SLEEP MONITORING VIA DEPTH VIDEO COMPRESSION & ANALYSIS

Cheng Yang[†], *Gene Cheung*[‡], *Kevin Chan*[§], *Vladimir Stankovic*[†]

[†]Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow, UK
[‡]National Institute of Informatics, Tokyo, Japan
[§]School of Medicine, University of Western Sydney,
Camden and Campbelltown Hospitals, Sydney, Australia

ABSTRACT

Quality of sleep greatly affects a person's physiological well-being. Traditional sleep monitoring systems are expensive in cost and intrusive enough that they disturb natural sleep of clinical patients. In this paper, we propose an inexpensive non-intrusive sleep monitoring system using recorded depth video only. In particular, we propose a two-part solution composed of depth video compression and analysis. For acquisition and compression, we first propose an alternating-frame video recording scheme, so that different 8 of the 11 bits in MS Kinect captured depth images are extracted at different instants for efficient encoding using H.264 video codec. At decoder, the uncoded 3 bits in each frame can be recovered accurately via a block-based search procedure. For analysis, we estimate parameters of our proposed dual-ellipse model in each depth image. Sleep events are then detected via a support vector machine trained on statistics of estimated ellipse model parameters over time. Experimental results show first that our depth video compression scheme outperforms a competing scheme that records only the eight most significant bits in PSNR in mid- to high-bitrate regions. Further, we show also that our monitoring can detect critical sleep events such as hypopnoea using our trained SVM with very high success rate.

***Index Terms*—** Sleep monitoring, depth video compression, depth image processing

1. INTRODUCTION

Everyone sleeps. It is well documented [1] that a sleep-deprived person carries a number of health-related risks, including increase in body weight, increased risk of diabetes and heart deceases, increased risk of psychiatric illness such as depression, etc. Further, it is not simply the *quantity* or duration of sleep that affects a person's physiological well-being, but also the *quality* of sleep. In particular, sleep-disordered breathing is common in the general population [2], and repeated episodes of *apnoea* (temporary suspension of external breathing) and *hypopnoea* (overly shallow breathing or low respiratory rate) can significantly disturb a person's sleep and interfere with his/her daily activities. It is thus paramount to closely monitor a patient's sleep, so that potential sleep problems can be quickly and correctly diagnosed and the right treatment prescribed. We address the sleep monitoring problem in this paper.

Existing sleep monitoring systems fall into two general categories. In the first category are vibration-sensing wristbands like Fitbit[1] and Jawbone UP[2]. While these are minimally intrusive devices, they are mostly designed to record sleep time, *i.e.*, the *quantity* of sleep rather than the *quality* of sleep. In the second category, there are full multi-sensing monitoring devices such as Philips Alice PDx[3]. While accurate in measuring various vital signs such as oxygen intake, airflow, etc, these are expensive and intrusive systems with multiple body straps and tubes. (PDx requires up to 10 minutes to set up.) The numerous sensors attached to various body parts tend to disturb natural sleep of a patient during monitoring (who had enough difficulties sleeping).

In this paper, we propose an inexpensive and non-intrusive sleep monitoring system based solely on depth video compression and analysis. Not relying on the lighting condition of a dark sleeping room, we use a MS Kinect camera that actively projects its own structured infrared beams to form captured depth images of the patient. Unlike vibration-sensing wristbands, our system can quantify the quality of one's sleep by detecting episodes of apnoea or hypopnoea during the night. Yet unlike full monitoring devices, our system is entirely non-contact, and thus is completely non-intrusive to the patient's sleep.

In particular, we propose a two-part solution composed of depth video compression and analysis. For compression, we first propose an alternating-frame video recording scheme, so that different 8 of the 11 bits in captured depth images are extracted at different instants for efficient encoding using H.264 video codec [3]. At decoder, the uncoded 3 bits in each frame can be recovered accurately via a block-based search procedure. For analysis, we first estimate parameters of our proposed dual-ellipse model in each depth image. Apnoea or hypopnoea are then detected via a *support vector machine* (SVM) [4] trained on statistics of estimated ellipse model parameters over time. Experimental results show first that our depth video compression scheme outperforms a competing scheme that records only the eight most significant bits in PSNR at mid- to high-bitrate regions. Further, we show also that our monitoring system can detect hypopnoea or apnoea using our trained SVM with very high success rate.

The outline of the paper is as follows. We first discuss related works in Section 2. We then overview our sleep monitoring system in Section 3. We discuss the two parts of our system, depth video compression and sleep event detection, in Section 4 and 5, respectively. Finally, we present experimental results and conclusion in Section 6 and 7.

