleaning process are given in the original paper. In evaluation, Ambroise and McLachlan [4] recommend using 10-fold rather than leave-one-out cross-validation for gene expression data analysis. In this research, 10-fold cross validation is utilized and C4.5 decision tree algorithm [127] is used as the base classifier. Furthermore, to investigate the influence of the number of base models used, we evaluate the classification accuracy of the ensembles with different numbers of base classifiers (from 10 to 70 classifiers in steps of 10). The experimental results show that the modified algorithms all perform better than the corresponding original algorithms. The cw-AdaBoost algorithm consistently performs best over all 12 gene expression datasets, and is our recommended variant. In order to compare the performances of the seven algorithms on 12 datasets with different number of iterations, we first generate the cross validated average accuracy  $E_i$  of the algorithms on a specific iteration number  $i$ , to represent the average performance of 7 algorithms with iteration  $i$ . Given  $A_i(m)$  is 10-fold cross validation accuracy of the algorithm  $m$  with iteration  $i$ ,  $E_i = A_i(m)/7$ . We then create a performance index  $P_m$  to compare the relative performance of

Table C.2: 10-fold Cross Validation Accuracy% for single dataset(Breast Cancer)

base classifiers	10	20	30	40	50	60	70	$P_m$
Bagging	85.57	85.57	88.66	88.66	90.72	90.72	89.69	-0.71
Arcing	82.47	80.41	80.41	80.41	80.41	80.41	80.41	-8.52
Boosting	81.44	84.54	84.54	84.54	84.54	85.57	85.57	-4.83
MultiBoost	83.51	86.60	87.63	91.75	93.81	93.81	95.88	1.20
cw-Arcing	89.69	91.75	93.81	93.81	94.84	94.84	95.88	4.29
cw-AdaBoost	90.72	95.88	95.88	95.88	94.84	93.81	98.97	5.91
cw-MultiBoost	90.72	87.63	89.69	91.75	91.75	95.88	95.88	2.67
Average $E_i$	86.30	87.48	88.66	89.54	90.13	90.72	91.75	

the algorithm  $m$ . ( $P_m = (\sum_{i=10,20,\dots,70}(A_i(m) - E_i))/7$ .) Table C.2 illustrates the performance index on a single dataset, and Table C.3 displays the relative performance indices  $P_m$  on 12 gene expression datasets, showing that the new methods (particularly cw-AdaBoost, which has the best results for nine datasets, and the second best for two others) obtain consistently high performance index values.

Using the Wilcoxon signed rank test to compare the performance of cw-AdaBoost with cs-Arcing (the second best algorithm), we obtain the Wilcoxon statistic  $W=52$  with  $N=12$  samples, yielding the  $z$ -value  $2.02 > 1.96$ , and therefore conclude that the performance is significantly better at the 97.5% one-sided confidence level. Stronger results are obtained in comparing cw-AdaBoost with the algorithms, so that we conclude that cw-AdaBoost has superior performance. (The 10-fold cross validation results on the other 11 datasets are presented in the section C.6.)

A distinctive feature of gene expression data is that error free base model can be generated, causing the low generalization issue discussed above. The results are consistent with our theories, which are: (1) arcine keeps producing identical base models after an error free base model is built, and therefore the diversity of base models of arcine deteriorates. Thus, the variance error increases afterwards; (2) the stopping condition of boosting terminates further constructions of base classifiers and prevents further reduction of the variance error; (3) the performance of the multiboost is adversely affected by assigning  $\log 10^{10}$  to the decision power of an error free base model, leading to a small number of error free base classifiers dominating the decision output.

Table C.3: Performance Index  $P_m$  on 12 Gene Expression Datasets

	$AML_3$	Brain	Breast	CNS	Colon	$DLBCL_t$
Bagging	-1.39	-1.22	-0.71	-0.34	-5.73	-1.83
Arcing	-7.14	-2.37	-8.52	-9.63	-1.12	-7.77
Boosting	-4.36	-4.08	-4.83	1.56	6.02	1.14
MultiBoost	3.37	-0.37	1.20	-0.82	-7.11	1.51
cw-Arcing	3.37	2.49	4.29	3.23	8.10	1.69
cw-AdaBoost	3.17	4.49	5.91	5.13	6.26	2.99
cw-MultiBoost	2.98	1.06	2.67	4.42	-6.42	2.25
	Lung	$DLBCL_o$	$Prostate_o$	$MLL_2$	$Prostate_t$	Subtype
Bagging	-0.72	-4.27	-15.86	-2.88	-0.94	0.61
Arcing	-0.72	-9.20	-3.61	-1.87	-1.18	-6.56
Boosting	-0.17	-2.20	-6.33	-0.87	-0.05	1.57
MultiBoost	1.02	0.90	-5.65	-1.47	-1.99	0.78
cw-Arcing	0.94	5.09	12.04	0.32	2.45	0.17
cw-AdaBoost	1.49	5.33	18.84	4.28	1.72	1.75
cw-MultiBoost	1.17	4.34	12.72	2.49	-0.62	1.66

### C.4.1 Variance and Bias

In this section we present an analysis of the bias and variance of the algorithms, using the breast cancer dataset, and Kohavi and Wolpert’s [91] approach to variance and bias decomposition. There are 97 instances with 835 attributes in the original breast cancer data set  $D$ . Utilizing Kohavi and Wolpert’s approach [91], a sample of size 40 without replacement from the original data  $D$  is taken to produce a training set source. From the remainder, a test set of size 40 is sampled without replacement. There are 50 training samples to produce 50 trained ensemble models, which are then applied to the test set, and the bias and variance are calculated from the predictions on the test set.

Table C.4 tabulates the experimental results on the bias, variance and error of the ensemble models. Fig C.4 illustrates the differences in the performance of the original algorithms and the modified method. In general, the results show that the modified algorithms perform well on both bias and variance reduction. The top row, which compares bagging, boosting and the new algorithm show that, for this data set, boosting and bagging have equal performance on variance, but as expected boosting has lower bias. The cw-AdaBoost algorithm matches this low bias, and is able to continue boosting, further reducing the bias. The cw-AdaBoost algorithm also has noticeably lower variance than either boosting or bagging, indicating that the weight perturbation approach is highly effective. The second row show that the standard arcing algorithm deteriorates after around 30 iterations, at which point it keeps producing identical base models after an error free base model is built, leading to growing variance error. Boosting similarly stops improving after about 50 iterations. MultiBoost restarts learning every 10 iterations, and so it benefits from variance reduction in comparison to the original Boosting algorithm, but converges slowly; the proposed modification cw-MultiBoost variant achieves faster variance reduction.

