Unconstrained Video Monitoring of Breathing Behavior and Application to Diagnosis of Sleep Apnea

Ching-Wei Wang*, *Member, IEEE*, Andrew Hunter, *Member, IEEE*, Neil Gravill, and Simon Matusiewicz

***Abstract*—This paper presents a new real-time automated infrared video monitoring technique for detection of breathing anomalies, and its application in the diagnosis of obstructive sleep apnea. We introduce a novel motion model to detect subtle, cyclical breathing signals from video, a new 3-D unsupervised self-adaptive breathing template to learn individuals' normal breathing patterns online, and a robust action classification method to recognize abnormal breathing activities and limb movements. This 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. Moreover, shallow and abdominal breathing patterns do not adversely affect the performance of the method, and it is insensitive to environmental settings such as infrared lighting levels and camera view angles. The experimental results show that the technique achieves high accuracy (94% for the clinical data) in recognizing apnea episodes and body movements and is robust to various occlusion levels, body poses, body movements (i.e., minor head movement, limb movement, body rotation, and slight torso movement), and breathing behavior (e.g., shallow versus heavy breathing, mouth breathing, chest breathing, and abdominal breathing).**

***Index Terms*—Action recognition, behavior analysis, breathing monitoring, obstructive sleep apnea (OSA).**

I. Introduction

OBSTRUCTIVE Sleep Apnea (OSA) [1] is a condition with severe complications including: reduction in cognitive function, cardiovascular disease, stroke, fatigue, and excessive day time sleepiness. OSA is characterized by repetitive obstruction of the upper airways during sleep, resulting in oxygen desaturation and frequent arousal events, characterized by violent awakening. Although OSA affects around 4% of men and 2% of women [1], [2] the majority of affected individuals, perhaps 80–90% [3], [4], remain undiagnosed.

The gold standard diagnostic tool for OSA is Polysomnography (PSG), which measures a wide range of parameters, including brain waves (EEG), eye movements, skeletal muscle activation, electrocardiogram (ECG)/heart rate, airflow, respiratory effort, and blood oxygen saturation using a range of sensors. However, PSG is costly, labor-intensive (not least in analyzing the data), and invasive, which may disturb sleep and compromise the findings.

A popular cost-effective, less invasive alternative combines pulse oximetry (to measure blood oxygen saturation levels [SpO2] and heart rate) with infrared (IR) video monitoring. The clinician identifies suspicious areas on the pulse oximetry trace (defined by a dip of more than 4% in the oxygen saturation level) and reviews the corresponding video data to reach a diagnosis. However, the pulse oximetry traces of some OSA patients do not show all the abnormalities, forcing the clinician to review a significant amount of the video data. To reduce the workload, some existing video systems [5] try to detect patient movement, utilizing patterned sheets and IR light to detect gross degrees of motion, which at least identifies periods of activity, even if it does not identify what the activities are. However, if the patterned cover is removed by the patient, the system fails. There is thus a growing interest in alternative, more robust, automated approaches to the diagnostic assessment of OSA.

The contact-type approaches include thoracic–abdominal bands [7], which track changes in the body circumference during the respiratory cycle, stick-on electrodes such as the ECG method [8], the nasal temperature probe [9], and contact-type microphone for audio analysis to monitor tidal volumes from the human breathing activity [10]. The main disadvantages of these approaches are consequent on their invasiveness: they may be uncomfortable, which disturbs sleep and compromises results; and patient movement may dislodge sensors or compromise readings.

The noninvasive techniques include noncontact audio analysis [11], [12], vibration sensors [13], thermal imaging [14]–[16], and Doppler radar sensors [17], [18] designed to identify breathing. A major challenge for noncontact type audio analysis is the extraction of the breathing sounds from the sensor signals contaminated by the environmental noise [11] [12]. The vibration sensors require an expensive specialized hardware and impose positional and postural constraints. The thermal imaging techniques have been used to capture a breathing signal, by detecting the breath as it is expelled [14]–[16], and the radar sensors have

Manuscript received January 9, 2013; revised August 19, 2013; accepted August 27, 2013. Date of publication August 29, 2013; date of current version January 16, 2014. This work was jointly supported by United Lincolnshire Hospitals NHS Trust and the University of Lincoln. A research Ethics approval was gained from Derbyshire Research Ethics Committee (REC number 08/H0401/12). *Asterisk indicates corresponding author.*

*C.-W. Wang is with the Graduate Institute of Biomedical Engineering, National Taiwan University of Science and Technology, Taipei 106, Taiwan (e-mail: cweiwang@mail.ntust.edu.tw).

A. Hunter is with the University of Lincoln, Lincoln, LN6 7TS, U.K. (e-mail: ahunter@lincoln.ac.uk).

N. Gravill and S. Matusiewicz are with the Medical Physics Department and Medicine School of the United Lincolnshire Hospital, Lincoln, LN2 5QY, U.K. (e-mail: Neil.gravill@ulh.nhs.uk; Simon.matusiewicz@ulh.nhs.uk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBME.2013.2280132

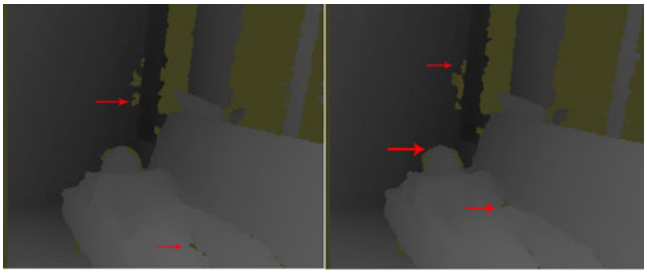


Fig. 1. Unstable signals appear in the depth images using a low-cost-3-D camera (Kinect).

been proposed for monitoring of the cardiac and respiratory motion [17], [18]. However, in both the methods, there are strict positional constraints (the mouth/nose region must be targeted), and the region of interest must not be occluded. These requirements are not easily fulfilled when monitoring humans during sleep.

