

A Novel Approach to Detect the Obscured Upper Body in application to Diagnosis of Obstructive Sleep Apnea

Ching-Wei Wang* and Andrew Hunter

Abstract—Obstructive Sleep Apnea is increasingly seen as a common and important condition, contributing to sleep disturbance and consequential daytime sleepiness. According to recent research findings, the best predictors of morbidity in individual patients, as assessed by improvements with CPAP (continuous positive airway pressure therapy), are nocturnal oxygen saturation and movements during sleep. Video monitoring and interpretation is less well developed both from a processing and analysis viewpoint due to the relative computational complexity of video processing analysis, as well as application-related technical challenges, involving night vision, obscured human body, variation on human size and behavior and massive video and audio data. In this paper, we propose a novel approach to detect and track upper body parts of the covered human body for real time processing. The locations of the upper body parts allow further analysis of covered body postures and human activity recognition in sleep. The experimental results show that the proposed model is promising to estimate the head and torso locations of the covered human body with various postures, body activities and filming environmental settings.

Index Terms—biomedical pattern analysis, covered human body, image analysis, obstructive sleep apnea.

1. INTRODUCTION

Obstructive Sleep Apnea is increasingly seen as a common and important condition, contributing to sleep disturbance and consequential daytime sleepiness. This has potentially serious consequences for the individual, employers and society as a whole. A wide range of parameters including EEG (electroencephalogram) sleep staging, snoring, change in airway resistance, airflow and respiratory effort as well as oxygen saturation and body movement during “normal sleep” in a Sleep Lab lead to an understanding of the pathophysiology of sleep apnea [1]. According to recent research findings [2]—[4], the best predictors of morbidity in individual patients, as assessed by improvements with CPAP (continuous positive

airway pressure therapy), are nocturnal oxygen saturation and movements during sleep. In addition, these parameters have the advantage of being relatively non-intrusive.

Video Monitoring has been adopted to assist diagnosis on obstructive sleep apnea. Sivan *et al* [6] indicate that results from traditional Polysomnography, which requires intrusive measurements and costly measurement equipment, are highly correlated with video test results. Although pulse oximetry is a well-established technique to analyze oxygen saturation, video monitoring and interpretation is less well developed both from a processing and analysis viewpoint. This is due to the relative computational complexity of video analysis, as well as application-related technical challenges, involving night vision / low illumination, body obscured by cover, variation on human size and behavior and massive video and audio data.

Recognition of covered human body activity appears to be a challenging task. Existing monitoring techniques in the sleep lab [5] utilize motion sensors, patterned sheets and infrared light to compute gross degrees of motion from video recorded throughout the night. However, gross motion suggests only periods of time with movements rather than identifying what the activities are, which still require clinicians to look into substantial amounts of video data for analyzing detailed human sleep activities. This is thus a time-consuming and expensive process.

Laser rangefinders are commonly used in 3D object geometry capture. A major 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. The pressure sensitive mattress is an alternative non-intrusive approach to identify occurrence of movements, and the technique was proposed for monitoring patients’ respiratory activities [22]. However, it is rarely used for activity recognition. To the authors’ best knowledge, the pressure sensitive mattress approach does not seem to have been utilized to analyze human activities. In this project, we did not adopt it because of the high development cost, which also requires additional hardware instead of utilizing existing measurement equipment, such as a video monitoring system. We have evaluated a thermal imaging system [20] for obtaining the covered body posture. However, due to heat retention properties, 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

Manuscript received October 16, 2007. This work is supported jointly by the United Lincolnshire Hospital NHS Trust and the University of Lincoln, UK.

Ching-Wei Wang is with the Video Surveillance and Machine Perception Research Centre, University of Lincoln, Brayford Pool, Lincoln LN6 7TS, United Kingdom (phone: +44 (0)1522837107; fax: +44 (0)1522886771; e-mail: cweiwang@lincoln.ac.uk).

Andrew Hunter is with the Video Surveillance and Machine Perception Research Centre, University of Lincoln, Brayford Pool, Lincoln LN6 7TS, United Kingdom (e-mail: ahunter@lincoln.ac.uk).

1 illustrates the issue, showing an image shortly after a leg movement. The image of the leg on the right indicates the current position; the ghost on the left indicates its previous location.

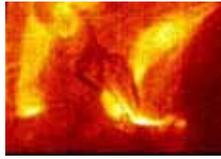


Fig. 1. Test on a Thermal Imaging System, which fails to locate a true body posture since the heat remains on the bed after movements. The right leg in the image is the real leg and the left one is the heat left on the bed.

There is clearly a need for more automated non-invasive methods to recognize human activities during sleep. Thus, our research is focused on developing an intelligent video monitoring system, which will be capable of identifying body postures for covered humans and recognizing human activities during sleep, such as leg movement, body rotation, arm movement, and head movement. To minimize sleep disturbance, monitoring is done using invisible IR lights and cameras.

This research addresses the problem of detecting and segmenting the covered human body. In our case, difficulties arising from night vision, the shifts of the cover surface with movements, obscure body borders under cover, and wrinkle noises 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 [7], [8] because of the large degree of occlusion for long periods. Current research in machine vision [9]–[14] for monitoring or tracking occluded objects focuses on temporary occlusion rather than constant obscuration. The main technical difficulty here is to observe objects constantly under fully or partially cover. Jaeggli *et al.* [15] developed a learned statistical model to predict unobserved features based on partial measurements. However, the method requires data without occlusion to train first and initialize the model, which is not applicable in our case, as each patient comes for one night only. Huang and Jiang [7] presented an iterative method of weighted region consolidation to track a camouflaged object in an outdoor environment. As obscured or camouflaged objects become visible only (or mainly) while in motion, the method locates the object based on pixels with high motion probabilities. However, movements of covered objects lead to motion of the surface around rather than the exact area of the objects, leading to difficulties in segmentation in comparison to camouflaged objects.