## C.5 Conclusion

This work has introduced modifications to three boosting methods to generate efficient boosting models for training high dimensional datasets with low numbers of instances. Training this type of dataset (particularly with unstable base classifiers like decision tree) tends to generate error free base models and causes malfunctions on conventional Boosting, Arcing, and MultiBoost, leading to low generalization. The modified algorithms, which use a weight perturbation method combined with sequential update, and discards the

Table C.4:  $Bias^2$ ,  $Variance$  and  $Error$ 

$Bias^2$	10	20	30	40	50	60	80
Bagging	.1381	.129	.1307	.1281	.132	.1308	.1272
Boosting	.113	.1102	.1079	.1118	.1063	.1076	.1079
cw-AdaBoost	.1135	.1074	.1045	.1027	.1011	.0981	.0974
Arcing	.1408	.1244	.1155	.1089	.1108	.1121	.1142
cw-Arcing	.1269	.1094	.1063	.1057	.1064	.1038	.1053
MultiBoost	.1292	.1137	.1098	.1058	.1067	.1038	.1017
cw-MultiBoost	.1236	.1117	.1138	.1049	.1082	.1024	.1074
$Variance$	10	20	30	40	50	60	80
Bagging	.1662	.1549	.1542	.1523	.1492	.1483	.1467
Boosting	.1596	.1495	.1464	.1453	.1448	.1453	.1453
cw-AdaBoost	.1285	.11	.1015	.1012	.0997	.0978	.0937
Arcing	.1636	.1529	.1519	.1608	.1717	.178	.1858
cw-Arcing	.1191	.1163	.1155	.112	.112	.1122	.1117
MultiBoost	.1695	.1413	.1359	.1291	.1283	.1259	.1239
cw-MultiBoost	.1279	.1213	.1078	.1042	.1013	.1011	.0962
$Error$	10	20	30	40	50	60	80
Bagging	.3077	.287	.2881	.2835	.2842	.2821	.2768
Boosting	.2758	.2628	.2572	.26	.254	.2558	.2561
cw-AdaBoost	.2446	.2196	.2081	.206	.2028	.1979	.193
Arcing	.3077	.2804	.2705	.273	.286	.2937	.3039
cw-Arcing	.2484	.2281	.2242	.22	.2207	.2182	.2193
MultiBoost	.3021	.2579	.2484	.2375	.2375	.2323	.2281
cw-MultiBoost	.254	.2354	.2239	.2112	.2116	.2056	.2056

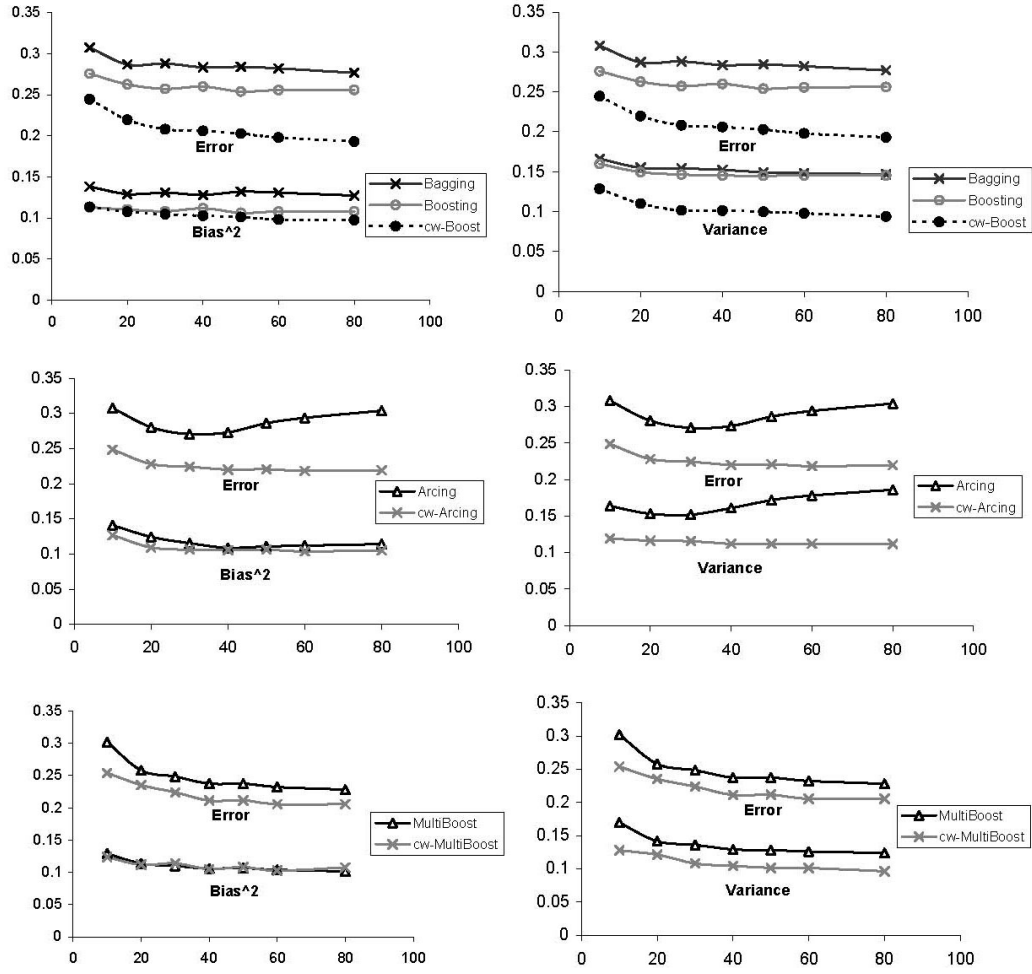


Figure C.4: Comparison on  $Bias^2$ ,  $Variance$  and Error between the original algorithms and the algorithms with the proposed modifications.

stopping condition, allows ensemble generation to continue, further lowering variance. We have introduced the cw-AdaBoost algorithm, which demonstrates superior performance on 12 gene expression data where it performs consistently well. We thus recommend it for wider use.