2. RELATED WORK

Depth image / video compression is a popular research topic due to the now standard texture-plus-depth format (coding of color and

[1] http://www.fitbit.com/

[2] https//jawbone.com/up/

[3] http://alicepdx.respironics.eu

depth images from the same captured viewpoints) [5] for free viewpoint video [6]. A large portion of these works [7, 8] proposed new coding tools like graph transforms exploiting the *piecewise smooth* (PWS) signal characteristic of depth maps; *i.e.* there exist sharp object boundaries and slowly varying surfaces away from the boundaries. In our work, for real-time implementation at the encoder we use H.264 [3] to encode different 8 of 11 bits captured by a Kinect camera for different frames. At the decoder, the PWS characteristic is exploited to recover the three uncoded bits. To the best of our knowledge, we are the first to propose an efficient H.264 implementation of 8-bit depth video compression given Kinect-captured 11-bit images and show demonstrable superior coding performance.

General pose detection from recorded images is a long-standing problem in computer vision [9], and recently algorithms are tailored to depth images [10] as well. Our sleep monitoring application is unique in that only depth (no color) images are available. Further, because the patient sleeps on a bed with similar physical distance to the depth sensor, the lack of clear separation between foreground objects and background means generic techniques like [10] do not work well. However, in our application the human pose (sleeping) is known *a priori*, and thus we propose a dual-ellipse model to detect sleep events via an SVM-based analysis.

There exists two other works in the literature [11, 12] that also used captured depth video for sleep monitoring. [11] claimed that a Time of Flight (ToF) camera was used to detect chest and abdomen movements for apnoea detection, but there is no description of what ToF camera was used and how chest and abdomen movements were deduced from collected depth measurements; for example, how to distinguish between depth measurements of the bed and of the patient, and how to detect chest / abdomen movements when the patient is sleeping upright versus sideway. There is also no performance analysis of the proposal against ground truth data. This renders a direct comparison with [11] impossible. Further, there is no depth video recording component, meaning a doctor cannot visually inspect and verify the patient's sleep afterwards.

In contrast, [12] described in detail a sleep monitoring system with a single Kinect camera, where chest movements are detected by tracking over time the closest depth measurement of the patient to a virtual camera directly above the patient. We differ from [12] in two respects. First, we propose an *end-to-end system* that includes an efficient depth video coding scheme; [12] did not propose any coding algorithm. Second, unlike [12] we propose a more accurate dual-ellipse model, so that individual chest and abdominal movements can be tracked, as typically recommended in standard sleep medicine [1], even if the patient is sleeping sideway.

3. SYSTEM OVERVIEW

We first overview our sleep monitoring system, which we have set up at a sleep clinic to capture depth videos of patients with suspected sleep problems. See Fig. 1 for an illustration. The system is composed of a first-generation MS Kinect depth capturing camera and a Lenovo X220 laptop. The camera is set up at a higher elevation above and away from the head of the patient lying down. This camera location gives an un-obstructed view of the patient's upper body (torso), which is important for our analysis. The Kinect camera captures depth image of resolution 640×480 at 30fps at 11-bit pixel precision. Because Kinect relies on active projection of structured lights to estimate depth, it can operate in the dark, which is ideal for monitoring sleep in a dark room.

The first part of our sleep monitoring system is the real-time capturing and compression of depth video. Recorded depth video can subsequently be used by doctors and patients for visual inspection of detected sleep events—a crucial double-checking mechanism for medical professionals and a valuable educational tool for patients. The recorded video can also be used to detect other sleep-related events beyond apnoea, such as irregular leg movements, frequent turning / tossing, etc [1]. In the second part, using the recorded depth video we track the breathing cycles of the patient by deriving parameters of our proposed dual-ellipse model. We discuss these components in order next.

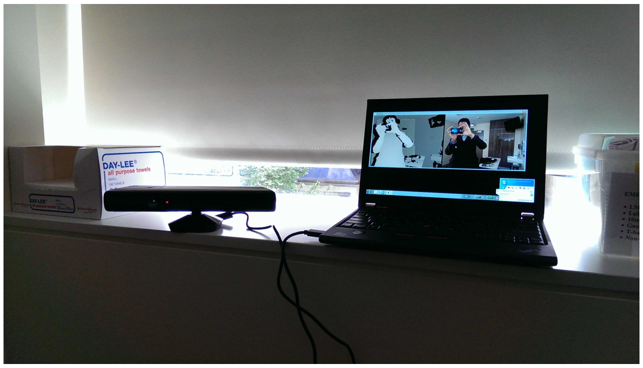
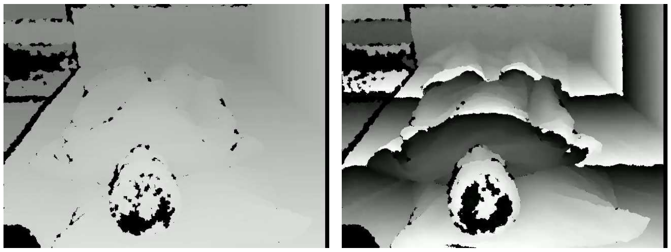


Fig. 1: Depth video capturing system at a sleep clinic: a MS Kinect camera is attached to a laptop computer. Example depth and infrared captured images are shown on the screen.