In this paper, we investigate the use of IR video in detecting apnea events. This has the advantages of using standard, low-cost hardware, and being noninvasive. However, there are several major technical issues: the breathing motion is barely perceptible (due to obscuration by bed clothing and the subtlety of the breathing movements), and being cyclical the movements are prone to self-occlusion. Consequently, the standard motion detection and activity recognition methods do not function well. An interesting alternative exists in modern low-cost 3-D cameras (e.g., Kinect, Xtion Pro, CamBoard nano, or Gesture Camera), which have been suggested for the respiratory motion detection [19], [20]. However, our preliminary analysis indicates that the image signal from Kinect is much more unstable than the IR image (see Fig. 1), making filtering of noise and detection of subtle breathing patterns more difficult. The use of a standard IR camera alone, which may already be available in sleep labs, is also helpful in reducing the complexity of the technology.

The standard motion detection methods include difference of frames (DOF) and optic flow. DOF can be formulated as follows:

$$D(t) = |I(t) - I(t-k)| \tag{1}$$

where $I(t)$ is the intensity image at frame/time t, k is the selected time interval, and $D(t)$ is the frame difference. If $k = 1$, $D(t)$ is the difference of consecutive frames. The difference is thresholded to produce a binary difference map

$$B(x,y,t) = \begin{cases} 1, & \text{if } D(x,y,t) > \alpha \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

where α is a selected threshold.

As breathing movement is so subtle, the value of α in (2) must be set to such a small value (e.g., $\alpha = 1$) to detect differences that noise problems become excessive, particularly as IR sensors suffer from high noise levels [22].

The optic flow tends to fail in regions with a largely homogeneous appearance [23], which is typical of bed clothing. Furthermore, objects that move in a straight line but oscillate forward and backward tend to have low salience [21]. These issues make optic flow unsuitable for our problem domain.

Activity recognition extracts a compact representation of spatiotemporal features and uses this to classify activities. A popular recent approach treats the video sequence as a 3-D space-time volume (of intensities, gradients, optical flow, or other local features). Efros *et al*. [24] perform action recognition by correlating optical flow measurements from low-resolution videos. Bobick and Davis [25] proposed 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; they introduced the motion history image (MHI) and motion energy image (MEI), spatiotemporal models that can be matched to stored models of known actions. 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 dissimilar. In our domain, there is little constraint on the subject's sleeping posture and the shape of breathing varies. MEI and MHI are derived from DOF, and indeed Bobick and Davis [25] suggest that a more robust motion detection mechanism is required in situations where the test subject moves slowly. In addition, the MHI is vulnerable to spatial motion self-occlusion occurring within a temporal window due to overwriting. The extensions to the MHI have been designed to handle self-occlusion [26], [27], but as these are based on DOF techniques, they remain unsuitable for our domain. Gorelick *et al*. [28] also use spatiotemporal volumes for action recognition, modeling human actions as silhouettes of a moving torso and protruding limbs undergoing articulated motion. These space-time shapes may be used to classify actions.

Another popular technique is to track space-time interest points [29] to generate spatial-temporal "words" [30] (using the bag of words representation), and to classify these using probabilistic techniques [31], [32]. However, the lack of distinctive patterns on the bed cover makes this unsuitable for our domain.

This paper presents a new real-time IR video monitoring technique for detecting abnormal breathing activities. It extends our previous work published in short form in [6]. Here, we introduce improved models for both the motion detection and activity recognition that are less sensitive to noise than our earlier approach. Our evaluation demonstrates that these achieve high accuracy in recognizing the abnormal breathing events and the body movement. The organization of the paper is as follows. The proposed algorithm is introduced in Section II. Section III shows the experimental results on 15 video sequences featuring simulated apnea episodes and four clinical clips featuring actual episodes. Section IV concludes the paper.

II. Breathing Detection

This section presents a new IR video monitoring approach for anomalous breathing behavior detection. 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. The technique works with subjects either fully or partially obscured by a bed cover.

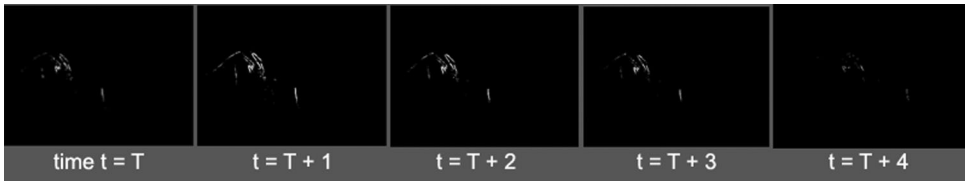


Fig. 2. Sample activity map generated using PLIM, illustrated across a breathing cycle. The subject's breathing is detected around the edge of the subject and folds in the bed clothes. The activity level grows and falls in a cyclical manner through the breathing cycle.

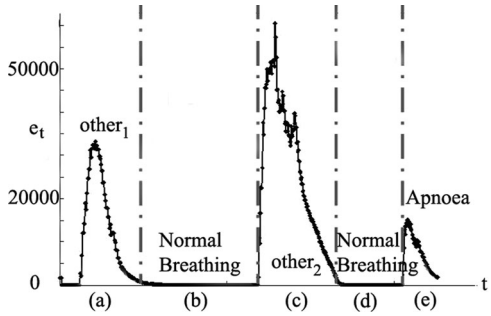


Fig. 3. Activity level, e_t, may be used to detect events, with normal breathing (b,d), apnea events (e), and body movements (a,c) giving rise to significantly different levels of e_t.

The system monitors the degree of motion; while this remains below a threshold, the subject is undergoing *normal breathing*; when it exceeds the threshold, a *motion event* has occurred. The system learns a template to characterize normal breathing motion patterns online, and uses this to classify motion events as *body movements, normal breathing episodes, deep breathing episodes,* or *apnea episodes*.

