



# A novel depth image analysis for sleep posture estimation

Maryam S. Rasouli D<sup>1</sup> · Shahram Payandeh<sup>1</sup>

Received: 20 June 2017 / Accepted: 14 April 2018 / Published online: 30 April 2018  
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

## Abstract

Recognition of sleep posture and its changes are related to information monitoring in a number of health-related applications such as apnea prevention and elderly care. This paper uses a less privacy-invasive approach to classify sleep postures of a person in various configurations including side and supine postures. In order to accomplish this, a single depth sensor has been utilized to collect selective depth signals and populated a dataset associated with the depth data. The data is then analyzed by a novel frequency-based feature selection approach. These extracted features were then correlated in order to rank their information content in various 2D scans from the 3D point cloud in order to train a support vector machine (SVM). The data of subjects are collected under two conditions. First when they were covered with a thin blanket and second without any blanket. In order to reduce the dimensionality of the feature space, a T-test approach is employed to determine the most dominant set of features in the frequency domain. The proposed recognition approach based on the frequency domain is also compared with an approach using feature vector defined based on skeleton joints. The comparative studies are performed given various scenarios and by a variety of datasets. Through our study, it is shown that our proposed method offers better performance to that of the joint-based method.

**Keywords** Sleep posture estimation · Human posture estimation · Machine learning · Depth data

## 1 Introduction

Sleep is one of the most important parts of human life which can strongly affect both physical and mental state. The quality of sleep can determine the mental outlook and performance of an individual in his/her daily activities. This can be an important factor in yielding to various health problems like diabetes, obesity, and depression (Falie et al. 2008). There are several indicators to diagnose sleep quality. Among them, the sleep posture and its changes can be an important indicator for sleep disorders such as snoring, night sweats, narcolepsy and especially sleep apnea (Hao et al. 2013) (i.e. recurring of restricted airflow leading difficulty in breathing during sleep). In this regard, monitoring sleep posture and frequency of its changes are widely

utilized for recognition and treatment of sleep apnea (Menon and Kumar 2013; Oksenberg et al. 2012; Oksenberg and Silverberg 1998; Sutherland and Cistulli 2015). In addition, in Gordon et al. (2007), authors show that supine postures can deteriorate the spine and increase back problems. Sleep disorders are even more important among elderlies. In Yao and Cheng (2008), authors illustrate the notable influence of sleep quality on the depression and social networking among independent older people. In Karlsen et al. (1999), insomnia is introduced as one of the most important factors to cause depression among older adult suffering from Parkinson. These factors are mentioned to be more significant than the severity of the disease itself. In Barbara and Ancoli-Israelb (2001) and Bixler et al. (1998) authors focused on investigating the influence of aging on sleep disorders. They show that sleep apnea is more severe among elderly so by increasing the average age of population the need for sleep monitoring also raises.

Sleep posture can also play an important role in recovering from serious surgical procedures. Nurses and hospital staff need to monitor the patient continuously during their sleep, in order to prevent any further problems due to the exertion of extra pressure on affected areas. However,

---

✉ Maryam S. Rasouli D  
rasoulid@sfu.ca

Shahram Payandeh  
payandeh@sfu.ca

<sup>1</sup> Networked Robotics and Sensing Laboratory, School of Engineering Science, Simon Fraser University, Burnaby, Canada

reliance on personnel's observations cannot guarantee a timely warning on the changes in the sleep posture (Liu et al. 2013). Thus, designing and developing of an automatic and less-human-dependent system for sleep posture monitoring is necessary to improve patient's life quality, promote care quality, and finally allow integration of sleep posture monitoring with the medical diagnosis system (Torres et al. 2016).

Sleep posture estimation is a special case of human posture estimation. In literature, many methods have been employed to solve this general problem. However, estimation of human posture during sleep introduces some additional constraints and limitations to the general problem of human posture estimation. Variations in illumination conditions, a weak contrast between the subject and background (especially the bed), and privacy concerns are just some of the challenges in sleep posture estimation.

Visual sensing method through RGB cameras is one of the most popular approaches for human posture estimation. However, these methods need some level of illumination or some source of light to capture an appropriate image from the subject and in general, they can be considered as more privacy-invading sensing approaches.

In Panini and Cucchiara (2003), Wachs et al. (2009) and Fujiyoshi et al. (2004) authors utilize a 2D projection of a single camera. In Panini and Cucchiara (2003) authors first detected and tracked human silhouette in a single camera view. They then try to define four main body postures using a statistical example-based classifier. This is done through a probabilistic map of each predefined posture during the training phase and then map and classifies each of the human postures. In order to improve their proposed method, they present a head detector which investigates body's boundaries using Jarvis-Convex Hull scan and light tracking phase.