4. DEPTH VIDEO CODING

Each depth image captured by a first-generation MS Kinect sensor contains 11-bit precision pixels. Baseline profile for H.264—the most prevalent and optimized profile in H.264—supports only 8-bit precision, however[4]. Thus, we propose an alternating frame coding scheme to extract different 8 of 11 available bits in each captured pixel of different frames for encoding. At the decoder, we recover the uncoded 3 bits using our proposed recovery scheme. The reasons we can recover the uncoded 3 bits with high accuracy are: i) depth maps are known to be PWS, and ii) in a typical sleep video, only slow motion exists across frames. We discuss the encoding and decoding procedures next.

4.1. Encoder Selection of 8 Coding Bits



(a) MSB frame

(b) LSB frame

Fig. 2: Examples of MSB and LSB frames.

The encoder selects different 8 bits for each depth frame $\mathbf{Z}_t$ of time instant t for encoding as follows. If $t \bmod M = 0$, then the 8

[4] Only High 4:4:4 Profile supports 11 to 14 bits precision.

most significant bits (MSB) of 11 captured bits are selected for encoding. Otherwise, the 8 *least significant bits* (LSB) of 11 available bits are selected. M is the *reference picture selection* (RPS) parameter used during H.264 video encoding [3]; *i.e.*, a P-frame $\mathbf{Z}_t$ can choose any one of previous frames $\mathbf{Z}_{t-1}, \ldots, \mathbf{Z}_{t-M}$ as predictor for differential coding. MSB frames and LSB frames are very different; missing details in LSBs, MSB frames are very smooth, while LSB frames suffer from overflow due to missing MSBs. See Fig. 2 for an illustration. However, our proposed encoding scheme ensures that each MSB or LSB frame $\mathbf{Z}_t$ can find a similar previous frame $\mathbf{Z}_{t-i}$ as predictor for differential coding, thus achieving good coding efficiency (shown in Section 6).

4.2. Decoder Recovery of Full 11 Bits

At the decoder, we recover the uncoded 3 MSBs in an LSB frame as follows. We first segment an LSB frame into *smooth regions*, *i.e.* spatial regions where adjacent pixels do not differ by more than a pre-defined threshold δ. Pixels in the same smooth region will share the same to-be-recovered 3 MSBs.

Next, we identify potential *overflow* pixels due to encoding of LSBs only—pixels that were similar to adjacent pixels before removal of 3 MSBs. Specifically, given smooth region boundary pixel location $\mathbf{p}$ in depth map $\mathbf{Z}_t$, we check if adding one significant bit 2^8 would bring it closer to within δ of one of its neighbors, *i.e.*:

$$\min_{\mathbf{q} \in \mathcal{N}_\mathbf{p}} \left| \mathbf{Z}_t(\mathbf{p}) + 2^8 - \mathbf{Z}_t(\mathbf{q}) \right| \leq \delta \tag{1}$$

where $\mathcal{N}_\mathbf{p}$ is the set of adjacent pixels to $\mathbf{p}$. If this is the case, then $\mathbf{p}$ is a potential overflow pixel. To check if $\mathbf{p}$ is an overflow pixel (or simply an object boundary), we perform *motion estimation* (ME) using the most recent MSB frame $\mathbf{Z}_\tau$. Specifically, given a $R \times R$ block $B_\mathbf{p}$ with center at $\mathbf{p}$ of the current frame $\mathbf{Z}_t$ as target, we compute:

$$\min_{\mathbf{v}} \left| \mathbf{Z}_\tau(B_{\mathbf{p}+\mathbf{v}}) \bmod 2^5 - \left\lfloor \frac{\mathbf{Z}_t(B_\mathbf{p})}{2^3} \right\rfloor \right| + \mu|\mathbf{v}| \tag{2}$$

where the 5 LSBs in block $B_{\mathbf{p}+\mathbf{v}}$ of $\mathbf{Z}_\tau$ and the 5 MSBs in block $B_\mathbf{p}$ of $\mathbf{Z}_t$ are compared—only 5 bits are common between MSB and LSB frames. Note that we add the magnitude of the *motion vector* (MV) $\mathbf{v}$ as a regularization term. μ is a parameter that trades off between the block differential and the regularization term. The regularization term is important, because for PWS images, there can be multiple vectors $\mathbf{v}$ with very small block differences. It is reasonable because the majority of the frames in sleep video have little motion.