## C.6 Full Experimental Results

iteration	10	20	30	40	50	60	70
<i>Prostate<sub>Outcome</sub></i>							
Bagging	66.67	61.90	61.90	61.90	61.90	61.90	61.90
Arcing	66.67	76.19	76.19	80.95	80.95	71.43	71.43
Boosting	76.19	71.43	71.43	71.43	71.43	71.43	71.43
MultiBoost	57.14	76.19	71.43	76.19	76.19	76.19	76.19
cw-Arcing	90.48	95.24	90.48	90.48	90.48	85.71	90.48
cw-AdaBoost	100.00	100.00	95.24	95.24	100.00	95.24	95.24
cw-MultiBoost	80.95	95.24	95.24	85.71	95.24	95.24	90.48
AMLALL							
Bagging	91.67	94.44	94.44	95.83	95.83	94.44	95.83
Arcing	88.89	88.89	88.89	88.89	88.89	88.89	88.89
Boosting	91.67	91.67	91.67	91.67	91.67	91.67	91.67
MultiBoost	95.83	100.00	100.00	100.00	100.00	100.00	100.00
cw-Arcing	98.61	100.00	98.61	100.00	100.00	100.00	98.61
cw-AdaBoost	100.00	98.61	98.61	98.61	98.61	100.00	100.00
cw-MultiBoost	97.22	100.00	98.61	100.00	98.61	100.00	98.61
Brain Tumor							
Bagging	86.00	88.00	86.00	84.00	86.00	82.00	82.00
Arcing	82.00	84.00	84.00	84.00	84.00	84.00	84.00
Boosting	80.00	84.00	82.00	82.00	82.00	82.00	82.00
MultiBoost	72.00	82.00	88.00	88.00	88.00	90.00	92.00
cw-Arcng	82.00	86.00	90.00	90.00	92.00	90.00	90.00
cw-AdaBoost	88.00	92.00	92.00	88.00	92.00	90.00	92.00
cw-MultiBoost	80.00	84.00	92.00	90.00	88.00	88.00	88.00
ColonTumor							
Bagging	82.26	82.26	79.03	79.03	80.65	80.65	80.65
Arcing	91.94	85.48	83.87	83.87	83.87	83.87	83.87
Boosting	90.32	90.32	91.94	93.54	93.54	93.54	93.54
MultiBoost	80.65	79.03	79.03	79.03	79.03	79.03	79.03
cw-Arcing	96.77	93.55	91.94	93.55	93.55	95.16	96.77
cw-AdaBoost	93.55	93.55	93.55	91.94	91.94	91.94	91.94
cw-MultiBoost	80.65	80.65	85.48	74.19	79.03	79.03	80.65

<i>DLBCL<sub>Tumor</sub></i>							
Bagging	93.51	92.21	92.21	92.21	92.21	92.21	92.21
Arcing	92.21	84.42	85.71	85.71	85.71	85.71	85.71
Boosting	96.10	96.10	96.10	94.81	94.81	94.81	94.81
MultiBoost	92.21	94.81	97.40	97.40	96.10	96.10	96.10
cw-Arcing	90.90	96.10	97.40	94.81	97.40	97.40	97.40
cw-AdaBoost	92.20	97.40	97.40	98.70	97.40	98.70	98.70
cw-MultiBoost	93.51	96.10	96.10	96.10	98.70	97.40	97.40
<i>DLBCL<sub>Outcome</sub></i>							
Bagging	79.31	87.93	89.66	94.83	91.38	93.10	94.83
Arcing	86.21	86.21	86.21	84.48	84.48	84.48	84.48
Boosting	89.66	89.66	93.81	93.10	93.10	93.10	93.10
MultiBoost	84.48	94.83	96.55	96.55	98.28	98.28	98.28
cw-Arcing	98.28	98.28	100.00	100.00	100.00	100.00	100.00
cw-AdaBoost	100.00	100.00	100.00	100.00	100.00	100.00	98.28
cw-MultiBoost	91.34	100.00	100.00	100.00	100.00	100.00	100.00
Lung Cancer							
Bagging	97.24	97.24	97.24	97.24	97.24	97.24	97.24
Arcing	97.24	97.24	97.24	97.24	97.24	97.24	97.24
Boosting	97.79	97.79	97.79	97.79	97.79	97.79	97.79
MultiBoost	97.79	98.90	98.90	98.90	99.45	99.45	99.45
cw-Arcing	98.90	98.90	99.45	99.45	98.90	98.34	98.34
cw-AdaBoost	99.45	99.45	99.45	99.45	99.45	99.45	99.45
cw-MultiBoost	97.79	98.90	99.45	99.45	99.45	99.45	99.45
<i>MLL<sub>Leukemia</sub></i>							
Bagging	90.23	90.23	90.23	91.67	91.67	91.67	91.67
Arcing	91.67	93.06	93.06	91.67	91.67	91.67	91.67
Boosting	93.06	93.06	93.06	93.06	93.06	93.06	93.06
MultiBoost	93.06	93.04	91.67	93.06	93.06	91.67	91.67
cw-Arcing	95.84	95.84	95.84	93.06	93.06	93.06	93.06
cw-AdaBoost	95.84	100.00	95.83	98.61	100.00	98.61	98.61
cw-MultiBoost	91.67	94.44	98.61	95.83	97.22	98.61	98.61
CNS							
Bagging	88.33	88.33	86.67	88.33	86.67	86.67	88.33
Arcing	78.33	78.33	78.33	78.33	78.33	78.33	78.33
Boosting	88.33	88.33	90.00	90.00	90.00	90.00	90.00
MultiBoost	85.00	86.67	85.00	86.67	86.67	90.00	90.00
cw-Arcing	86.67	91.67	93.33	91.67	91.67	91.67	91.67
cw-AdaBoost	91.67	91.67	95.00	93.33	93.33	93.33	93.33
cw-MultiBoost	93.33	88.33	91.67	93.33	93.33	93.33	93.33

<i>Prostate<sub>Tumor</sub></i>							
Bagging	94.92	95.48	95.48	95.48	94.92	94.92	94.92
Arcing	94.92	94.92	94.92	94.92	94.92	94.92	94.92
Boosting	96.05	96.05	96.05	96.05	96.05	96.05	96.05
MultiBoost	88.24	94.12	94.85	95.59	95.59	95.59	94.85
cw-Arcing	98.87	98.31	98.31	98.31	98.31	98.87	98.87
cw-AdaBoost	97.74	97.74	98.31	97.18	97.74	97.74	98.31
cw-MultiBoost	93.38	94.85	96.32	95.59	95.59	97.06	95.59
SubtypeALL							
Bagging	89.91	91.13	91.13	91.13	90.83	91.13	91.13
Arcing	88.07	89.91	85.63	81.65	80.73	80.12	80.12
Boosting	89.30	92.05	92.97	92.05	91.74	92.66	92.35
MultiBoost	86.85	88.99	91.44	92.35	92.66	92.66	92.66
cw-Arcing	89.30	90.21	89.91	90.83	91.13	91.13	90.83
cw-AdaBoost	90.21	92.05	92.66	92.05	92.05	92.35	92.97
cw-MultiBoost	88.07	92.35	92.05	92.05	92.35	93.58	93.27

# Appendix D

## Full Statistical Test Results

This appendix presents the detailed statistical test results summarized in section 4.6.3. The three paired methods were compared, including “RTPose vs Ramanan”, “RTPose vs MatchPose” and “MatchPose vs Ramanan”, using different types of data (all images, high illumination, low illumination, with cover, without cover). The head, torso and a lower body part (RUL) were selected for statistical tests.