A. Motion Detection for Breathing Analysis

To address the limitations of DOF methods with respect to the detection of breathing, we have developed a persistent luminous impression model (PLIM). PLIM is derived from the concept of background modeling, which updates a model of the background to discount transitory noise sources while allowing adaptation to long-term changes [33], [34]; however, the PLIM is tuned to detect subtle motion rather than to segment foreground objects. In background differencing, the background model is updated over time to *avoid* accumulated errors. In contrast, the PLIM is designed to *accumulate* errors to enhance the breathing signals and to differentiate between the breathing activity and the body movement. The PLIM incorporates slow adaptation, allowing pose changes to be accommodated while allowing cyclical movements to be detected; see Fig. 2. A simple measure of activity level can be extracted from the PLIM, and used to identify motion events; see Fig. 3.

Given an $M \times N$ image, and frame rate of F frames/s, the PLIM is initialized using the image values

$$P(x, y, 0) = I(x, y, 0). \tag{3}$$

At time t, the PLIM is updated using

$$\Delta(x, y, t) = I(x, y, t) - P(x, y, t-1) \tag{4}$$

$$P(x, y, t) = P(x, y, t-1) + \begin{cases} 1, & \Delta(x, y, t) > 0 \\ 0, & \Delta(x, y, t) = 0 \\ -1, & \Delta(x, y, t) < 0 \end{cases}. \tag{5}$$

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

$$A(x, y, t) = \begin{cases} 1, & \text{if} \quad I(x, y, t) - P(x, y, t) > \alpha \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

where α is the detection threshold, a parameter of the model. During normal sleep, breathing causes subtle movements away from and back toward any arbitrarily chosen starting point. These are observable as a cyclical growth and decline of regions in the PLIM activity map; see Fig. 2. We define the activity level, e_t, as the number of detected pixels in the activity map, $A(x, y, t)$, computed using (6) at time t

$$e_t = \sum_{(x,y)} A(x, y, t). \tag{7}$$

B. State Algorithm for Action Segmentation

Normal breathing is (barely) perceptible in the e_t level; however, motion events manifest as significant perturbations; see Fig. 3. We segment motion events by identifying the start and end time, t_s and t_e, where e_t rises above and subsequently falls below thresholds, described below. The sequences between motion events, where the activity level 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 e_t time series, but this is insufficient to distinguish some movements (e.g., slight head movements) from apnea episodes. A more sophisticated approach using online breathing templates is introduced in the next section.

C. Templates for Normal Breathing Activity

The sleeping subject undergoes protracted periods of normal breathing, where there is only slight movement in particular areas (e.g., around the rib cage, shoulder, throat, mouth, or abdomen, depending on the posture and individual breathing behavior). By identifying *where* this movement occurs, it is possible to distinguish even quite subtle body movements from breathing, and to classify breathing actions as apnea, moderate deep breathing, or normal breathing episodes. We use a template-based method to capture the regions of movement corresponding to normal breathing. However, as the subject tends

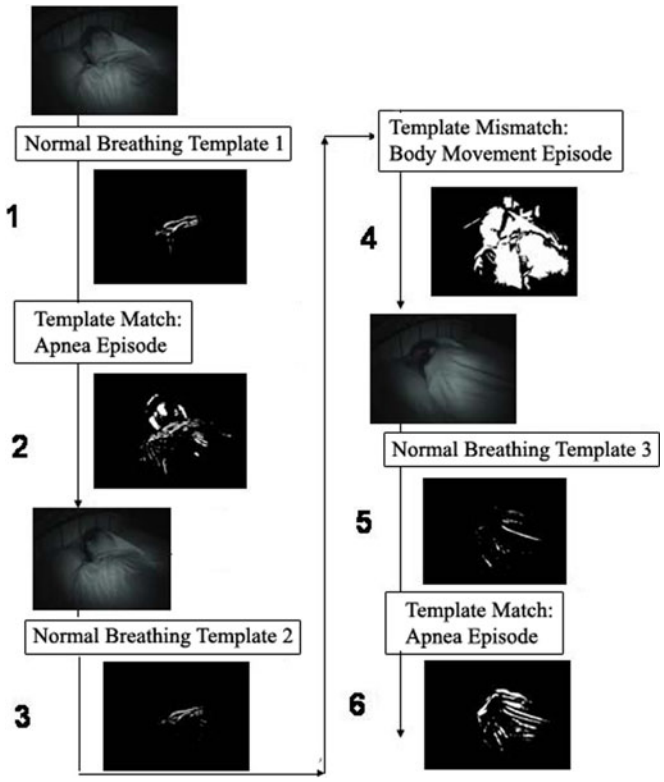


Fig. 4. Sample sequence illustrating adaptive online breathing template construction and activity recognition. (1) Initially, a normal breathing template is constructed, (2) an apnea episode occurs and is identified using the template-matching model, (3) adaption of the existing breathing template continues, (4) a body movement is detected based on the template matching result - template mismatch, (5) a normal breathing template is reconstructed by the proposed adaptive online breathing template construction model, and (6) another apnea episode is detected using the template matching model with the newly constructed template.

to change body pose periodically during sleep, the breathing motion region changes over time too; the template model is therefore reconstructed after body pose changes.

The method is related to MHI, described previously, but given the partial, noisy, and occasional signals, we use a simple binary template augmented with an online construction algorithm to produce a self-adapting normal breathing template based on an individual breathing behavior. A blank template is initially created, and when the state switches to *normal breathing*, the adaptive construction proceeds, until the algorithm switches to the *motion event* state. If the motion event is classified as a breathing event, which implies that the body pose remains the same, the previous template is retained and used when the state changes back to *normal breathing*. On the other hand, if the motion event is a *body movement*, a new template is created to capture breathing activity in the new pose. Fig. 4 illustrates the new adaptive online template construction process and recognition of events in a particular scenario.