Given the best MV $\mathbf{v}_\mathbf{p}$ computed in (2), we then check if $B_{\mathbf{p}+\mathbf{v}_\mathbf{p}}$ is smooth in $\mathbf{Z}_\tau$. If so, then pixel $\mathbf{p}$ in $\mathbf{Z}_t$ is indeed an overflow bit, and we merge the smooth region of $\mathbf{p}$ with the corresponding neighboring smooth region; *i.e.*, the merged smooth region will share the same MSBs. If not, then this is actually an object boundary, and we copy the 3 MSBs in $B_{\mathbf{p}+\mathbf{v}}$ of $\mathbf{Z}_\tau$ to *all* pixels in the smooth region containing $\mathbf{p}$.

5. SLEEP EVENT DETECTION

In this section we discuss how apnoea or hypopnoea can be detected using the recorded depth video with full 11 bits recovered. The event detection part is divided into three steps. In the first step, each captured depth pixel from the captured camera view is mapped to a virtual camera view located horizontally from the top of the patient's head (*head-on view*). See Fig. 3 for an illustration. (To reduce the computational time, the depth observations of the background are filtered out by setting manually in the first frame a region of interest—a rectangle that covers the body area.) Each pixel with coordinate (u, v, d) in the virtual view is then classified into two different cross sections of the patient's torso based on the depth value d. In the second step, for each cross section, the optimal ellipse that is closest to the set of observations (u, v)'s is chosen. In the third step, we detect apnoea or hypopnoea by correlating them with the changes in derived ellipse parameters over time. We describe these three steps in order next.

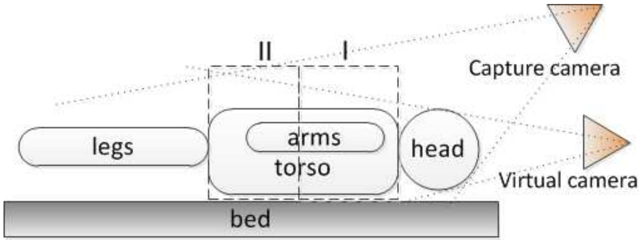


Fig. 3: Side view of sleep patient. Torso is divided into two cross sections, each modeled by an ellipse.

5.1. Perspective Change

In the first step, we first map each observed depth value z at coordinate (i, j) of the camera view to a new triple (u, v, d) in the head-on view. We know that the camera coordinate $\mathbf{x}$ is related to a 3D world coordinate $\mathbf{X}$ as follows [13]:

$$\mathbf{x} = \mathbf{KR}\left[\mathbf{I} \,|\, -\mathbf{C}\right]\mathbf{X} \tag{3}$$

where $\mathbf{K}$, $\mathbf{R}$ and $\mathbf{C}$ are the intrinsic camera matrix, rotation matrix, and translation matrix respectively. These parameters can be computed using standard camera calibration procedures[5] [14]. We thus back-project a captured pixel (i, j, z) to a 3D world coordinate $\mathbf{X}$ using an inverse of (3), then re-project from $\mathbf{X}$ to a virtual camera coordinate (u, v, d).

Given the computed coordinates (u, v, d)'s in the head-on view, we classify observations into two cross sections that correspond to the patient's chest and abdomen. It is recommended in standard sleep medicine [1] to track chest and abdominal movements for detection of apnoea; in *central apnoea*, there is a lack of respiratory effort and hence a corresponding lack of chest and abdominal movements, while in *obstructive apnoea* there can be very slight movements in chest and abdomen but in opposite phase. Though we do not distinguish between central and obstructive apnoea in this paper (central apnoea takes place only 0.4% of the time), we nonetheless follow the medical recommendation and track chest and abdominal movements separately.

5.2. Ellipse Parameter Estimation

An ellipse in 2D space—one whose major and minor axes coincide with the u- and v-axes—can be described as:

$$\left(\frac{u}{a}\right)^2 + \left(\frac{v}{b}\right)^2 = r^2 \tag{4}$$

with parameterization $\mathbf{p}(\phi) = r(\alpha \cos\phi, \beta \sin\phi)$. a and b are called the *major* and *minor radius*, respectively. For simplicity, we

[5] Camera calibration software can be downloaded here: http://www.vision.caltech.edu/bouguetj/calib_doc/htmls/ref.html

will assume the center of the ellipse is at origin $(0,0)$. Thus the parameter θ_i of an ellipse i can be characterized by $\theta_i = (a_i, b_i, r)$, where r is determined based on the waist measurement of the patient.