**A1. All images – head detection:** The contingency tables to compare “RTPose vs Ramanan”, “RTPose vs MatchPose” and “MatchPose vs Ramanan” in detecting the head are:

47	15	28	34	47	9
186	307	28	465	185	313

The outcomes ( $\frac{(|n_{01}-n_{10}|-1)^2}{n_{01}+n_{10}}$  = 143.65, 0.33 and 158.28) show that overall RTPose and MatchPose perform significantly better than Ramanan, and RTPose and MatchPose have similar performance.

**A2. All images – torso detection:** The contingency tables to compare “RTPose vs Ramanan”, “RTPose vs MatchPose” and “MatchPose vs Ramanan” in detecting the torso are:

11	15	24	2	54	8
352	177	38	491	309	184

The outcomes ( $\frac{(|n_{01}-n_{10}|-1)^2}{n_{01}+n_{10}}$  = 307.96, 30.77 and 284.31) show that overall RTPose performs significantly better than Ramanan, MatchPose performs significantly better than Ramanan, and RTPose performs significantly better than MatchPose in detecting the torso.

**A3. All images – LUL detection:** The contingency tables to compare “RTPose vs Ramanan”, “RTPose vs MatchPose” and “MatchPose vs Ramanan” in detecting the left upper leg are:

140	23	34	129	112	17
205	187	95	297	232	193

The outcomes  $(\frac{(|n_{01}-n_{10}|-1)^2}{n_{01}+n_{10}} = 143.48, 4.65 \text{ and } 184)$  show that overall RTPose performs significantly better than Ramanan, MatchPose significantly better than Ramanan, and MatchPose significantly better than RTPose in detecting the left upper leg.

**B1. High illumination – head detection:** The contingency tables to compare “RTPose vs Ramanan”, “RTPose vs MatchPose” and “MatchPose vs Ramanan” are:

45	10	27	28	40	11
98	308	24	382	103	307

The outcomes  $(\frac{(|n_{01}-n_{10}|-1)^2}{n_{01}+n_{10}} = 69.69, 0.25 \text{ and } 72.83)$  show that RTPose and MatchPose both perform significantly better than Ramanan, and RTPose and MatchPose have similar performance.

**B2. High illumination – torso detection:** The contingency tables to compare “RTPose vs Ramanan”, “RTPose vs MatchPose” and “MatchPose vs Ramanan” are:

10	13	23	0	49	11
202	236	37	401	163	238

The outcomes  $(\frac{(|n_{01}-n_{10}|-1)^2}{n_{01}+n_{10}} = 164.4, 34.81 \text{ and } 131.17)$  show that in high illumination data RTPose performs significantly better than Ramanan, MatchPose performs significantly better than Ramanan, and RTPose performs significantly better than MatchPose in detecting the torso.

**B3. High illumination – LUL detection:** The contingency tables to compare “RTPose vs Ramanan”, “RTPose vs MatchPose” and “MatchPose vs Ramanan” are:

117	21	33	105	89	22
247	76	78	245	276	75

The outcomes ( $\frac{(|n_{01}-n_{10}|-1)^2}{n_{01}+n_{10}} = 188.79, 3.89$  and  $214.36$ ) show that in high illumination data RTPose performs significantly better than Ramanan, MatchPose performs significantly better than Ramanan, and MatchPose performs significantly better than RTPose in detecting the left upper leg.

**C1. Low illumination – head detection:** The contingency tables to compare “RTPose vs Ramanan”, “RTPose vs MatchPose” and “MatchPose vs Ramanan” are:

2	5	1	6	0	6
43	45	5	83	45	44

The outcomes ( $\frac{(|n_{01}-n_{10}|-1)^2}{n_{01}+n_{10}} = 28.14, 0$  and  $27.88$ ) show that in low illumination data RTPose performs significantly better than Ramanan, MatchPose performs significantly better than Ramanan, and RTPose has similar performance to MatchPose in detecting the head.

**C2. Low illumination – torso detection:** The contingency tables to compare “RTPose vs Ramanan”, “RTPose vs MatchPose” and “MatchPose vs Ramanan” are:

1	2	1	2	0	2
18	73	1	90	19	73

The outcomes ( $\frac{(|n_{01}-n_{10}|-1)^2}{n_{01}+n_{10}} = 11.23, 0$  and  $12.12$ ) show that in low illumination data RTPose performs significantly better than Ramanan, MatchPose performs significantly better than Ramanan, and RTPose has similar performance to MatchPose in detecting the torso.

**C3. Low illumination – LUL detection:** The contingency tables to compare “RTPose vs Ramanan”, “RTPose vs MatchPose” and “MatchPose vs Ramanan” are:

22	2	1	23	17	2
26	44	18	52	31	44

The outcomes ( $\frac{(|n_{01}-n_{10}|-1)^2}{n_{01}+n_{10}} = 18.41, 0.52$  and  $23.89$ ) show that in low illumination data RTPose performs significantly better than Ramanan, MatchPose performs significantly better than Ramanan, and RTPose has similar performance to MatchPose in detecting the left upper leg .

**D1. With cover – head detection:** The contingency tables to compare “RTPose vs Ramanan”, “RTPose vs MatchPose” and “MatchPose vs Ramanan” are:

42	12	25	29	47	7
179	280	29	431	174	285

The outcomes ( $\frac{(|n_{01}-n_{10}|-1)^2}{n_{01}+n_{10}}$  = 144.36, 0 and 152) show that RTPose performs significantly better than Ramanan, MatchPose has similar performance to RTPose, and MatchPose performs significantly better than Ramanan in detecting the head from the data with occlusion.