The template construction algorithm needs to capture intermittent and limited breathing motion signals while discarding noise. To suppress noise, the signals are included in the template only if they appear at least twice within a certain period, and are retained if they repeat reasonably often. The signals that stop

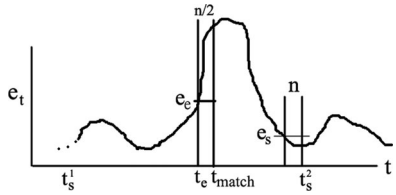


Fig. 5. Switching algorithm state. Assuming the algorithm has switched to normal breathing state at t_s^1, the rise of e_t above the threshold e_e at time t_e, for $n/2$ steps to t_{match} switches state to a motion event and triggers template matching at t_{match}. The fall of e_t below the threshold e_s for n steps triggers the switch to normal breathing at t_s^2.

repeating are usually discarded, except when the number of signals on the template is low, in which case they are retained, as it is then more important to accumulate data than to avoid noise.

The binary template T_t is updated at each time step t, using an auxiliary integer-valued cumulative image T_t^g with values in the range $[0, 255]$. Defining the template quality level, $q_t = \sum_{x,y}(T_t)$ as the number of set pixels in the template, each pixel of T_t is updated as follows:

$$T_t^g = \begin{cases} 255, & \text{if } A_t = 1 \text{ and } T_{t-1}^g > 0 \\ \delta, & \text{if } A_t = 1 \text{ and } T_{t-1}^g = 0 \\ T_{t-1}^g - \epsilon, & \text{if } A_t = 0 \text{ and } q_{t-1} > \lambda \text{ and } T_{t-1}^g > 0 \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

$$T_t(x,y) = \begin{cases} 1, & \text{if } T_t^g(x,y) > \delta \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

where $\delta = 100, \epsilon = 4$, and $\lambda = 0.0012\,\mathrm{WH}$ are empirically-determined parameters of the algorithm, and we have omitted the pixel indices (x, y) for brevity. The template quality threshold λ, is also ultimately used to determine whether the template contains sufficient information to be used for motion event classification.

D. State Transition Rules

The start point t_s of a normal breathing cycle is triggered when the activity level, e_t, drops below the selection threshold λ for $n = 10$ time steps. The end t_e point of the normal breathing cycles is triggered when the activity level rises above the adaptive threshold, e_e, given by

$$e_e = \max(q(t^*)\nu, \lambda) \tag{10}$$

where $\nu = 1.3$ is determined empirically and t^* is a periodic time sample taken on every mth frame ($m = 40$ for 15 frames/s video clips). Hence, the activity level must rise by 130% within a short period $\leq m$ to indicate the end of a normal breathing episode.

The template is compared with the activity map at time t_{match}, after t_e, where the activity level has risen above the adaptive threshold, e_e, for $n/2$ time steps; see Fig. 5. At frame 0, when no template has been built, e_e is temporarily set as $e_e = \lambda\nu^2$, which is defined empirically and soon replaced by the values generated based on the individual breathing pattern. Obtaining e_e, e_s, the end time t_e to terminate the current template construction, the template matching time t_{match}, and the starting time t_s^{new} to

TABLE I
EFFECTIVE VALUES OF MODEL PARAMETERS

α	λ	γ_1	γ_2	ν	n	κ	β
$8 \sim 10$	11.7Δ	.03	.004	1.3	10	.26	4.6
10	$10.5 - 16.5\Delta$	.03	.004	1.3	10	.26	4.6
10	11.7Δ	$.05 - .025$	.004	1.3	10	.26	4.6
10	11.7Δ	.03	$.003 - .005$	1.3	10	.26	4.6
10	11.7Δ	.03	.004	1.3	$8 - 12$	0.26	4.6
10	11.7Δ	.03	.004	1.3	10	$.1 - .3$	$4 - 5$

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

build the next normal breathing template are defined below with a stabilizing factor n for the switching status, which is defined empirically ($n = 10$)

$$t_e = \arg\min_{t \in T_1} t \tag{11}$$

where $T_1 = \{t : e_t \geq e_e \text{ and } t > t_s^{\text{old}}\}$ and t_s^{old} is set to 0 in the beginning

$$t_{\text{match}} = \arg\min_{t \in T_2} |\frac{n}{2} - \sum_{l=t_e}^{t} 1| \tag{12}$$

where $T_2 = \{t : e_t \geq e_e \text{ and } t > t_e\}$

$$t_s^{\text{new}} = \arg\min_{t \in T_3} |n - \sum_{l=t_{\text{match}}}^{t} 1| \tag{13}$$

where $T_3 = \{t : e_t \leq e_s \text{ and } t > t_{\text{match}}\}$.

E. Action Recognition by Template Matching

The motion events are classified using one of the two techniques, based on the breathing template if it is usable (i.e., $q_t \geq \lambda$), or the activity level otherwise. The activity map is compared with the template using a normalized matching score, $s = w_2/w_1$, where w_1 is the proportion of the template intersecting the activity map, and w_2 the proportion of the activity map not intersecting the template

$$w_1 = \frac{\sum(T \bigcap A)}{\sum T} \tag{14}$$

$$w_2 = \frac{\sum(\sim T \bigcap A)}{\sum A} \tag{15}$$

$$s = w_2/w_1. \tag{16}$$

The action is classified using two empirically defined thresholds, $\gamma_1 = 0.03$ and $\gamma_2 = 0.004$

$$\text{action} = \begin{cases} o_1, & \text{if } s \geq \gamma_1 \\ o_2, & \text{if } \gamma_2 \leq s < \gamma_1 \\ o_3, & \text{if } s < \gamma_2 \end{cases} \tag{17}$$

where o_1 is a *body movement* event, o_2 an *apnea* event, and o_3 a *deep breathing* event.