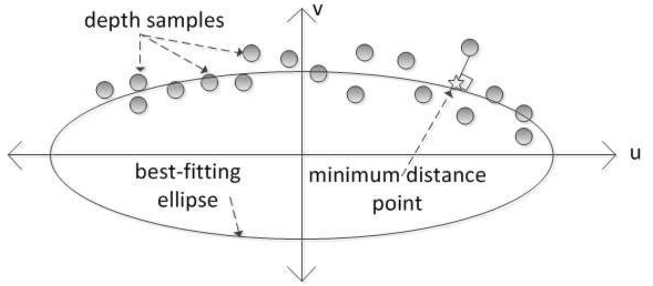


Fig. 4: Best-fitting ellipse from multiple depth observations of the cross section. The closest ellipse point to each observation is perpendicular to the tangent of ellipse at that point.

5.2.1. *Problem Formulation*

Let $\mathbf{o}_i = \{\mathbf{o}_{i,1}, \ldots, \mathbf{o}_{i,N}\}$ be the set of N observations of ellipse i, where $\mathbf{o}_{i,n}$ is a triple (u, v, d) denoting the point's location (u, v) in 2D space and corresponding depth value d as viewed from the head-on view.

To estimate parameters θ_i of an ellipse i given observations $\mathbf{o}_i$, we formulate a *maximum likelihood* (ML) problem—instead of finding θ_i that maximizes $Pr(\theta_i \mid \mathbf{o}_i)$, we solve the following:

$$\max_{\theta_i} Pr(\mathbf{o}_i \mid \theta_i) \tag{5}$$

We assume a jointly Gaussian noise model for our observed depth data, as done in [15], so that if true body part i has ellipse parameter θ_i, then $Pr(\mathbf{o}_i|\theta_i)$ is:

$$Pr(\mathbf{o}_i|\theta_i) = \frac{1}{(2\pi)\,\frac{N}{2}\,|\mathbf{H}|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}\mathbf{d}_{\theta_i}(\mathbf{o}_i)^T\mathbf{H}^{-1}\mathbf{d}_{\theta_i}(\mathbf{o}_i)\right] \tag{6}$$

where $\mathbf{H}$ is the covariance matrix, and $\mathbf{d}_{\theta_i}(\mathbf{o}_i)$ is a vector composed of minimum distances between each observation $\mathbf{o}_{i,n}$ and an ellipse of parameter θ_i.

Instead of solving the maximization problem in (5), we solve the following equivalent minimization problem:

$$\begin{aligned} &\min_{\theta_i} \quad -\log Pr(\mathbf{o}_i|\theta_i) \\ &\min_{\theta_i} \quad \mathbf{d}_{\theta_i}(\mathbf{o}_i)^T\mathbf{H}^{-1}\mathbf{d}_{\theta_i}(\mathbf{o}_i) \end{aligned} \tag{7}$$

If we now assume each observation is independent from the others, then variance $\mathbf{H}$ is a diagonal matrix, and (7) can be simplified to:

$$\min_{\theta_i} \sum_{n=1}^{N} h_n^{-1} \left(d_{\theta_i}(\mathbf{o}_{i,n})\right)^2 \tag{8}$$

In practice, h_n^{-1}, the inverse variance of observation n, is assigned an appropriate value depending on how occluded the observation $\mathbf{o}_{i,n}$ is likely to be. For example, $\mathbf{o}_{i,n}$ at the top of the chest away from limbs will have a small variance h_n.

5.2.2. *Optimization Algorithm*

Before we can solve (8), we first need to properly define how minimum distance $d_{\theta_i}(\mathbf{o}_{i,n})$ between observation $\mathbf{o}_{i,n}$ and ellipse with parameter θ_i can be computed. To find the exact point (s,t) on the ellipse with parameter θ_i that is closest to observation (u,v) involves solving a *quartic* (fourth degree) equation with four possible solution candidates, and the minimum distance point is then chosen as the final solution [16]. This is clearly too computation-expensive for us when the number of observations N is large.

Instead, we make the observation[6] that a necessary condition for $\mathbf{p}(\phi)$ to be closest to $\mathbf{o}_{i,n} = (u,v)$ is that $\mathbf{p}(\phi) - \mathbf{o}_{i,n}$ must be perpendicular to the tangent at $\mathbf{p}(\phi)$; *i.e.*,

$$\begin{aligned} (\mathbf{o}_{i,n} - \mathbf{p}(\phi)) \cdot \mathbf{p}'(\phi) &= 0 \\ \left(\alpha^2 - \beta^2\right) r\cos\phi\sin\phi - u\,\alpha\sin\phi + v\,\beta\cos\phi &= 0 \end{aligned}$$

See Fig. 4 for an illustration. We can then solve the above equation using a Newton method, with an initial guess $\phi^0 = \tan^{-1}\left(\frac{\alpha\, v}{\beta\, u}\right)$.