**D2. With cover – torso detection:** The contingency tables to compare “RTPose vs Ramanan”, “RTPose vs MatchPose” and “MatchPose vs Ramanan” are:

11	12	23	0	54	7
343	147	38	453	300	152

The outcomes ( $\frac{(|n_{01}-n_{10}|-1)^2}{n_{01}+n_{10}}$  = 306.8, 35.78 and 278.15) show that RTPose performs significantly better than Ramanan, MatchPose performs significantly better than Ramanan, and RTPose performs significantly better than MatchPose in detecting the torso from the data with occlusion.

**D3. With cover – LUL detection:** The contingency tables to compare “RTPose vs Ramanan”, “RTPose vs MatchPose” and “MatchPose vs Ramanan” are:

137	17	31	123	105	16
181	178	90	269	213	179

The outcomes ( $\frac{(|n_{01}-n_{10}|-1)^2}{n_{01}+n_{10}}$  = 133.92, 4.89 and 167.78) show that RTPose performs significantly better than Ramanan, MatchPose performs significantly better than Ramanan, and MatchPose performs significantly better than RTPose in detecting the left upper leg from the data with occlusion.

**E1. Without cover – head detection:** The contingency tables to compare “RTPose vs Ramanan”, “RTPose vs MatchPose” and “MatchPose vs Ramanan” are:

2	6	3	5	2	1
17	16	0	34	18	21

The outcomes ( $\frac{(|n_{01}-n_{10}|-1)^2}{n_{01}+n_{10}}$  = 4.58, 4.84 and 13.69) show that MatchPose performs significantly better than RTPose, MatchPose performs significantly better than Ramanan, and RTPose performs significantly better than Ramanan in detecting the head from the data without occlusion.

**E2. Without cover – torso detection:** The contingency tables to compare “RTPose vs Ramanan”, “RTPose vs MatchPose” and “MatchPose vs Ramanan” are:

0	3	1	2	0	1
9	30	0	39	9	32

The outcomes ( $\frac{(|n_{01}-n_{10}|-1)^2}{n_{01}+n_{10}}$  = 2.28, 0.21 and 4.88) show that MatchPose performs significantly better than Ramanan, RTPose has similar performance to Ramanan, and MatchPose and RTPose have similar performances in detecting the torso from the data without occlusion.

**E3. Without cover – LUL detection:** The contingency tables to compare “RTPose vs Ramanan”, “RTPose vs MatchPose” and “MatchPose vs Ramanan” are:

2	6	3	5	7	1
24	10	5	28	19	15

The outcomes ( $\frac{(|n_{01}-n_{10}|-1)^2}{n_{01}+n_{10}}$  = 9.68, 0.09 and 14.51) show that RTPose performs significantly better than Ramanan, MatchPose performs significantly better than Ramanan, and RTPose has similar performance to MatchPose in detecting the left upper leg from the data without occlusion.

# Appendix E

## Auxiliary Forms and Documents

A number of documents are attached, including

- Information Sheet for Symptomatic Volunteer
- Information sheet for Non-Symptomatic Volunteer
- General Consent Form for Symptomatic Volunteer
- General Consent Form for Non-Symptomatic Volunteer
- Audio and Video Consent Form
- GP Letter for Symptomatic Volunteer
- GP Letter for Non-Symptomatic Volunteer
- Questionnaire: to assess sleep quality of the volunteer and pre-classify the volunteer as a normal participant or as a symptomatic patient
- Advertisement: to recruit volunteers

Centre Number: ULHT  
Study Number: 08/H0401/12

## **Information Sheet – Symptomatic Volunteer**

### **The Role of Movement Monitoring in the Assessment of Sleep Disorders – a pilot study**

You have been identified by your Medical Consultant as someone who could help us by taking part in this research study. Before you agree to take part you need to understand why the research is being done and what it would involve for you. Please take time to read the following information carefully.

### **Part 1: What you need to know**

#### **What is the purpose of the study?**

The idea of this research is to find out if computerised video recordings can be used to identify people who have sleep problems where they stop breathing (sleep apnoea).

#### **Why have I been invited to take part?**

You are being asked to participate as a member of a group of people who do have sleep apnoea, so we can compare the results of people who do not have the condition with those who do.

#### **Do I have to take part?**

No. It is up to you to decide. Read this information and if you are interested in taking part please ring the hospital on 01522 573684 to let us know. If you ring between 9:00 and 13:00 Monday-Friday someone should be available to agree a time to ring you back to discuss the arrangements for the overnight study (an answer phone telling you it is the newborn hearing screening office will be heard at other times). We will ask you a few questions about your health to check you are suitable to be in the study and answer any questions you have. If you are suitable and are still interested we will then ask you to sign a consent form to show you have agreed to take part.

Even after you have signed the consent form you can change your mind and decide not to go ahead, without giving a reason.

You can ring the number above to discuss anything about the research with one of the researchers.

#### **What will happen to me if I take part?**

If you agree to take part you will be given a date when you will spend a night at Lincoln County Hospital, sleeping in a normal bed in a single room (called a 'Sleep Study Room'). You will wear the normal testing system that patients wear. It is a small device which will record the oxygen level in your blood – using a little gadget that just shines a light through your finger (see picture).



**NB This system is non-invasive i.e. you will NOT be required to have any injections.**

The research part will be done using 3 video cameras on the walls/ceiling with the pictures and sound stored on a computer system. The audio/video recording will be started by hospital staff once you are ready to settle down for the night. You will be able to get up during the night e.g. to use the adjacent toilet facilities. In the morning, you will be offered a drink, the testing equipment will be taken off and you can leave.

**What will I have to do?**

You will have been given details of where to go in the hospital and the date and time to attend. You will need to bring with you suitable nightwear ( e.g. pyjamas or tee shirt and shorts) and anything you would normally take with you for a night away (e.g. medication). After the night in the sleep study room you will not have to do anything else.

**What will happen to my details and the recordings?**

Your details will be kept confidential within the hospital and only the audio/video and oxygen level data will taken to the University identified by a study number. The data will be kept safe and we won't show it to anyone other than those involved with the research without asking your permission. At the end of the research we will send you a summary of what we have found if you want us to.

The video data will be analysed by researchers at the University of Lincoln, using a special programme to see if they can tell the difference between people who are sleeping normally and those who have sleep apnoea (sleep apnoea is stopping and starting breathing during sleep). How well the computer analysis works will be judged against the opinion of a doctor who knows about sleep apnoea and checked against the normal testing system.

**What about expenses?**

If you take part in the study you will receive £50 from the university in order to cover your expenses including your trips to the hospital.