Methods based on 2D projected images from RGB data have some limitations, including viewpoint dependency, occlusion sensitivity, environment features dependency, and environmental noise. To overcome some of the limitations of these 2D approaches, a 3D model of the human body was further utilized. For example (Boulay et al. 2006), generates some 3D models of human postures based on some of the feature points on the human body (e.g. abdomens, shoulders, elbows, knees, hips and pelvis) to extract 2D projection. To generate each 3D model, they introduce three Euler angles for each articulation in each posture and finally employ the Mesa library to illustrate 3D models. Then they detect foreground objects from a 2D projected view to map 2D points onto various 3D models of human postures.

Using multiple cameras is an alternative approach to capture a 3D image of the body postures. In Pellegrini and Iocchi (2007), authors track and detect the human body in stereo images. They, first, extract the body and then analyze the tracked body to detect the body's posture. They

categorize main postures into five different groups (up, sit, bent, on-knee, laid) and also define three principal points in the head, pelvis, and legs. They calculate the angles of the vectors connecting these principal points and current height of the body to utilize them for the classification. They define a training phase to set these parameters in their algorithm.

Other sensing modalities are introduced in order to track human's body and estimate its postures. For example Luštrek and Kaluža (2009), takes advantage of using infrared motion capture system in order to tag different parts of human bodies including shoulders, elbows, wrists, hips, knees, and ankles. This is then used to recognize four different behaviors, falling, lying down, sitting down, standing/walking, sitting and lying. They then defined some attributes such as the angles between different parts of the body as a part of the machine learning algorithm for detecting fall.

Using Kinect I and II sensor is another efficient way to estimate human posture since it can provide the depth map of the environment. Kinect II takes advantage of IR laser and time of flight sensor to measure the distance, while Kinect I employs IR light patterns. In addition, Kinect II provides more Filed of view and more resolution compare to Kinect I.

In Shotton (2013), authors employ the Kinect I to estimate the human postures in a single depth map image. They define a feature for some selected pixel in the image based on the depth value of the pixel and its neighbors to create a randomized decision forest and train the forest to determine the position of each pixel in the body. Then, they utilize a local mode finding method base on mean shift to accumulate the result of each pixel to estimate the final skeleton position (they define 31 different labels for different parts of the body). They also utilize a comprehensive dataset including a wide variety of postures. However, lack of the sleep postures is tangible in this dataset.

In Xiao et al. (2012), authors use 31 labels for different parts of the body and generate a function to calculate the features of the depth data for each image. After generating the features, they propose an algorithm based on machine learning using the randomized decision forest classifier. Finally, they present a human skeleton joint model to collect information across all pixels and determine the resultant estimation of the human posture. The paper also suggests the application of their method to detection of various other postures, but the results are limited to posture estimation of a calibration cube.

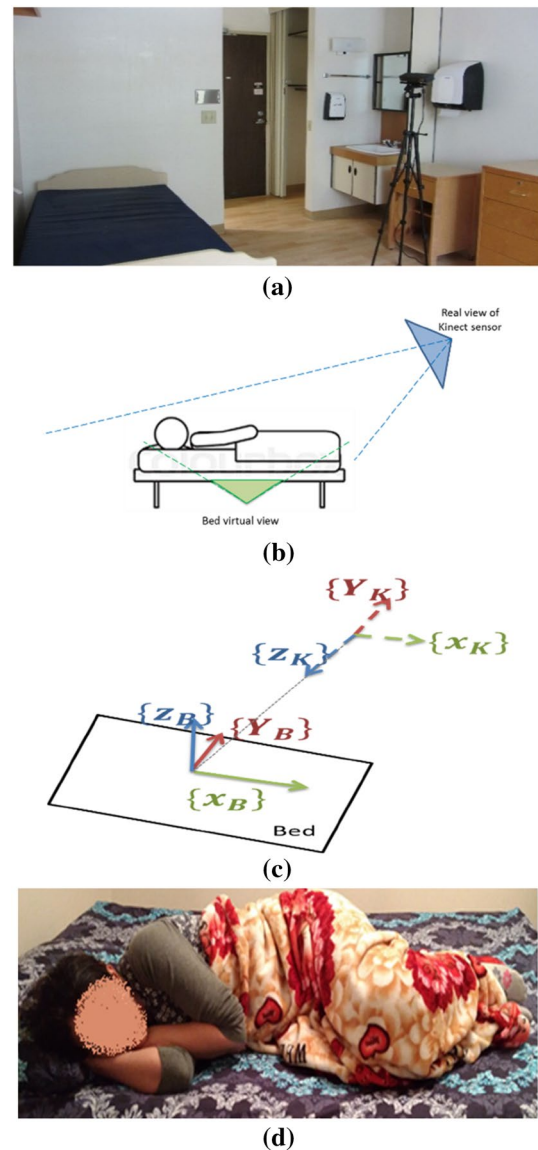
The capability of the depth sensor to work in low illumination makes it a suitable candidate for sleep posture estimation and investigation of sleep disorders. In addition, since such sensor can also operate in dark setting, its operation does not create any other discomfort due to the presence of light source and also their integration offers another low-cost alternative to other sensing modalities (Yang et al. 2014; Torres et al. 2015; Lee et al. 2015; Falie et al. 2008). (Falie

et al. 2008) and Yang et al. (2014) focused on detecting apnea with Kinect sensor. Apnea and hypopnea are detected in Yang et al. (2014) by concentrating on detecting abdominal and chest movement using Kinect I. Authors fit a dual ellipse to the observed depth data and use the axes of the ellipse for providing graph based classification. They also define an optimization problem to de-noise the collected depth data and remove the unwanted temporal flickering.