Using the above method, we can efficiently compute $d_{\theta_i}(\mathbf{o}_{i,n})$ for each observation $\mathbf{o}_{i,n}$, and hence the objective (8). To find the optimal ellipse parameter θ_i, we perform a local search where each of a and b are perturbed by a small amount $\pm\gamma$ to see if the objective (8) has decreased. If so, we continue the perturbation in the same direction until the objective can no longer be decreased further.

5.3. Sleep Event Detection

We discuss next how hypopnoea is detected based on the estimation of the parameters of a best-fitting ellipse for each cross section of a patient's torso. Hypopnoea is a condition where the patient experiences overly shallow breathing or abnormally low respiratory rate. Low respiratory rate means the changes of ellipse parameters are much slower than normal.

First, we split the sequence into 10-second windows. This size was chosen as a good tradeoff between complexity and performance, since the respiratory rate of both normal breathing and overly-shallow-breathing-hypopnoea are approximatively 3 breaths per 10 seconds. For each 10-second window, we compute the standard deviations of all ellipse parameters. These parameters are used to train a nonlinear SVM with a Gaussian *Radial Basis Function* (RBF) kernel. Accordingly, during the test phase, we compute the standard deviations of all ellipse parameters for each testing 10-second window and test for the episodes of hypopnoea and normal breathing using the trained RBF SVM.

6. EXPERIMENTATION

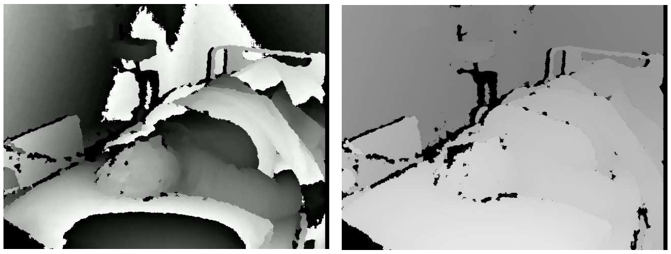
6.1. Experimental Setup

We captured depth videos of 6 patients, diagnosed with *obstructive sleep apnoea* (OSA) [1], at a sleep clinic during October and November 2013. Besides our depth video capturing, each patient was also connected to a professional-grade sleep monitoring system (expensive and intrusive) that measures various vital signs. This provided ground truth data for training our SVM and validation of our results.

[6] http://www2.imperial.ac.uk/˜rn/distance2ellipse.pdf

6.2. Experimental Results

We first validate our proposed block-based search procedure to recover the 3 uncoded MSBs in an LSB frame. Fig. 5 shows an example of the decoded LSB frame and the recovered LSB frame. First, we can see in Fig. 5(a) that due to overflow, there are discontinuities even within the same physical object. We see in the recovered LSB frame in Fig. 5(b) that the overflow problem is correctly resolved, resulting in a much smoother and natural looking depth image.



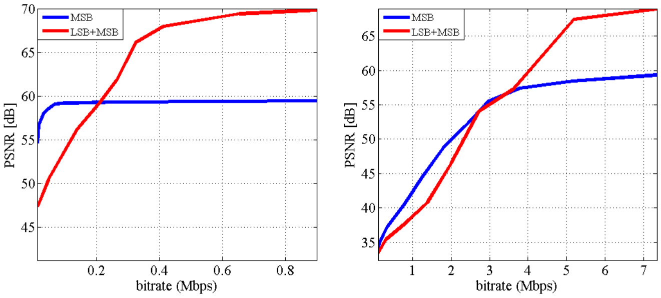
(a) original LSB frame (b) recovered LSB frame

Fig. 5: Examples of decoded LSB frame and recovered LSB frame.

Next we compare the compression performance of our LSB-MSB coding scheme to the scheme that compresses only the 8 MSBs of each depth frame using the same H.264/AVC codec. As a performance metric we use peak signal-to-noise ratio (PSNR), calculated as:

$$\mathrm{PSNR} = 10\log_{10}\frac{(2^{11}-1)^2\cdot M\cdot N}{\sum_{i=1}^{M}\sum_{j=1}^{N}[\mathrm{X}(i,j)-\mathrm{Y}(i,j)]^2}$$

where X and Y are two $M\times N$ 11-bit depth images. Uncompressed 11-bit depth images were used as ground truth, and for the 8-MSB coding scheme, three zero bits were appended to the decompressed 8-bit values.



(a) Video sequence 1. (b) Video sequence 2.

Fig. 6: Compression performance for two video sequences.

Fig. 6 shows the coding performance as PSNR averaged over all frames of the two coding schemes for two video sequences. The second video sequence has many background objects, hence it is more difficult to compress. The results indicate that our LSB-MSB coding scheme outperforms 8-MSB coding scheme for up to 10dB, at mid- to high-bitrate regions that are of interest for event detection.