The matching score is a measure of the degree of novelty of the action with respect to the template, normalized for both the activity map and the template size; the time complexity is $\mathcal{O}(p)$ where $p \geq q$.

F. Simple Action Recognition Model

When the breathing template is insufficient to support matching (e.g., due to shallow breathing), $q_t < \lambda$, we instead classify the motion events using the duration d and the activity level value e_{t_m}

$$d = t_e - t_s \tag{18}$$

$$\text{action} = \begin{cases} o_1, & \text{if } e_{t_m} \geq \theta_m \text{ or } d \geq \theta_d \\ o_2, & \text{if } d \geq \theta_d/2 \\ o_3, & \text{otherwise} \end{cases} \tag{19}$$

where thresholds $\theta_m = \kappa$ WH and $\theta_d = \beta F$, where $\kappa = 0.26, \beta = 4.6$ are defined empirically.

G. Adjustable Parameters

The algorithm has a number of adjustable parameters. An initial set of effective parameter values was heuristically determined; then, each parameter was experimentally varied in turn. The operating values for these were determined using three video clips from the simulated datasets, which contains various events including overbreathing and body movement events. Where parameters produced effective performance over a range of values, the most effective value was chosen for each parameter.

Table I illustrates the range of effective parameter values. The reported results utilize the values ($\alpha = 10, \lambda = 0.0012$ WH, $\gamma_1 = 0.03$, $\gamma_2 = 0.004$, $\nu = 1.3$, $n = 10$, $\kappa = 0.26$, $\beta = 4.6$). The algorithm is not very sensitive to the settings of most of these parameters provided they are within the effective range; we discuss the more sensitive parameters below.

The front end motion detector parameter α influences the motion detection results. When α is small (e.g., $\alpha = 6$), more motion is captured, as is noise; when α is too high, all motion is filtered out. As a result, the selection of α is important and can influence the settings of other parameters such as λ. An effective range of $(8 \sim 10)$ was identified, and a large value ($\alpha = 10$) chosen to filter out high IR noise.

Another important and relatively sensitive parameter, λ, determines whether to use the template-matching method or the simple action recognition model. A range of λ values were tested $(0.00105 \sim 0.00165)$, and a low value ($\lambda = 0.00117$) was selected in order to utilize the template as often as possible. Other parameters ($\gamma_1, \gamma_2, \nu, n, \kappa, \beta$) are set using the mean of the effective range.

The same parameter values were used successfully on two separate datasets, which have significantly different environmental settings—-including the illumination, camera viewpoint and angle, camera distance to the subject, and bed and clothing configuration—which indicate that the proposed method is robust.

III. Experiments

We have evaluated the new technique in identification of normal breathing, apnea events, and body movement events. The apnea events are identified as the overbreathing event that occurs at the end of every apnea episode. The body movement is also used as an indicator of waking up by clinicians: if a body movement directly follows an overbreathing event, it supplies additional evidence of an apnea episode. Consequently, the evaluation of the proposed technique is based on detection of the overbreathing events and body movement events.

Two datasets are used in our evaluation: the simulation dataset and the clinical dataset, which use different models of camera, and in different settings with different camera positions with respect to the subject.

The simulation dataset (15 video clips) features actors simulating a wide variety of motion and body movement events. This allows us to evaluate a range of scenario with various occlusion levels, body poses, body movements (i.e., minor head movement, limb movement, body rotation, and slight torso movement), breathing behavior (e.g., shallow versus heavy breathing, mouth breathing, chest breathing, and abdominal breathing) and sequences of linking events (i.e., apnea–body movement and body movement–apnea). Two Sony IR camcorders (DCR-HC-30E) were utilized, with three different shooting angles, at 15 frames/s and a resolution of 320×240. In order to simulate the sleep-lab environment, there was no visible lighting in the filming room and the subjects were partially covered by a sheet. The experimental data were collected from two subjects with three main postures (i.e., lying on the back, lying on one side facing the camera, and lying on the other side with their back to the camera). The data were collected on different days, from multiple camera positions, with the subjects wearing different clothing. The activities, such as normal breathing, obstructive apnea, 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 a human observer who defined all motion events except for deep breathing events, including the frame numbers of the beginning and end of each event. The deep breathing activity is marked as normal breathing.

The clinical evaluation system is installed in the sleep lab of the Lincoln County Hospital. The video system contains three IR cameras: two wall-mounted cameras on each side of the bed targeting on the upper body of the patient from different angles, and one on the ceiling capturing the full body view. In these experiments, the wall-mounted cameras were used. Three symptomatic subjects (one severe and two moderate) and six nonsymptomatic subjects were recruited to spend one night sleeping in the sleep lab for 8 h video recording. For the symptomatic data, five video clips are randomly sampled from the 8 h recordings of the severe OSA patient; four video clips are randomly sampled from the moderate OSA sufferers (two from each). Each clip lasting 15 min, containing 22500 frames. Six video clips are randomly sampled, one from each of the nonsymptomatic subjects. To produce a reference standard, the data were manually marked by the author, who is trained by the medical experts from the Lincoln County Hospital to identify apnea episodes.