In this paper, we propose a novel approach to detect and track the upper body parts of a covered human body in real time. The location of the upper body parts allow further analysis on covered body postures and human activity recognition in sleep. The method includes noise proof preprocessing approaches for extracting obscured object features effectively, a directed head-torso measurement model, which consists of a hierarchical-boosting head detector, a novel head tracker, an interacting mechanism between head detector and head tracker, and a novel torso finder (See Figure 2). The experimental

results show that the proposed model is promising to estimate the head and torso locations of the covered human body precisely with various postures, body activities and smoothness of the cover.

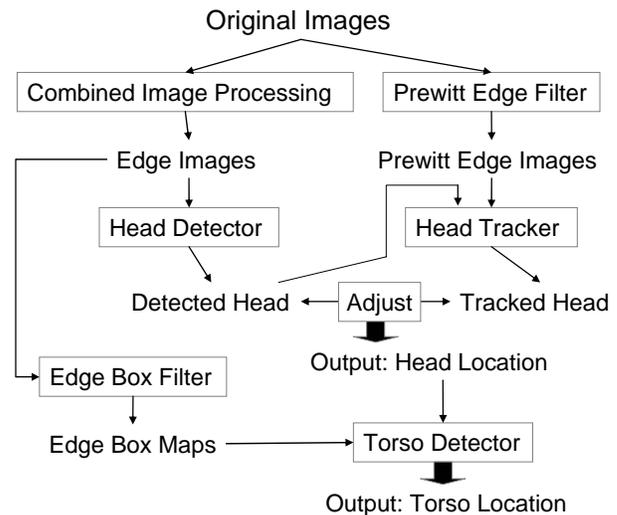


Fig. 2. System Framework

The structure of this paper is as follows. In section 2 we describe preprocessing methods used for feature extraction. Section 3 presents a robust head detector, and section 4 provides the details of a novel head tracker. The novel torso finder is introduced in section 5. The experimental results are shown in section 6. Finally, some failing examples and issues are discussed in section 7, and the conclusion is given in section 8.

2. FEATURE EXTRACTION

As the surface of the cover has uniform texture and color, segmentation methods utilizing texture or color information are not applicable. Hence, the edge information is employed here for extracting object features. Nonetheless, due to the nature of the sheet, large amounts of wrinkle noise are produced. Thus, the first step is to explore a solution that obtains important edges and removes redundant ones at the same time. A combined image processing technique is proposed in section 2.A, producing edge images, which then are used for detecting the head in section 3.

As the remaining edges in the edge images tend to be discontinuous and scattered, an edge box map is created for shape abstraction, which proves to generate a robust and clear outline of the human body. The method is demonstrated in section 2.B. The edge box map is utilized in searching for the torso part in section 5.

A. Edge Image

We 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 Kernels, Prewitt Kernels, Kirsch Compass Kernels and Laplacian [18] inevitably produce noisy information from wrinkles. Due to the horizontal layout of the bed, we propose an oriented horizontal edge detector to

effectively generate object edges aligned with the body, and to remove noise. The oriented horizontal edge detector can be formulated in Equation 1.

$$I'(x, y) = I(x, y) \otimes g(\omega, \nu)$$

$$= \sum_{\omega=-N}^N \sum_{\nu=-N}^N I(x-\omega, y-\nu) g(\omega, \nu) \quad (1)$$

, where $I(x, y)$: input image ,
 $I'(x, y)$: output image ,
 $2N + 1 =$ size of $g(\omega, \nu)$

$$g = \begin{bmatrix} -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

In addition, in order to preserve stability of the edge quality and avoid influences from environment factors, a Gaussian blur filter [19] is applied before the edge detector. Moreover, the image dimension is reduced initially. There are two advantages using down-scale images: scene abstraction; and an increase in computational speed. Figure 3 displays the sequence of the combined image processing techniques for generating edge images, and Figure 4 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.



Fig. 3. Image Processing for Edge Image

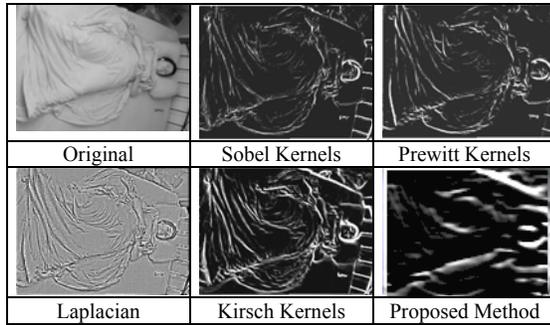


Fig. 4. Resulting images by general edge detectors and the proposed method.

B. Edge Box Map

As the targeted human body is covered, a portion of the edge information is lost. Hence, we develop an edge box map approach for further shape abstraction, to deal with the issue of discontinuous and scattered edges. Figure 5 displays two edge box maps calculated from edge images, which extract edge information from the original images in the left column. The resulting edge box maps are then utilized by the torso detector to find the covered torso part. The torso detector is introduced in section 5.