Recall from Section V that there are two ellipses of interest each described with two radii: the chest-ellipse, with major and minor radius a_1 and b_1, and the abdomen-ellipse, with radii a_2 and b_2. All four radii can be used for classification. Hence, we test 6 different classification methods that are based on different combinations of extracted features: (a_1, a_2), (a_1, b_1), (a_2, b_2), (a_1, b_2), (a_2, b_1), and (b_1, b_2). For each method, during the training phase, we train 2D RBF SVMs with a scaling factor of 1. The training data for (a_i, b_j), $i = 1, 2, j = 1, 2$, RBF SVM is given by:

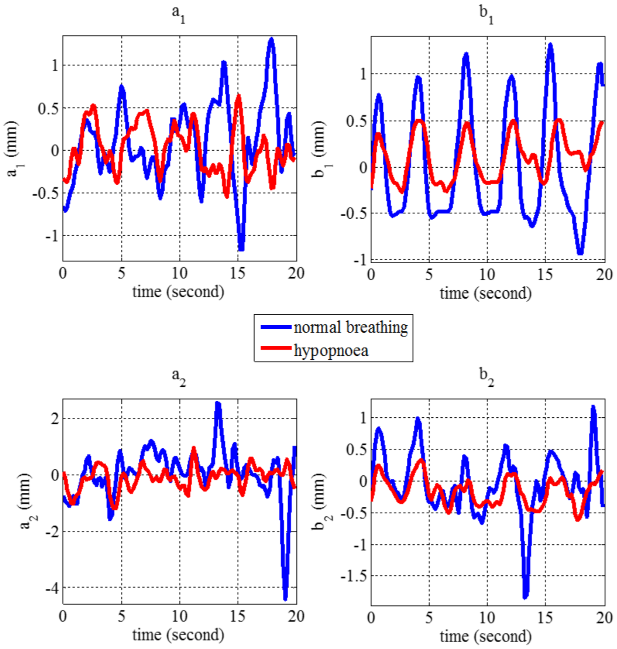


Fig. 7: 20-sec samples of normal breathing and hypopnoea of one single subject in upright sleeping position showing variations of the four ellipse parameters across time. a_1 and b_1 are the major and minor radiuses of the chest-ellipse, respectively; a_2 and b_2 are the major and minor radiuses of the abdomen-ellipse, respectively.

$$\{(\sigma_{a_{il}}, \sigma_{b_{jl}}), \psi_l\}, l = 1, ..., L,$$
$$(\sigma_{a_{il}}, \sigma_{b_{jl}}) \in \mathbb{R}^2,$$
$$\psi_l \in \{\text{hypopnoea, normal breathing}\},$$

where L is the total number of available training samples, $\sigma_{a_{il}}$ and $\sigma_{b_{jl}}$ are the standard deviations (STD) of a_i and b_j in a 10-second window, respectively.

Additionally, we test the performance of a more complex classification scheme that uses all four features (a_1, b_1, a_2, b_2) and 4D RBF SVM with the scaling factor of 1, and the following training data:

$$\{(\sigma_{a_{1l}}, \sigma_{b_{1l}}, \sigma_{a_{2l}}, \sigma_{b_{2l}}), \psi_l\}, l = 1, ..., L,$$
$$(\sigma_{a_{1l}}, \sigma_{b_{1l}}, \sigma_{a_{2l}}, \sigma_{b_{2l}}) \in \mathbb{R}^4,$$
$$\psi_l \in \{\text{hypopnoea, normal breathing}\}.$$

Fig. 7 shows variations of the four ellipse parameters of one single subject in upright sleeping position over a period of 20 sec. For each parameter an example of normal breathing and an episode of hypopnoea are shown. Given very small movement is detected during breathing, high precision is required.

To evaluate the performance of these RBF SVMs, we use precision (P), recall (R), F-Measure (F$_M$), and Accuracy (ACC) which are defined as:

$$\mathrm{P} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}},$$

$$\mathrm{R} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}},$$

Table 1: Evaluation of the trained RBF SVMs on classification of hypopnoea and normal breathing.

Metric	(a_1, b_1)	(a_1, a_2)	(a_1, b_2)	(a_2, b_1)	(b_1, b_2)	(a_2, b_2)	(a_1, b_1, a_2, b_2)
TP	25	25	25	24	25	25	25
FP	0	2	2	0	0	3	0
TN	25	23	23	25	25	22	25
FN	0	0	0	1	0	0	0
Precision	1.000	0.926	0.926	1.000	1.000	0.893	1.000
Recall	1.000	1.000	1.000	0.960	1.000	1.000	1.000
F-Measure	1.000	0.962	0.962	0.979	1.000	0.943	1.000
Accuracy	1.000	0.960	0.960	0.980	1.000	0.940	1.000

$$\mathrm{F}_M = 2 \cdot \frac{\mathrm{P} \cdot \mathrm{R}}{\mathrm{P} + \mathrm{R}},$$

$$\mathrm{ACC} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FN} + \mathrm{FP} + \mathrm{TN}},$$

respectively, where true positive (TP) denotes that a hypopnoea testing sample is correctly classified into hypopnoea, false positive (FP) denotes that a normal breathing testing sample is incorrectly classified as hypopnoea, true negative (TN) denotes that a normal breathing testing sample is correctly classified as normal breathing, and false negative (FN) denotes that a hypopnoea testing sample is incorrectly classified as normal breathing.