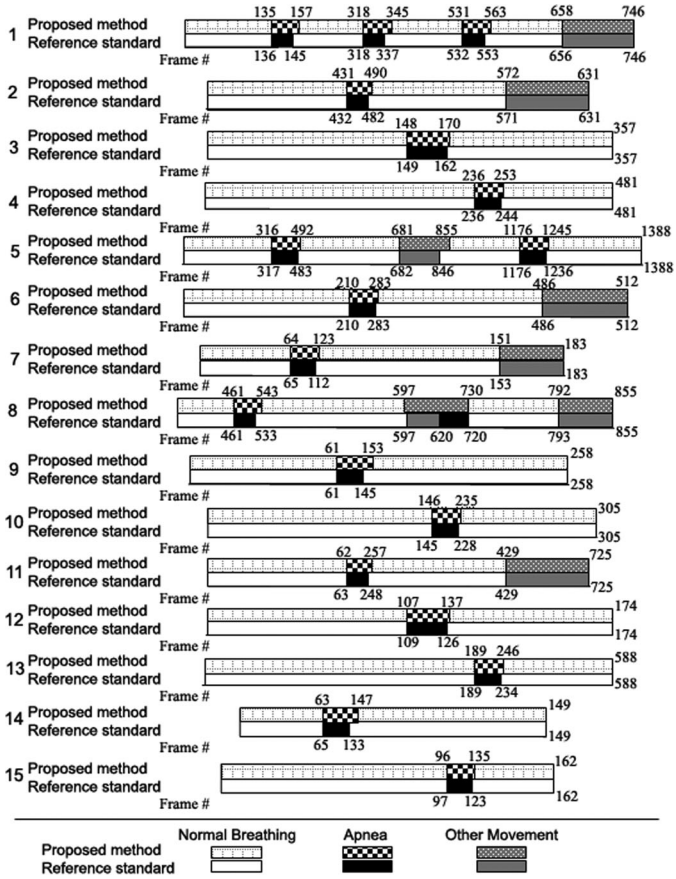


Fig. 6. Experimental results of the simulated data: action classification outcome and reference standard.

The output of the algorithm is a list of apnea and body movement episodes, each with the associated beginning and end frame numbers. These episodes are compared to the reference standard. We define an event to be correctly recognized if the majority of frames ($> 85\%$) covered by the estimated event have the correct labeling. Fig. 6 illustrates the classification process on the simulation dataset.

Fig. 7 shows the quantitative classification results in the form of a confusion matrices [24], [35], for both datasets. The rows represent the reference standard, the columns the algorithm's results. On the simulation dataset, the diagonal average of the confusion matrix is 95.5%, demonstrating that the method achieves high accuracy in recognizing apnea episodes and body movements. We observe that the method misses apnea episodes occurring directly after a body movement episode, as shown in video clip 8, as it segments temporally contiguous episodes as

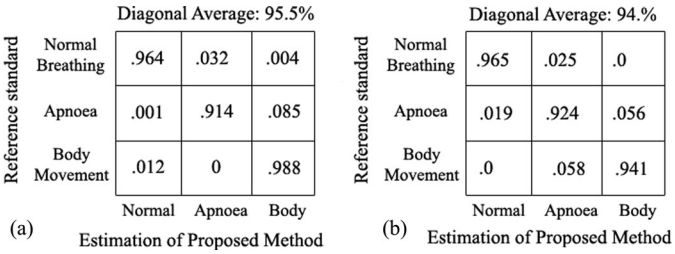


Fig. 7. Confusion matrix of action classification on (a) Simulated data. (b) Clinical data.

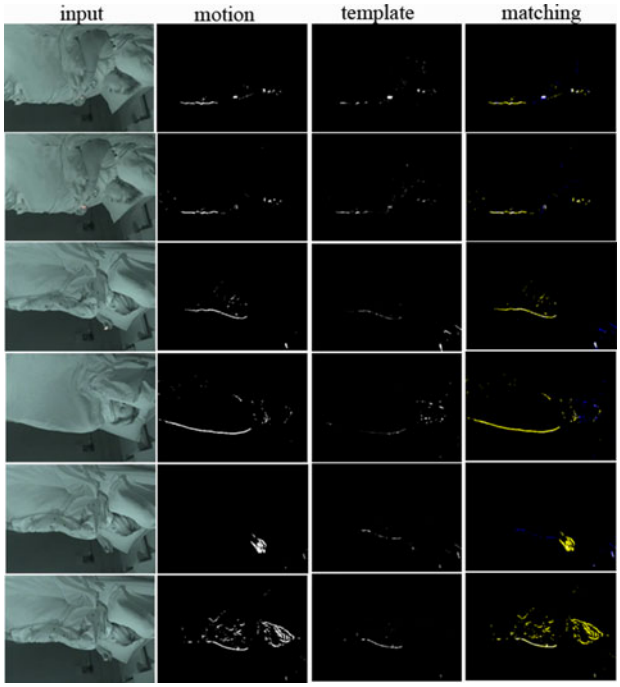


Fig. 8. Template matching examples. Each row contains the raw image I, activity map A, online constructed action template T, matching result at a given time step (yellow: $\sim T \wedge A$; blue: $T \wedge \sim A$; white: $T \wedge A$). The upper three rows are apnea episodes, showing relatively low matching levels; the lower three rows are body movement episodes.

one. In practice, this scenario is implausible as apnea does not occur when the patient is awake, and body movement indicates waking up. On the other hand, if minor body movements happens right after an apnea episode (e.g., in video clip 11, and frequently in the clinical data), the method classifies the entire event as an apnea episode. In such cases, the human observer also defines the entire session as an apnea episode.

For the clinical dataset, the diagonal average is 94%, demonstrating high accuracy in recognizing apnea and body movements episodes for the real clinical data. It is worth noting that the nonsymptomatic patients may experience some apnea episodes (a normal occurrence), and some such episodes were identified.

Some template matching outputs for the clinical dataset are shown in Fig. 8.

a) Classification of symptomatic and nonsymptomatic subjects: The apnea–hypopnea index is generally used for evaluation of the severity of OSA in PSG studies, and is calculated as the average number of apneas (airflow during breath reduced by >90%) plus hypopneas (airflow during breath reduced by between 50% and 90%), per hour of sleep. It is normal for the nonsymptomatic subjects to have a few apnea episodes during sleep, and generally the pulse oximetry traces of the nonsymptomatic subjects also show a small number of oxygen desaturation episodes (ODI<5 h^{-1}). The distinction between the symptomatic subjects and nonsymptomatic subjects is that the number of apnea episodes is considerably higher for the former (the greater the number is, the more severe the OSA patients suffer).