Given an $(m \times n)$ edge image and an $(s \times s)$ edge box, we generate an edge box map $B(i, j)$, which can be formulated as follows.

$$B(i, j) = \begin{cases} 1 & \text{if } \sum f_{ij}(k, l) \geq \tau \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

$$\text{, where } i \in \{1, \dots, \lfloor m/s \rfloor\} \quad j \in \{1, \dots, \lfloor n/s \rfloor\} \quad k, l \in \{1, \dots, s\}$$

$$f_{ij}(k, l) = \begin{cases} 1 & \text{if } I((i-1)s+k, (j-1)s+l) \geq \nu \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

where $I(x, y)$ is the intensity value of the edge image at location (x, y) . In this work, the following values are used based on our experiments. ($s = 3, \tau = 50, \nu = 30$).

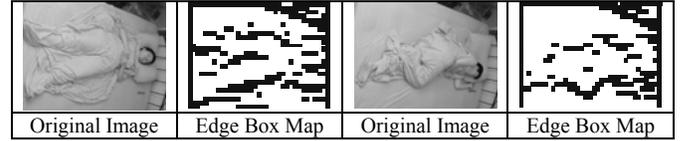


Fig. 5. Original Images and Edge Box Maps.

3. HEAD DETECTOR: HIERARCHICAL ENSEMBLES

There has been considerable work on face detection in computer vision research over the past ten years. However, most of the face detection systems require at least portions of the face to be shown, such as both eyes. Patients may sleep on their side, presenting only half or less of the face. Here, a head detector invariant to facial direction is built, utilizing an ensemble machine learning algorithm [17], which is presented below.

(1) Ensemble Learning Algorithm

Given $(i_1(x, y), c_1) \dots (i_m(x, y), c_m)$

, where $i_j \in I \quad c_j \in C = \{T, F\} \quad x, y \in \{1 \dots 14\}$

Initialise $D_1(j) = \frac{1}{m}$, where $D(j)$: weight of i_j

For $t = 1, \dots, T$

• Find base classifier $h_t : I \rightarrow \{T, F\}$ that

$$h_t = \arg \min_{h_k \in H} \varepsilon_k, \text{ where } \varepsilon_k = \sum_{j=1}^m D_t(j) [c_j \neq h_k(i_j)]$$

• prerequisite : $\varepsilon_k \neq 0 \wedge \varepsilon_k < 0.5$, otherwise resample data

• $\beta_t = \frac{1}{2} \ln \frac{1 - \varepsilon_t}{\varepsilon_t}$, where $\beta_t \in R$

• $D_{t+1}(j) = \frac{D_t(j) \exp(-\beta_t c_j h_t(i_j))}{Z_t}$

, where $\exp(-\beta_t c_j h_t(i_j)) \begin{cases} < 1, & c(j) = h_t(i_j) \\ > 1, & c(j) \neq h_t(i_j) \end{cases}$

Output the final classifier:

$$H(i) = \text{sign} \left(\sum_{t=1}^T \beta_t h_t(i) \right)$$

(2) Re-sample Data Algorithm

Generate random number $r(1), \dots, r(m)$

$$P(a) = \frac{\sum_{i=1}^a r(i)}{\sum_{j=1}^m r(j)}$$

$$\varpi(b) = \sum_{j=1}^b D(j)$$

If $P(a) < \varpi(b)$, add instance b into output dataset and then compare $P(a+1)$ and $\varpi(b+1)$. Otherwise, compare $P(a)$ and $\varpi(b+1)$.

Importantly, we employ edge information instead of original pixels as the input source, avoiding influences from different facial appearance, expression and direction. This helps to learn more robust patterns and rules. The entire machine learning model for head detection is constituted of 3-layer hierarchical ensembles. Each ensemble consists of 10 decision tree base classifiers. The training procedure is decomposed as follows, and the overview of the training scheme is presented in Figure 6(b).

- Step 1: Prepare Training Data (See Figure 6(a))
 - Step 1.1: Select 15 different frontal head images
 - Step 1.2: Select 15 different head images on sides
 - Step 1.3: Randomly select 10 different non-head images
 - Step 1.4: Set the head images as class T and non-head images as class F.
- Step 2: Train 1st layer ensemble machine learning head model using the data collected in Step 1.
- Step 3: Test 1st layer machine learning model and generate a number of false positive instances
- Step 4: Prepare Training Data for 2nd layer model
 - Step 4.1: Randomly select 10 different false positive instances from Step 3
 - Step 4.2: Set the instances from step 4.1 as class F.
- Step 5: Train 2nd layer ensemble machine learning head model using the data collected in Step 1 and Step 4.
- Step 6: Test 2 layers hierarchical machine learning models and generate a number of false positive instances
- Step 7: Prepare Training Data for 3rd layer model
 - Step 7.1: Randomly select 10 different false positive instances from Step 6
 - Step 7.2: Set the instances from step 7.1 as class F.
- Step 8: Train 3rd layer ensemble machine learning head model using the data collected in Step 1, Step 4 and Step 7.
- Step 9: Terminate with 3 layers ensemble machine learning models.