In Table 1, we show the corresponding numerical performance for all trained RBF SVMs, which indicates that our monitoring system can detect hypopnoea with a very high success rate. Indeed, using 25 test samples, three SVMs classified all episodes of hypopnoea correctly. The results also show that lower complexity 2D SVMs, using (a_1, b_1) or (b_1, b_2) features, classified all instances of hypopnoea without any classification mistake.

7. CONCLUSION

Existing sleep monitoring systems are expensive and intrusive enough that they negatively affect the quality a patient's sleep. In this paper, we propose a non-intrusive three-part video monitoring system based solely on depth video recording and analysis. In the first part, for efficient compression we propose an alternating frame coding scheme, where different 8 of 11 available bits from captured depth images are extracted for H.264 real-time encoding. The uncoded 3 bits are recovered via a block-based search procedure at decoder. Meanwhile, we show that our LSB-MSB coding scheme outperforms 8-MSB coding scheme at mid- to high-bitrate region. In the second part, we estimate ellipse parameters for the patient's chest and abdomen, and using the estimated ellipse parameters, we detect hypopnoea via trained RBF SVMs with a very high success rate. Experimental results confirm that our proposed sleep monitoring system can be effective in detecting important sleep-disordered breathing events.

8. REFERENCES

[1] A. Malhotra and D. P. White, "Obstructive sleep apnoea," in *The Lancet*, July 2002, vol. 360, no.9328, pp. 237–245.

[2] P. Peppard et al., "Prospective study of the association between sleep-disordered breathing and hypertension," in *The New England Journal of Medicine*, May 2000, vol. 342, no.19.

[3] T. Wiegand et al., "Overview of the H.264/AVC video coding standard," in *IEEE Transactions on Circuits and Systems for Video Technology*, July 2003, vol. 13, no.7, pp. 560–576.

[4] C. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.

[5] P. Merkle, A. Smolic, K. Mueller, and T. Wiegand, "Multi-view video plus depth representation and coding," in *IEEE International Conference on Image Processing*, San Antonio, TX, October 2007.

[6] M. Tanimoto, M. P. Tehrani, T. Fujii, and T. Yendo, "Free-viewpoint TV," in *IEEE Signal Processing Magazine*, January 2011, vol. 28, no.1.

[7] G. Shen, W.-S. Kim, S.K. Narang, A. Ortega, J. Lee, and H. Wey, "Edge-adaptive transforms for efficient depth map coding," in *IEEE Picture Coding Symposium*, Nagoya, Japan, December 2010.

[8] W. Hu, G. Cheung, X. Li, and O. Au, "Depth map compression using multi-resolution graph-based transform for depth-image-based rendering," in *IEEE International Conference on Image Processing*, Orlando, FL, September 2012.

[9] M. W. Lee and R. Nevatia, "Body part detection for human pose estimation and tracking," in *IEEE Workshop on Motion and Video Computing*, Austin, TX, February 2007.

[10] J. Shotton et al., "Real-time human pose recognition in parts from single depth images," in *IEEE International Conference on Computer Vison and Pattern Recognition*, Collorado Springs, CO, June 2011.

[11] D. Falie, M. Ichim, and L. David, "Respiratory motion visualization and the sleep apnea diagnosis with the time of flight (tof) camera," in *1st WSEAS International Conference on Visualization, Imaging and Simulation*, Bucharest, Romania, November 2008.

[12] M.-C. Yu et al., "Breath and position monitoring during sleeping with a depth camera," in *International Conference on Heath Informatics*, Vilamoura, Portugal, February 2012.

[13] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2003.

[14] J. Keikkila and O. Silven, "A four-step camera calibration procedure with implicit image correction," in *IEEE Conference on Computer Vision and Pattern Recognition*, San Juan, Puerto Rico, June 2007.

[15] W. Sun et al., "Rate-distortion optimized 3D reconstruction from noise-corrupted multiview depth videos," in *IEEE International Conference on Multimedia and Expo*, San Jose, CA, July 2013.

[16] P. Rosin, "Analysing error of fit functions for ellipses," in *Pattern Recognition Letters*, 1996, vol. 17, no.14, pp. 1461–1470.