TABLE II
VAHI VALUES

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

OSA severity obtained from the ODI value using pulse oximetry; DB: Deep breathing; Body: Body Movement.

Apart from experiments on classification accuracy of individual event-based recognition, we further tested subject-based classification performance. 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 v (VAHI) which reflects the severity of the subject OSA: see Table II. We treat detected deep breathing episodes as potential hypopnea events, and sum the number of apnea episodes, a, and $0.5\times$the number of deep breathing episodes, d, and divide the total by the ratio of the length of the video clip to an hour, l

$$v = \frac{a + 0.5d}{l}. \tag{20}$$

Table II shows that the VAHI values of the symptomatic video clips are distinct from the nonsymptomatic ones. Using Spearman's rho statistical analysis, the VAHI values are significantly correlated to the OSA diagnosis generated by pulse oximetry device ($p = 0.002$). In the binary classification of the symptomatic and nonsymptomatic video, the VAHI values are highly correlated to the OSA diagnosis ($p < 0.001$) based on Spearman's rho statistical analysis. In one clip, a nonsymptomatic subject had a disturbed sleep and showed a number of body movement episodes and nine apnea episodes (of which five are minor body movements but misclassified as apnea episodes, and the other four are overbreathing episodes; this is normal as noted previously).

IV. Conclusion and Discussion

We have presented a novel approach to detect breathing signals and to recognize abnormal breathing activity from IR video, and have analyzed the method in identification of episodes of OSA . The technique runs in real time, is robust to occlusion by a standard hospital bed cover or sheet, variances in patterns of breathing and subject appearance, and substantial changes of camera view relative to the subject. This preliminary study indicates that it has good performance on both the simulated and clinical data. The algorithm uses a novel persistence luminance model that helps to reinforce subtle breathing movements, an activity level to segment the video, and a novel activity template to classify motion events.

One limitation of the presented method is the number of heuristically determined parameters of the algorithm. For future work, we will investigate automated methods to determine sensitive parameters values and adapt them to individual scenarios, and potentially subjects. We also plan to augment video analysis of human breath activity by adding audio analysis. In addition, we would like to further investigate human sleep behavior by combining the computer vision approach with a (contact-type) EEG technique and exploring methods for using depth data acquired from Kinect. A more extensive analysis with a wider range of cameras and cross validation of clinical diagnosis is required to justify clinical trials and interventions. Furthermore, it would also be interesting to apply the newly developed methods to other breathing monitoring problems.

References

[1] G. J. Gibson, "Obstructive sleep apnoea syndrome: Underestimated and undertreated," *Brit. Med. Bull.*, vol. 72, pp. 49–64, 2004.

[2] 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 the American Thoracic Society," *Chest*, vol. 124, no. 4, pp. 1543–1579, 2003.

[3] J. Hossain and C. Shapiro, "The prevalence, cost implications, and management of sleep disorders: An overview," *Sleep Med. Rev.*, vol. 6, no. 2, pp. 85–99, 2002.

[4] T. Young, L. Evans, L. Finn, and M. Palta, "Estimation of the clinically diagnosed proportion of sleep apnea syndrome in middle aged men and women," *Sleep*, vol. 20, pp. 705–706, 1997.

[5] Visi: Visi-3 Digital Video System. (2013). [Online]. Available:http://www.stowood.co.uk/Brochures/Visi%20Brochure.pdf

[6] C. W. Wang, A. Ahmed, and A. Hunter, "Vision analysis in detecting abnormal breathing activity in application to diagnosis of obstructive sleep apnoea," in *Proc. IEEE Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2006, vol. 1, pp. 4469–4473.

[7] I. Svetlana, H. Y. Mammo, W. A. John, E. H. Michael *et al.*, "A gated deep inspiration breath-hold radiation therapy technique using a linear position transducer," *Appl. Clin. Med. Phys.*, vol. 6, no. 1, pp. 61–70, 2005.

[8] G. B. Moody, R. G. Mark, M. A. Bump, J. S. Weinstein, A. D. Berman, J. E. Mietus, and A. L. Goldberger, "Clinical validation of the ECG-Derived Respiration (EDR) technique," *Comput. Cardiol.*, vol. 13, pp. 507–510, 1986.

[9] K. Storck, M. Karlsson, P. Ask, and D. Loyd, "Heat transfer evaluation of the nasal thermistor technique," *IEEE Trans. Biomed. Eng.*, vol. 43, no. 12, pp. 1187–1191, Dec. 1996.

[10] D. H. Hunsaker and R. H. Riffenburgh, "Snoring significance in patients undergoing home sleep studies," *Otolaryngol.-Head Neck Surg.*, vol. 134, pp. 756–760, 2006.

[11] C.-M. Cheng, Y.-L. Hsu, C.-M. Young, and C.-H. Wu, "Development of a portable device for tele-monitoring of snoring and OSAS symptoms," *Tellemed. e-Heallth*, vol. 14, no. 1, pp. 55–68, 2008.

[12] A. K. Ng, K. Y. Wong, C. H. Tan, and T. S. Koh, "Bispectral analysis of snore signals for obstructive sleep apnea detection," in *Proc. IEEE Eng. Med. Biol. Soc.*, 2007, pp. 6195–6198.

[13] D. P. Randall, "Remote respiratory monitor," in *Proc. IEEE 8th Annu. Symp. Comput.-Based Med. Syst.*, 1995, pp. 204–211.

[14] S. Y. Chekmenev, H. Rara, and A. A. Farag, "Non-contact, wavelet-based measurement of vital signs using thermal imaging," *Int. J. Graph. Vis. Image Process.*, vol. 6, pp. 25–30, 2005.

[15] R. Murthy, I. Pavlids, and P. Tsiamyrtzis, "Touchless monitoring of breathing function," in *Proc. IEEE 26th Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2004, pp. 1196–1199.