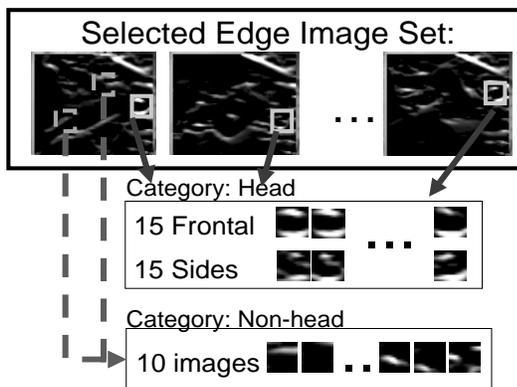


Fig. 6 (a). Collect Training Data Procedure: The format of the training data is intensity values of a 14×14 edge image. $I[x, y]$, where x is valid from 1 to 14 and y is valid from 1 to 14.

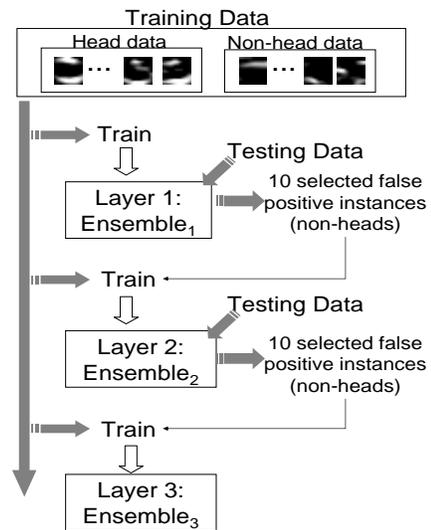


Fig. 6 (b). Training Scheme. Hierarchical structure allows continuously refining patterns and rules by focusing on the false positive instances from previous learning experiences.

The hierarchical structure allows patterns and rules to be continuously refined by focusing on the false positive instances from the previous learning experience. The proposed method utilizes a relatively small number of instances (60 images) to build the machine learning models, and moreover all training data is selected from one single video clip. In evaluation, the experimental results show that the head detector works robustly for locating the head in testing all 17 video data, which are recorded with 4 different filming angles and in 2 different environments. Figure 7 presents the head searching scheme.

Search head over the Region with starting point (x, y) , width (w) and height (h) .

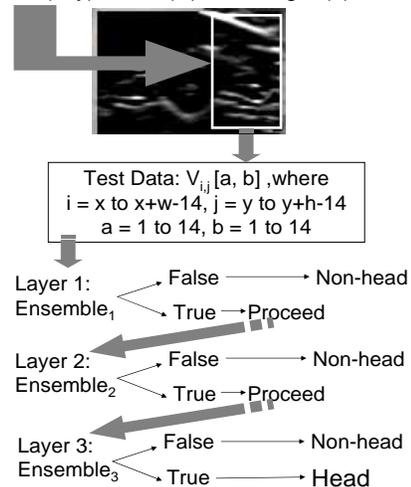


Fig. 7. Search Scheme of the Head Detector.

4. NOVEL HEAD TRACKER

The purpose of the head tracker is to find an optimal head region in a local searching area, which is derived from the location previously identified either by the head detector or the head tracker. Figure 8(a) illustrates the procedure of the head

tracker first starting searching the head in the local area, and Figure 8(b) illustrates the definition of the local search area, which is an expanded area in association with the estimated head location calculated from the head detector. Firstly, the image is processed by a Prewitt Edge Detector, and secondly a constrained area is computed from the detected position. Thirdly, we create a statistic model for searching an optimal location in the constrained area. Fourthly, a final adjustment mechanism is proposed to ensure a good final estimated position by interacting results between the head detector and the head tracker. In addition, a motion detection mechanism is proposed to speed up system computation by avoiding unnecessary actions.

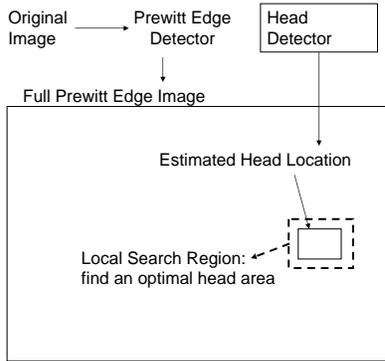


Fig. 8 (a). Overview of the head tracking.

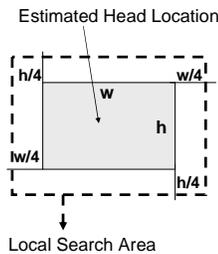


Fig. 8 (b). Define Local Search Area.

1) *Image Pre-processing*: As the aim of local search is to find an optimal target's region, simulating human brain operation, the local search conducts search in a more detailed aspect, and thus requires more detailed information than detection. Therefore, instead of using the horizontal edge detector, which largely abstracts the image, the Prewitt edge detector is employed to pre-process image for the head tracker. Figure 9 compares three edge detectors, including horizontal edge detector, the Kirsh Kernels and the Prewitt Kernels. For a clear illustration, the head locations are superimposed on the images in Figure 9 in order to help readers to identify the head around area. The images show that the Prewitt Kernels perform best to preserve edges over the head while maintaining lowest level of other edges around the head, allowing targeting a precise head position more easily. In the next section, we create a statistic model to calculate a region with the most likelihood as a head.

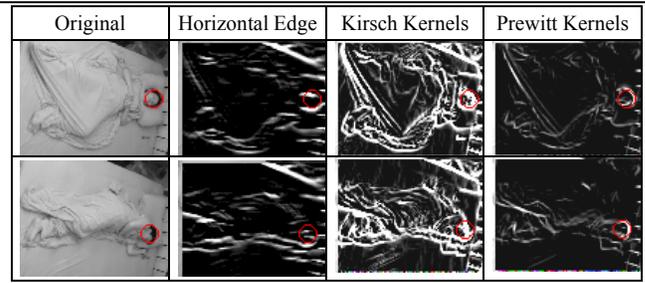


Fig. 9. Comparison on various edge detectors around the head region.