**What are the possible disadvantages and risks of taking part?**

As a symptomatic volunteer we would expect to find evidence of sleep apnoea in your recordings. There is a however a very small chance of us seeing something else that may indicate another condition as well such as nocturnal seizures. If you have agreed to being informed of any such findings we will advise you of any such findings and provide your GP with any details that are requested. Such findings have the benefit of picking up such conditions early but may have implications for future employment, driving or insurance.

**What are the possible benefits of taking part?**

There are not intended to be any benefits for you in taking part in this study but the information we get from the study should help us to diagnose people with sleep apnoea in the future. As mentioned in the risk section we might find an unexpected condition which you could benefit from being picked up and treated early.

If the information in Part 1 has interested you and you are considering participating, please read the additional information in Part 2 before making any decision

## **Part 2: Additional information**

### **What will happen if you wish to show anyone else my recording?**

Even though you will have signed a consent form before the video data is recorded we will not show any of the recording which will result in you being recognised to anyone not involved with the research without coming back to you to ask for your specific permission. If we do this we will explain why we wish to show the recording and you will have the opportunity to view the recording and you are free to refuse to give your consent. If you give your consent for the video to be used, you will be free to withdraw that consent at any time. Although we will respect your wishes and stop using the recording if you withdraw that consent, if the use has already resulted in publication then withdrawing consent may not be effective.

### **What if there is a problem?**

#### **Complaints**

If you have any concern about any aspect of the study, you should ask to speak to one of the research team who will do their best to deal with your concern (Dr Neil Gravill, Consultant Clinical Scientist at the hospital on 01522 573684 or Ching-Wei Wang, PhD Researcher at the university on 01522 837107). If you remain unhappy and you wish to complain formally, you can do this through the Research and Development Dept at the hospital or via the NHS Complaints Procedure. Details can be obtained from the hospital.

#### **Harm**

In the (extremely unlikely) event that something does go wrong and you are harmed during the research and it is due to someone's negligence then you may have grounds for a legal action for compensation against United Lincolnshire Hospitals NHS Trust. The normal National Health Service complaints mechanism will still be available to you.

### **What will happen to the results of the research study?**

The computer analysis work will be done by a research student at the university and will be used in their PhD thesis and the findings may also be published in scientific journals. If the research shows that the work is useful it may be used in a new piece of medical equipment.

*All research in the NHS is looked at by an independent group of people, called a Research Ethics Committee to protect your safety, rights, wellbeing and dignity. This study has been reviewed and given a favourable opinion by the Derbyshire Research Ethics Committee.*

Centre Number: ULHT  
Study Number: 08/H0401/12

## Information Sheet – Normal Volunteer

### **The Role of Movement Monitoring in the Assessment of Sleep Disorders – a pilot study**

You have shown an interest in taking part in this research study. Before you agree to take part you need to understand why the research is being done and what it would involve for you. Please take time to read the following information carefully.

#### **Part 1: What you need to know**

##### **What is the purpose of the study?**

The idea of this research is to find out if computerised video recordings can be used to identify people who have sleep problems where they stop breathing (sleep apnoea).

##### **Why have I been invited to take part?**

You are being asked to participate as a member of a group of people who do not have sleep apnoea, so we can compare the results of people who do not have the condition with those who do.

##### **Do I have to take part?**

No. It is up to you to decide. Read this information and if you are interested in taking part please ring the hospital on 01522 573684 to let us know. If you ring between 9:00 and 13:00 Monday-Friday someone should be available to agree a time for you to come up to the hospital to discuss the arrangements for the overnight study (an answer phone telling you it is the newborn hearing screening office will be heard at other times). At this visit we will ask you a few questions about your health to check you are suitable to be in the study and answer any questions you have. If you are suitable and are still interested we will then ask you to sign a consent form to show you have agreed to take part.

Even after you have signed the consent form you can change your mind and decide not to go ahead, without giving a reason.

You can ring the number above to discuss anything about the research with one of the researchers.

##### **What will happen to me if I take part?**

If you agree to take part you will be given a date when you will spend a night at Lincoln County Hospital, sleeping in a normal bed in a single room (called a 'Sleep Study Room'). You will wear the normal testing system that patients wear. It is a small device which will record the oxygen level in your blood – using a little gadget that just shines a light through your finger (see picture).



**NB This system is non-invasive i.e. you will NOT be required to have any injections.**

The research part will be done using 3 video cameras on the walls/ceiling with the pictures and sound stored on a computer system. The audio/video recording will be started by hospital staff once you are ready to settle down for the night. You will be able to get up during the night e.g. to use the adjacent toilet facilities. In the morning, you will be offered a drink, the testing equipment will be taken off and you can leave.

**What will I have to do?**

You will have been given details of where to go in the hospital and the date and time to attend. You will need to bring with you suitable nightwear ( e.g. pyjamas or tee shirt and shorts) and anything you would normally take with you for a night away (e.g. medication). After the night in the sleep study room you will not have to do anything else.

**What will happen to my details and the recordings?**

Your details will be kept confidential within the hospital and only the audio/video and oxygen level data will taken to the University identified by a study number. The data will be kept safe and we won't show it to anyone other than those involved with the research without asking your permission. At the end of the research we will send you a summary of what we have found if you want us to.

The video data will be analysed by researchers at the University of Lincoln, using a special programme to see if they can tell the difference between people who are sleeping normally and those who have sleep apnoea (sleep apnoea is stopping and starting breathing during sleep). How well the computer analysis works will be judged against the opinion of a doctor who knows about sleep apnoea and checked against the normal testing system.

**What about expenses?**

If you take part in the study you will receive £50 from the university in order to cover your expenses including your trips to the hospital.

**What are the possible disadvantages and risks of taking part?**

As a normal volunteer we would not expect to find any evidence of sleep apnoea in your recordings. There is a however a very small chance of us seeing something that may indicate this or another condition such as nocturnal seizures. If you have agreed to being informed of any such findings we will advise you of any such findings and provide your GP with any details that are requested. Such findings have the benefit of picking up such conditions early but may have implications for future employment, driving or insurance.

**What are the possible benefits of taking part?**

There are not intended to be any benefits for you in taking part in this study but the information we get from the study should help us to diagnose people with sleep apnoea in the future. As mentioned in the risk section we might find an unexpected condition which you could benefit from being picked up and treated early.