[16] Z. Zhu, J. Fei, and I. Pavlidis, "Tracking human breath in infrared imaging," in *Proc. IEEE 5th Symp. Bioinformat. Bioeng.*, 2005, pp. 227–231.

[17] K. Mostov and E. Liptsen, "Medical applications of shortwave FM radar: Remote monitoring of cardiac and respiratory motion," *Med. Phys.*, vol. 37, no. 3, pp. 1332–1338, 2010.

[18] C. Li and J. Lin, "Random body movement cancellation in Doppler radar vital sign detection," *IEEE Trans. Microw. Theory Tech.*, vol. 56, no. 12, pp. 3143–3152, Dec. 2008.

[19] J. Xia and R. A. Siochi, "A real-time respiratory motion monitoring system using KINECT: Proof of concept," *Med. Phys.*, vol. 39, no. 5, pp. 2682–2685, 2012.

[20] M. Alnowami, B. Alnowami, F. Tahavori, M. Copland, and K. Wells, "A quantitative assessment of using Kinect for Xbox360 for respiratory surface motion tracking," in *Proc. SPIE*, 2012, vol. 8316, pp. 1–10.

[21] L. Wixson, "Detecting salient motion by accumulating directionary-consistent flow," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 774–780, Aug. 2000.

[22] Y. Ran, I. Weiss, Q. Zheng, and L. S. Davis, "Pedestrian detection via periodic motion analysis," *Int. J. Comput. Vis.*, vol. 71, no. 2, pp. 143–160, 2007.

[23] A. Lipton, "Local application of optic flow to analyse rigid versus non-rigid motion," in *Proc. Int. Conf. Comput. Vis. Workshop Frame-Rate Vis.*, 1999, pp. 1–9.

[24] A. A. Efros, A. C. Berg, G. Mori, and J. Malik, "Recognizing action at a distance," in *Proc. Int. Conf. Comput. Vis.*, 2003, pp. 726–733.

[25] A. F. Bobick and J. W. Davis, "The recognition of human movement using temporal templates," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 3, pp. 257–267, Mar. 2001.

[26] A. B. Albu and T. Beugeling, "A three-dimensional spatiotemporal template for interative human motion analysis," *J. Multimedia*, vol. 2, no. 4, pp. 45–54, 2007.

[27] M. Valstar, M. Pantic, and I. Patras, "Motion history for facial action detection in video," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, 2004, pp. 635–640.

[28] L. Gorelick, M. Blank, E. Shechtman, M. Irani, and R. Basri, "Actions as space-time shapes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 12, pp. 2247–2253, Dec. 2007.

[29] P. Dollar, V. Rabaud, G. Cottrell, and S. Belongie, "Behavior recognition via sparse spatio-temporal features," in *Proc. Visual Surveill. Perform. Eval. Track. Surveill.*, 2005, pp. 65–72.

[30] J. C. Niebles, H. Wang, and F.-F. Li, "Unsupervised learning of human action categories using spatial-temporal words," *Int. J. Comput. Vis.*, vol. 79, pp. 299–318, 2008.

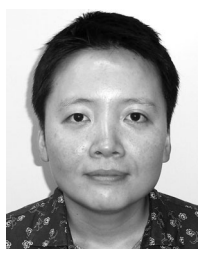
[31] T. Hofmann, "Probabilistic latent semantic analysis," in *Proc. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 1999, pp. 50–57.

[32] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, 2003.

[33] A. Makarov, "Comparison of background extraction based intrusion detection algorithms," in *Proc. Int. Conf. Image Process.*, 1996, vol. 1, pp. 521–524.

[34] N. T. Sibel and S. J. Maybank, "Fusion of multiple tracking algorithms for robust people tracking," in *Proc. Eur. Conf. Comput. Vis.*, 2002, vol. 4, pp. 373–387.

[35] R. Kohavi and F. Provost, "Special issue on applications of machine learning and the knowledge discovery process," *Mach. Learn.*, vol. 30, pp. 271–274, 1998.

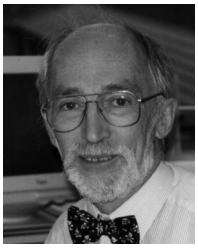


Ching-Wei Wang (M'08) received the M.Sc. degree (with Distinction) in computer science from the University of Glasgow, Glasgow, U.K., and the Ph.D. degree in computer science from the University of Lincoln, Lincoln, U.K.

She has years of working experiences in computer vision and artificial intelligence, and is currently an Associate Professor in the Graduate Institute of Biomedical Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan.

Andrew Hunter (M'09) received the B.Sc. and Ph.D. degrees form the University of Bath, Bath, U.K., in 1985 and 1989, respectively.

He is the Dean of Research, and holds a Chair in computer vision at the University of Lincoln, Lincoln, U.K., where he founded the Vision and Robotics Research Centre in 2004. He has published more than 70 academic papers in video surveillance, medical image processing, neural networks, and genetic algorithms. His research interests include FPGA-based neural vision systems, human monitoring for surveillance and assistive care, and retinal image processing.



Neil Gravill is a Consultant Clinical Scientist in the Medical Physics Department, United Lincolnshire Hospitals NHS Trust, U.K. He is the Head of the Clinical Measurement Service which provides a range of specialist diagnostic tests. His research interests include respiratory sleep assessment, measurement of the upper gastro intestinal physiology, and hearing assessment of newborns. He publishes occasional papers.

Simon Matusiewicz received the M.B.Ch.B. degree from the University of Leeds, Leeds, U.K.

He is a Consultant Physician in the United Lincolnshire Hospitals NHS Trust U.K., with an interest in respiratory medicine. He set up a clinical service for the diagnosis and management of obstructive sleep apnea in Lincoln, U.K., in 1997.

Dr. is a fellow of the Royal college of Physicians, London, U.K.