2) *Head Tracker Algorithm*: The aim of the local search algorithm is to target a region in the constrained area with the most likelihood as a head. The algorithm is explained as follows. The local area is firstly processed with a binary filter and being converted into a 2-dimensional binary map. Given a $(m \times n)$ area, we generate a binary map $B(i, j)$, which can be formulated as follows.

$$B(i, j) = \begin{cases} 1 & \text{if } I(i, j) \geq \tau \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

, where $I \in \{0 \dots 255\}$ $i \in \{x \dots x + m\}$ $j \in \{y \dots y + n\}$

(x, y) is the location of the search area and I is the intensity of individual pixel.

Secondly, a scoring mechanism is applied to summarize the valid points in each potential range. Given the binary map $B(i, j)$ and an $(w \times h)$ estimated head region from head detector, we generate a list of scores $S(v, r)$, which can be formulated in Equation 5.

$$S(v, r) = \sum_{q=r}^{r+h} \sum_{p=v}^{v+w} B(p, q) \quad (5)$$

, where $v \in \{x, \dots, x + m - w\}$ $r \in \{y, \dots, y + n - h\}$

(x, y) is the location of the local search area with size $(m \times n)$

Thirdly, the method votes for a location (k, g) with maximum score. The location is selected due to its most likelihood to be a head region.

$$(k, g) = \arg \max \left(S(v, r) \right) \quad (6)$$

, where $v \in \{x, \dots, x + m - w\}$ $r \in \{y, \dots, y + n - h\}$

3) *Checking & Adjusting Mechanism – Interaction between Head Detector & Head Tracker*: The head detector is composed of a hierarchical ensemble machine learning models and capable of locating the head, and the head tracker targets a more precise head region within the area previously identified either by the head detector or by the head tracker. The concept of the head tracker is to find an optimal area, which contains edges all over the head, and tracks the head while it moves. Most of time, the tracker tracks the head and moves toward where the head goes. However, occasionally the tracker goes toward the torso part. This situation happens in one of our experimental environments, in which the images are less

contrasted and the edges are more blur. Hence, in order to prevent the tracker from moving toward a wrong direction, we create a checking and adjusting mechanism to keep the tracker function well by amending the tracker's behavior based on the estimation of the head detector. Figure 10 illustrates the adjusting mechanism and the interaction between the head detector and the head tracker.

The mechanism first computes the distance between the estimated location from the head detector and the estimated

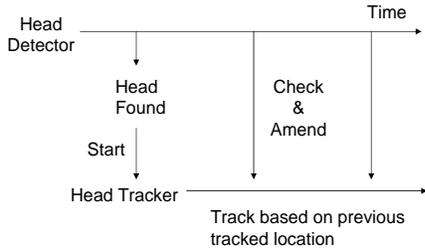


Fig. 10. Checking and Adjusting mechanism between the Head Tracker and Head Detector.

location from the head tracker, and then adjusts the tracker if the distance is too big. Given the estimated location from the head detector (x_1, y_1) , the estimated location from the head tracker (x_2, y_2) and the size of the estimated head region $(w \times h)$, Equation 7 computes the distance between the estimated location from the head detector and the estimated location from the head tracker, and the checking and adjusting mechanism is formulated in Equation 8. (x_3, y_3) is the resulting location after adjustment.

$$D_x = |x_1 - x_2|, \quad D_y = |y_1 - y_2| \quad (7)$$

$$x_3 = \begin{cases} (x_1 + x_2) / 2 & \text{if } D_x > w / 4 \\ x_2 & \text{otherwise} \end{cases} \quad (8)$$

$$y_3 = \begin{cases} (y_1 + y_2) / 2 & \text{if } D_y > h / 4 \\ y_2 & \text{otherwise} \end{cases}$$

4) *Incorporate with Motion Detection*: As the proposed head detector and tracker are able to target the head fast and precisely, the method is stable enough to include the motion detection for exempting searching when there is no motion over the head, saving the computing power, and stabilizing the resulting estimated location. The entire algorithm for detecting and tracking the head and the torso of covered human body is decomposed in the following steps.

- Step 1: Check if there is estimated head location existing.
- Step 2: If there is head found, process motion detection over the estimated head region. Otherwise, go to Step 4.
- Step 3: Check if there is motion over the estimated head region.
 - Step 3.1: If there is no movement, then remain the estimated head location and then process next frame. Otherwise, go to Step 3.2.
 - Step 3.2: Process local search for the head based on existing estimated head location.

- Step 3.3: Track the torso.
- Step 3.4: Process next frame.
- Step 4: Detect the head.
- Step 5: Check if a head is found by Step 4.
- Step 6: If a head is found, process local search for the head based on the detected head location. Otherwise, go to Step 7.
- Step 6.1: Detect the torso.
- Step 7: Process next frame.

5. DIRECTED TORSO FINDER

Humans are articulated objects composed of connecting parts. Hence, the coordinates of the head position are used as the starting point to locate the torso. The core idea to find the torso is to search for a relatively smooth region with reasonable distance and angle from the head, i.e. an area near to the head with a low interior edge box count. Hence, a constrained search range is used to keep limited distance and angle between the head and the torso. The directed torso detector searches over the edge box maps, which largely abstract the output images from edge images and hence reduce the data complexity. Figure 5 presents several edge box maps with the head region found, illustrating more clean views for segmentation of the torso part.