If the information in Part 1 has interested you and you are considering participating, please read the additional information in Part 2 before making any decision

## **Part 2: Additional information**

### **What will happen if you wish to show anyone else my recording?**

Even though you will have signed a consent form before the video data is recorded we will not show any of the recording which will result in you being recognised to anyone not involved with the research without coming back to you to ask for your specific permission. If we do this we will explain why we wish to show the recording and you will have the opportunity to view the recording and you are free to refuse to give your consent. If you give your consent for the video to be used, you will be free to withdraw that consent at any time. Although we will respect your wishes and stop using the recording if you withdraw that consent, if the use has already resulted in publication then withdrawing consent may not be effective.

### **What if there is a problem?**

#### **Complaints**

If you have any concern about any aspect of the study, you should ask to speak to one of the research team who will do their best to deal with your concern (Dr Neil Gravill, Consultant Clinical Scientist at the hospital on 01522 573684 or Ching-Wei Wang, PhD Researcher at the university on 01522 837107). If you remain unhappy and you wish to complain formally, you can do this through the Research and Development Dept at the hospital or via the NHS Complaints Procedure. Details can be obtained from the hospital.

#### **Harm**

In the (extremely unlikely) event that something does go wrong and you are harmed during the research and it is due to someone's negligence then you may have grounds for a legal action for compensation against United Lincolnshire Hospitals NHS Trust. The normal National Health Service complaints mechanism will still be available to you.

### **What will happen to the results of the research study?**

The computer analysis work will be done by a research student at the university and will be used in their PhD thesis and the findings may also be published in scientific journals. If the research shows that the work is useful it may be used in a new piece of medical equipment.

*All research in the NHS is looked at by an independent group of people, called a Research Ethics Committee to protect your safety, rights, wellbeing and dignity. This study has been reviewed and given a favourable opinion by the Derbyshire Research Ethics Committee.*



UNIVERSITY OF  
LINCOLN

Centre Number: ULHT

Study Number: 08/H0401/12

Trial Patient Identification Number:

## GENERAL CONSENT - Symptomatic Volunteers

### The Role of Movement Monitoring in the Assessment of Sleep Disorders – a pilot study

Thank you for volunteering to help with this project. In order to study the changes that occur in the movement of people with sleep disorders it is important to learn about people with known sleep disorders. This form and the information sheet you will have been given is intended to confirm you understand what is involved and you are happy to participate.

Please initial box to agree

1. I confirm that I have read and understand the information sheet dated ..... (version . ....) for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily. ☐
2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, without my medical care or legal rights being affected. ☐
3. I understand that anonymized data collected during the study may be looked at by responsible individuals from the University of Lincoln. ☐
4. Only Clinical staff from hospital will have access to my medical records. I give permission for these individuals to have access to my records. ☐
5. I am aware that participation will involve audio/video recording from which I may be recognisable. I understand that these recordings, or extracts from them, will not be made available to anyone other than those described without my separate consent. ☐
6. I agree to my GP being informed of my participation in the study. ☐
7. If anything is found during the study that may suggest any possible medical condition I wish to be advised of this. Contact details given overleaf. ☐
8. I agree to take part in the above study. ☐

Please complete contact details overleaf if you wish to receive a summary of the findings of this work

\_\_\_\_\_  
Name of Volunteer

\_\_\_\_\_  
Date

\_\_\_\_\_  
Signature

\_\_\_\_\_  
Name of Person taking consent

\_\_\_\_\_  
Date

\_\_\_\_\_  
Signature

\_\_\_\_\_  
Name Researcher

\_\_\_\_\_  
Date

\_\_\_\_\_  
Signature

When completed, copy for patient; copy for researcher site file; original to be kept in medical notes

My contact details are as follows.

Name: .....

Address .....

.....

.....

.....

Email address .....

Phone .....

I **would / would not** wish to receive a summary of the findings of this study. I understand this may not be available until 3 years time.

I would wish to receive this as a paper / electronic (email) copy



UNIVERSITY OF  
LINCOLN

Centre Number: ULHT

Study Number: 08/H0401/12

Trial Patient Identification Number:

## GENERAL CONSENT – Non-Symptomatic Volunteer

### The Role of Movement Monitoring in the Assessment of Sleep Disorders – a pilot study

Thank you for volunteering to help with this project. In order to study the changes that occur in the movement of people with sleep disorders it is important to know about people without sleep disorders. This form and the information sheet you will have been given is intended to confirm you understand what is involved and you are happy to participate.

Please initial box

1. I confirm that I have read and understand the information sheet dated .....  
(version . ....) for the above study. I have had the opportunity to consider  
the information, ask questions and have had these answered satisfactorily. ☐
2. I understand that my participation is voluntary and that I am free to withdraw at  
any time, without giving any reason, without my medical care or legal rights  
being affected. ☐
3. I understand that anonymized data collected during the study may be looked at  
by responsible individuals from the University of Lincoln. ☐
4. I am aware that participation will involve audio/video recording from which I  
may be recognisable. I understand that these recordings, or extracts from  
them, will not be made available to anyone other than those described without  
my separate consent. ☐
5. I agree to my GP being informed of my participation in the study. ☐
6. If anything is found during the study that may suggest any possible medical  
condition I wish to be advised of this. Contact details given overleaf. ☐
7. I agree to take part in the above study. ☐

Please complete contact details overleaf if you wish to receive a summary of the findings of this work.

Name of GP

Address of GP

Address of Volunteer (As registered with GP)

\_\_\_\_\_  
Name of Volunteer

\_\_\_\_\_  
Date

\_\_\_\_\_  
Signature

\_\_\_\_\_  
Name of Person taking consent

\_\_\_\_\_  
Date

\_\_\_\_\_  
Signature

\_\_\_\_\_  
Name Researcher

\_\_\_\_\_  
Date

\_\_\_\_\_  
Signature

When completed, copy for patient; copy for researcher site file; original to be kept in medical notes

My contact details are as follows.

Name: .....

Address .....

.....

.....

.....

Email address .....

Phone .....

I **would / would not** wish to receive a summary of the findings of this study. I understand this may not be available until 3 years time.

I would wish to receive this as a    paper    /    electronic (email) copy

Centre Number: ULHT  
Study Number: 08/H0401/12  
Trial Patient Identification Number:

## CONSENT FORM – Specific Use of Audio/Video

**The Role of Movement Monitoring in the Assessment of Sleep Disorders – a pilot study**

### Details of Audio/Video Recording

Recording Start Date:     /     /	Location:
Recording reference No:	
Specific use required:	

**Please initial box**

- I confirm that it has been explained to me that I am being asked for my consent to allow the audio/video recording, or extracts from them, of my participation in the above project to be used for the specific purpose given above.
- I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, without my medical care or legal rights being affected.
- Whilst the researchers will respect my wishes and stop using the recording if I withdraw consent I understand that if the use has resulted in publication then withdrawing consent may not be effective.