Moreover, the location of the search range is adjusted by the relative level of the head position on margins of the bed considering general body posture (See Figure 11). Also, the size of the region is proportional to the size of the head. Hence, the width of the search range is equal to $1\frac{1}{2} \times$ width of the head and the height of the range is $3 \times$ height of the head.

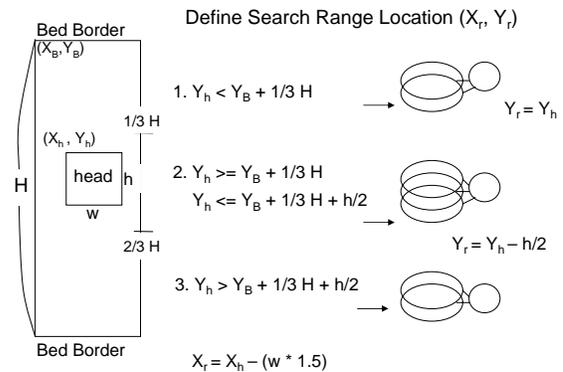


Fig. 11. The location of the torso search region (X_r, Y_r) is derived from the head location (X_h, Y_h) . $X_r = X_h - \Delta g$ and $Y_r = Y_h - \Delta k$. Given the head size (w, h) and the bed size (W, H) , $\Delta g = 1\frac{1}{2} \times w$ and $\Delta k (Y_h) = (1) 0, (2) h/2, \text{ otherwise}$.

A voting mechanism is created to select a region inside the search area as the torso location. The voting mechanism votes for a region with the least number of edge boxes as the most flat region inside the search area.

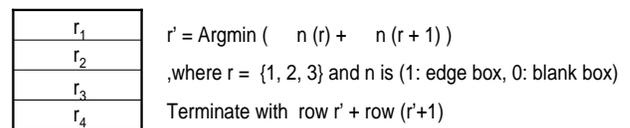


Fig. 12. Voting mechanism to select the smoothest area. $\sum n(r)$ is the accumulation of valid edge boxes within row r .

Given a $(W_r \times H_r)$ torso search range (X_r, Y_r) , the edge box size $(s \times s)$ and the estimated head size $(w \times h)$, we first generate a list of edge box scores E_U by dividing the range into four sub-zones and calculating the number of edges box insides (See Equation 9).

$$E_U = \sum_j \sum_i B(i, j) \quad (9)$$

where $B(i, j) = \begin{cases} 1, & \text{edge box} \\ 0, & \text{blank box} \end{cases}$

$$i \in \left\{ \frac{X_s}{s}, \dots, \frac{X_s + W_s}{s} \right\}$$

$$j \in \left\{ \left[\frac{Y_s + U \times \frac{3h}{4}}{s}, \dots, \left[\frac{Y_s + (U+1) \times \frac{3h}{4}}{s} \right] \right\}$$

$$U \in \begin{cases} \{0, \dots, 3\}, & \text{if } Y_h < Y_B + \frac{H}{3} \cup Y_h > Y_B + \frac{H}{3} + \frac{h}{2} \\ \{0, \dots, 2\}, & \text{otherwise} \end{cases}$$

We then select two connecting zones with the lowest edge box count as the estimated torso area, which can be formulated in Equation 10. The resulting torso location is (X_t, Y_t) with size $(W_t \times H_t)$, which can be formulated in Equation 11.

$$\varphi = \arg \min (E_U + E_{U+1}) \quad (10)$$

$$Y_t = Y_r + \varphi \times \frac{3h}{4}, \quad X_t = X_r \quad (11)$$

$$W_t = \frac{3w}{2}, \quad H_t = \frac{3h}{2}$$

where the estimated head size is $(w \times h)$.

6. EXPERIMENTS

In evaluation, the experiments are conducted on a number of video clips in two different environments with various body postures, human activities and camera shooting angles. The human activities include head movements, limb movements and body rotation. For a quantitative evaluation of the proposed method, we firstly systematically sample 2090 frames from 17 video clips, secondly generate 200 non-repeating random numbers between 1 and 2090, and then select 200 frames according to those random numbers. The randomly selected frames are manually marked to produce a reference standard by an independent individual. The output of the system is then compared to the reference standard. Given the overlapped area ζ of the manually marked area and the estimated area by the proposed algorithm ω , the precision P of each frame is equal to ζ / ω . Furthermore, a recognition rate is defined as $R(P) = n(P) / N$, where $n(P)$ is the sum of frames with precision P and N is the total number of frames sampled from individual video clip. Table 1 presents the recognition rates on the head and the torso with three levels of precision, i.e. $P \geq 0.7$, $P \geq 0.5$ and $P < 0.5$. The results show that the tracking and adjustment mechanisms obtain high recognition rates in the high precision category ($P \geq 0.7$), i.e. 0.96 and 0.97.

Table I. Recognition Rates.

R(P)	$P \geq 0.7$	$P \geq 0.5$	$P < 0.5$
Head	0.96	0.99	0.01
Torso	0.97	0.99	0.01

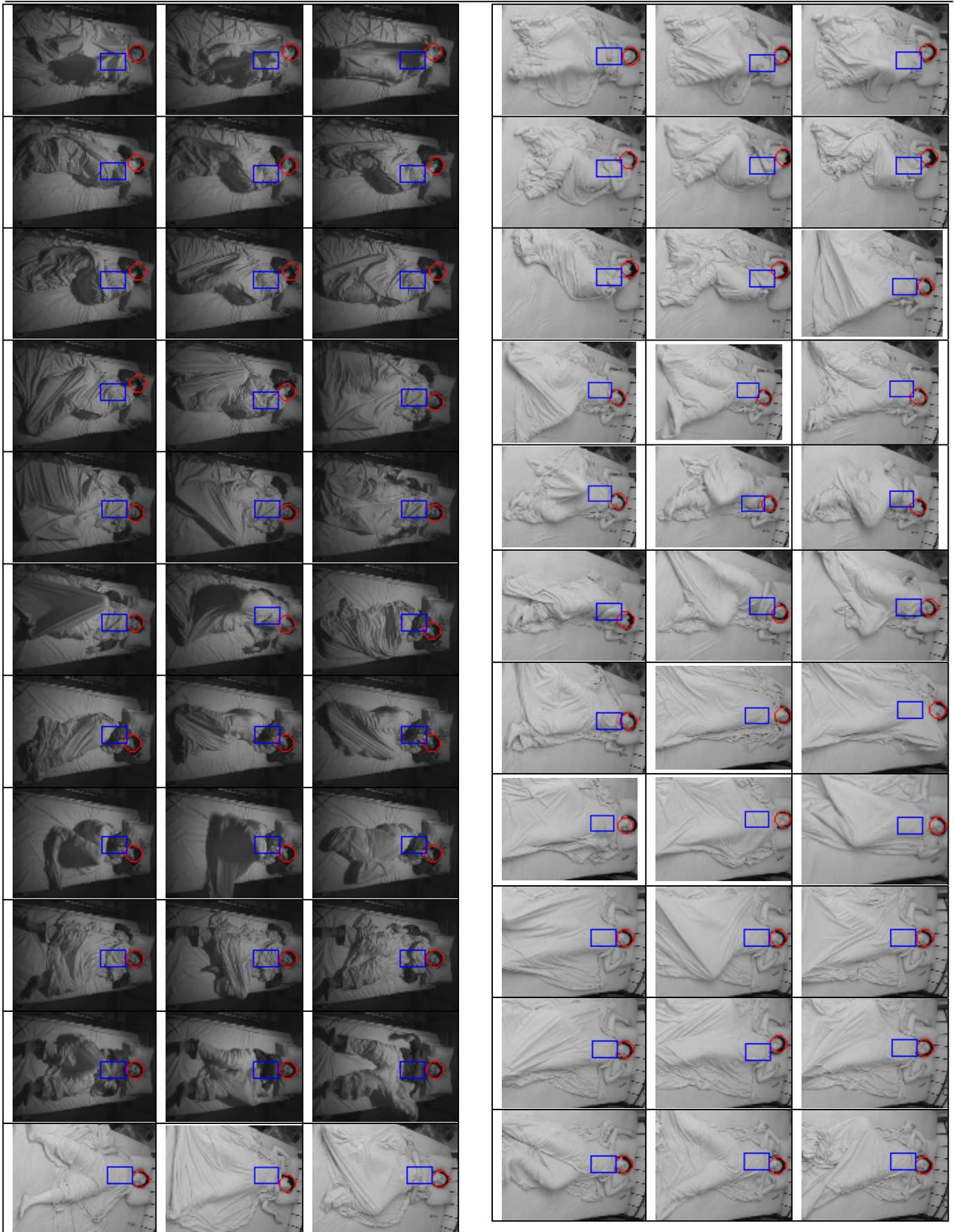
$R(P) = n(P) / N$, where $n(P)$ is the sum of frames with precision rate P and N is the total number of frames sampled.

Although the proposed model is trained and built by using a small number of frames from one single video clip, the experimental results on 200 randomly selected frames from 17 video clips show that the proposed approach works robustly despite different environmental set-up, camera shooting angles, and variations of body postures and movements. Table 2 illustrates some image outputs from our system, presenting that the approach is invariant to different body postures, human activities and environmental set-up.

Regarding the computing efficiency, the system is implemented in C# and able to process 25 fps on a P4 2.4GHz CPU power. The video was acquired with a resolution of 320*240, using a SONY infrared camcorder (DCR-HC-30E).

Table II. System Outputs.





7. DISCUSSIONS

Although the experimental results demonstrate success of our algorithm in estimating upper body parts from the obscured human body, there are a few research issues to be explored. As the aim of this research focuses on building a head-torso measurement model, the proposed model utilizes a simple first come first out strategy and gives higher priority to the middle area in the search order to find a head. Hence, the head position will be more accurately estimated with more complicated search algorithms. Furthermore, the estimation will be more reliable with a more complicated bidirectional head-torso measurement model. In addition, an efficient model is required in validating the estimation for quick responses to rapid movements. Figure 12 demonstrates the failing examples.



Fig. 12. Failing Examples.

8. CONCLUSION

Analysis of the covered human body is a challenging task. We have presented a novel technique for detecting and tracking the upper body parts on the covered human in real time. The proposed method allows further analysis of the activities of the covered human body. The locations of the upper body parts and the proposed edge box map can be utilized to identify the limb posture, allowing recovering / estimating a full human body posture and advanced recognition on covered human body activities, such as leg movements, arm movements, body rotation, getting up, head movements, going to bed, etc. Apart from assisting diagnosis of obstructive sleep apnea, the monitoring technique can be applied to other sleeping disorders or syndromes.