☐
☐
☐

\_\_\_\_\_  
Name of Patient/Participant

\_\_\_\_\_  
Date

\_\_\_\_\_  
Signature

\_\_\_\_\_  
Name of Person taking consent

\_\_\_\_\_  
Date

\_\_\_\_\_  
Signature

\_\_\_\_\_  
Name Researcher

\_\_\_\_\_  
Date

\_\_\_\_\_  
Signature

When completed, copy for patient; copy for researcher site file; original to be kept in medical notes

Participant's GP  
Practice address  
Practice address  
Practice address  
Practice address

Tel: 01522 512512

[www.ulh.nhs.uk](http://www.ulh.nhs.uk)

Direct Line: 01522 573678 Fax: 01522 529858  
Email: [neil.gravill@ulh.nhs.uk](mailto:neil.gravill@ulh.nhs.uk)

Re:

Participant's Name  
DOB  
NHS No (if known)  
Address

The above patient of yours who is under the care of Dr S Matusiewicz, Consultant Respiratory Physician has agreed that we can inform you that they have volunteered as a Symptomatic Participant in the following Research Project:

**The Role of Movement Monitoring in the Assessment of Sleep Disorders – a pilot study**

**Centre Number:** ULHT  
**Study Number:** 08/H0401/12

Which has been granted Ethics Committee approval by the Derbyshire Research Ethics Committee.

Your patient will be spending a night in Lincoln County Hospital sleep lab where they will be monitored by a Pulse Oximeter as well as Audio & Video monitoring. The Video recording will be analysed and used to produce a template of sleep movement in the participants who have no history of sleep disorder.

We would not anticipate any risks to your patient that you need to be aware of. Your patient's care will not be changed by their involvement in this study.

Yours sincerely

**Neil Gravill**  
**Consultant Clinical Scientist (Head of Clinical Measurement)**

Participant's GP  
Practice address  
Practice address  
Practice address  
Practice address

Tel: 01522 512512

[www.ulh.nhs.uk](http://www.ulh.nhs.uk)

Direct Line: 01522 573678 Fax: 01522 529858  
Email: [neil.gravill@ulh.nhs.uk](mailto:neil.gravill@ulh.nhs.uk)

Re:

Participant's Name  
DOB  
NHS No (if known)  
Address

The above patient of yours has agreed that we can inform you that they have volunteered as a Non-Symptomatic participant in the following Research Project:

**The Role of Movement Monitoring in the Assessment of Sleep Disorders – a pilot study**

**Centre Number: ULHT**  
**Study Number: 08/H0401/12**

Which has been granted Ethics Committee approval by the Derbyshire Research Ethics Committee.

Your patient will be spending a night in Lincoln County Hospital sleep lab where they will be monitored by a Pulse Oximeter as well as Audio & Video monitoring. The Video recording will be analysed and used to produce a template of sleep movement in the participants who have no history of sleep disorder.

We would not anticipate any risks to your patient that you need to be aware of. We will not be looking to diagnose any medical condition in your patient, although we will advise you if any possible health issues are indicated during the study.

Yours sincerely

**Neil Gravill**  
**Consultant Clinical Scientist (Head of Clinical Measurement)**

Centre Number: ULHT

Study Number: 08/H0401/12

Patient Identification Number for this trial:

## Questionnaire

**Project Title: The Role of Movement Monitoring in the Assessment of Sleep Disorders – a pilot study**

Please answer the following questions truthfully, to the best of your knowledge. The answers will be used to decide if it is appropriate for you to be included as a normal (non sleep apnoea) participant or as a symptomatic (sleep apnoea) patient for the purposes of this project. This questionnaire will be stored securely, together with your consent form in the hospital. Any of the information given to the university will be known only by your Patient Identification Number.

Date of Birth  /  /  Weight  Height  Gender  M / F

Do you snore regularly  Y / N

Do you suffer from insomnia or daytime sleepiness  Y / N

Are you aware of any sleep related disorder  Y / N

Score from Epworth Sleepiness Scale (attached)

What is your average nightly sleep time (in hours)

Do you have any of the following medical conditions:

CONDITION	YES	NO
Asthma	<input type="text"/>	<input type="text"/>
Chronic Lung Disease	<input type="text"/>	<input type="text"/>
Diabetes	<input type="text"/>	<input type="text"/>
Thyroid Problems	<input type="text"/>	<input type="text"/>
High Blood Pressure	<input type="text"/>	<input type="text"/>
Cardiac(Heart) condition	<input type="text"/>	<input type="text"/>

Do you take any regular Prescription Medication  Y / N

Details .....

\_\_\_\_\_  
Name of Patient/Participant

\_\_\_\_\_  
Date

\_\_\_\_\_  
Signature

\_\_\_\_\_  
Name of Person completing

\_\_\_\_\_  
Date

\_\_\_\_\_  
Signature

# EPWORTH SLEEPINESS SCALE

Name.....

How likely are you to doze off or fall asleep in the situations described in the box below, in contrast to feeling just tired?

This refers to your usual way of life in recent times.

Even if you haven't done some of these things recently try to work out how they would have affected you.

Use the following scale to chose the most appropriate number for each situation:-

0 = would never doze

1 = Slight chance of dozing

2 = Moderate chance of dozing

3 = High chance of dozing

Situation	Chance of Dozing
Sitting and reading	
Watching TV	
Sitting inactive in a public place (eg a theatre or a meeting)	
As a passenger in a car for an hour without a break	
Laying down to rest in the afternoon when circumstances permit	
Sitting and talking to someone	
Sitting quietly after a lunch without alcohol	
In a car, while stopped for a few minutes in traffic	

TOTAL (Copy onto Questionnaire)

# Sleep Volunteers Wanted

Volunteers are required to participate in a research study – *the role of movement monitoring in the assessment of sleep disorders*. The purpose of this research is to investigate the use of computerised video monitoring in support of the diagnosis of sleep disorders.

Your role would be to spend a night sleeping in the sleep laboratory at the Lincoln County Hospital. The overnight sleep will be captured by video cameras (with sound) and the recording will then be used as a normal sample data to compare against recordings from patients with sleep disorders.

For further information

Please contact:

Ching-Wei Wang

Email: [cweiwang@lincoln.ac.uk](mailto:cweiwang@lincoln.ac.uk)

Phone: 01522 837107