We will continuously develop and enhance methods to locate the rest of the body parts and to recognize human activities during sleep. Currently, we are investigating methods for tracking the covered human body parts. As the detection is applied to every frame in the initial stage, tracking will be added to associate temporal relationships among frames, to stabilize locations of detected body parts and to further improve the computational speed.

Amounts of experiments will be conducted using patients and volunteers in the Sleep Lab at the Lincoln County Hospital to determine performance on a large scale and to generalize across the variations in human sizes, shapes, and behavior. Furthermore, we will develop a system to diagnose movements characteristic of sleep disturbances.

ACKNOWLEDGMENT

We thank Yen-Ting, Fang for marking the head and the torso of the covered human on the 200 randomly selected images as the reference standard in evaluation of the proposed method.

REFERENCES

- [1] W. W. Flemons, M. R. Littner, J. A. Rowley, P. Gay, W. M. Anderson, D. W. Hudgel, R. D. McEvoy, and D. I. Loube, Home Diagnosis of Sleep Apnea: A Systematic Review of the Literature An Evidence Review Cosponsored by the American Academy of Sleep Medicine, the American Thoracic Society CHEST 124(4), 2003, pp. 1543-1579.
- [2] J. C. T. Pepperell, R. J. O. Davies, and J. R. Stradling, Sleep studies for sleep apnoea, *Physiological Measurement*, 23, 2002, pp. R39 – R74. L. S. Bennett, B. A. Langford, J. R. Stradling, and R. J. O. Davies, Sleep fragmentation indices as predictors of daytime sleepiness and nCPAP response in obstructive sleep apnea, *Am J Respir Crit Care Med*, 158, 1998, pp. 778 – 786.
- [3] R. N. Kingshott, M. Vennelle, C. J. Hoy, H. M. Engleman, I. J. Deary, and N. J. Douglas, Predictors of improvements in daytime function outcomes with CPAP therapy, *Am J Respir Crit Care Med*, 161, 2000, pp. 866 – 871.
- [4] Visi-3 Digital Video System, <http://www.stowood.co.uk/page26.html>
- [5] Y. Sivan, A. Kornecki, and T. Schonfeld, Screening obstructive sleep apnoea syndrome by home videotape recording in children, *Eur Respir J*, 9, 1996, pp. 2127 – 2131.
- [6] Z. Q. Huang and Z. Jiang, Tracking Camouflaged Objects with Weighted Region Consolidation, *Proceedings of the IEEE Digital Image Computing: Techniques and Applications*, 2005, pp. 161 – 168.
- [7] T. E. Boulton, R. Micheals, X. Gao, P. Lewis, C. Power, W. Yin, and A. Erkan, Frame-Rate Omnidirectional Surveillance & Tracking of Camouflaged and Occluded Targets, *Proceedings of the Second IEEE Workshop on Visual Surveillance*, 1999, pp. 48 – 55.
- [8] H. T. Nguyen and A. W. M. Smeulders, Fast Occluded Object Tracking by a Robust Appearance Filter, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26, 2004, pp. 1099 – 1104.
- [9] A. Ghafoor, R. N. Iqbal, and S. Khan, Robust image matching algorithm. *Proceedings of Video/Image Processing and Multimedia Communications*, 4th EURASIP Conference focused on Vol. 1, 2003, pp. 155 – 160.
- [10] H. L. Eng, J. Wang, A. H. Kam, and W. Y. Yau, A Bayesian framework for robust human detection and occlusion handling using human shape model. *Proceedings of the 17th International Conference on Pattern Recognition*, 2004, pp. 257 – 260.
- [11] J. Winn and J. Shotton, The Layout Consistent Random Field for Recognizing and Segmenting Partially Occluded Objects. *Proceedings of IEEE Computer Vision and Pattern Recognition*, 2006, pp. 37 – 44.
- [12] B. Chalmond, B. Francesconi, and S. Herbin, Using Hidden Scale for Salient Object Detection. *IEEE Transactions on Image Processing*, 2004, pp. 2644 – 2656.
- [13] R. T. Collins, Y. Liu, and Leordeanu, M., Online Selection of Discriminative Tracking Features. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 27, 2005, pp. 1631 – 1643.
- [14] T. Jaeggli, G. Caenen, R. Fransens, and L. V. Gool, Analysis of Human Locomotion Based on Partial Measurements. *Proceedings of IEEE Motion*, 2005, pp. 248 – 253.
- [15] J. R. Quinlan, Bagging, boosting, and C4.5. *Proceedings of the Thirteenth National Conference on Artificial Intelligence*, 1996, pp. 725 – 730.
- [16] C. W. Wang, New Ensemble Machine Learning Method for Classification and Prediction on Gene Expression Data. *Proceedings of 28th IEEE EMBS Annual International Conference*, 2006, pp. 3478 – 3481.
- [17] M. Seul, L. O'Gorman, and M. J. Sammon, *Practical Algorithms for Image Analysis with CD-ROM*. Cambridge University Press, 2000.
- [18] X. Jiang, *Advanced Techniques for Assessment Surface Topography*. Kogan Page, 2003.
- [19] Raytek Corporation, Fluke ThermoView Ti30 thermal imager, Available: <http://www.radir.com/thermalimager.htm>
- [20] S. Matusiewicz [Position: Consultant Physician], Personal Interview, Medical Physics Department of Lincoln County Hospital in United Kingdom, 2006
- [21] D. Mack, M. Alwan, B. Turner, P. Suratt, and R. Felder, 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 (D2H2)*, 2006.