In Yoshino and Nishimura (2016), authors define three different postures (supine, sideway, and prone position). They detected the head using color map under the assumption of dark hair color of the local majority. They divide a human body into four different equal regions (head, chest, stomach, and legs) which the detected head is utilized to determine other regions. They then compare histogram of the depth data of each region to estimate sleep postures. In Metsis et al. (2014), a Kinect I and an force sensing array (FSA) pressure mat are employed to extract features for machine learning to detect sleep posture changes during sleep. They classify the position of the target into five groups e.g. side (left and right) and back and supine and also sitting.

More detailed sleep posture recognition methods are beneficial in a more critical situation (i.e. hospital and ICU) and consequently need a more complex system to obtain a more accurate result for wider classification of sleep posture. In Torres et al. (2015, 2016), the proposed method is able to detect different sleep postures as fetal, log, yearner, soldier and faller. They use quite a complex system in their algorithm including multi-sensor (depth, RGB, and pressure sensors) and multi-view of the camera and Kinect sensors (top, head, and side). They are dependent on the pressure sensor to distinguish the presence of blanket as an occlusion.

In this paper, the problem of sleep posture estimation for various caregiving facilities such as the elderly care by using a single depth sensor. To effectively distinguish each sleep posture, the overall depth data are sampled with various 2D scans. These scans are then analyzed using FFT for the definition of feature vectors which are then utilized to train the support vector machine (SVM) classifier. As it is shown, the proposed method can detect the sleep posture even when the person is covered with a blanket. A blanket cover is used to better simulate the natural sleeping environment. In addition, our approach is independent of the tilt angle of the sensor. This feature is also important since a number of proposed approaches relays on the sensor to be installed directly on top of the bed and facing the bed (Metsis et al. 2014; Torres et al. 2015). It is shown that in our proposed method the depth sensor can be mounted on the sides of the bed having various adjustable tilting angle. This feature makes the installation of the sensor easier and thus make it appropriate for in-house monitoring or a care center which can place the sensor in a comfortable range away from the direct line of sight of the person (see Fig. 1).



**Fig. 1** **a** Environment setup located at a care center environment. Figure shows that the possible mounting location of the sensor including the side walls, facing the bed and away from the line of sight of the person (society). **b** Schematics of the placement of the sensor during the experiment and the definition of the virtual view (i.e. a view from the bed advantage point to map the body point of cloud to the bed frame). **c** Bed coordinate frame ( $x$ ,  $y$ , and  $z$  axes are along height, width and the normal of the bed, respectively) and the Kinect coordinate frame. **d** A sample subject under a thin blanket

The remaining of the paper is organized as follows. The overview of preliminaries including data preprocessing and 2D cross-sectional is presented in Sect. 2. The feature extraction method is described in Sect. 3. Section 4 includes the dataset preparation and the experimental results. Finally, Sect. 5 presents a discussion and concluding remarks.

## 2 Preliminaries

In this work, a single depth sensor (Kinect II sensor) is employed to estimate the sleep postures. A depth sensor estimates the distance from an object with respect to the image plane of the sensor (usually the measurements can be represented in millimeter using the sensor resolution of the depth map as  $512 \times 424$  obtained at 30 frames per second). Our approach is divided into two steps, first, is estimating main sleep postures and second is estimating the configuration of hands and legs during sleep.

Sleep postures can be divided into two main groups, namely side, and supine postures. In many sleep monitoring studies, it is usually sufficient to detect changes among these dominant postures. For example, in Menon and Kumar (2013) authors provide a review of the postures which can affect sleep apnea. In their review, they summarize the results of 14 different studies which all emphasize the importance of identifying two dominant postures during the monitoring process.

The classification of sleep postures is usually defined based on the position of hands and legs. In Sleep Position Gives Personality Clue (2003), the most common sleep postures are divided into six groups (Fig. 2). In general, four different hand positions can be identified. For example, in the soldier posture, hands are aligned with the body and are positioned close to the body (referred to the first position). In foetus and yearner postures, the position of hands is around the head and located on one side (referred to the second position). In log posture, hands are positioned downward and to the side (the third position). Finally, in freefaller and starfish hands are positioned around the head (the fourth position).

Similarly, three different positions can be recognized for legs. In the foetus posture, legs are curled up toward the stomach (the first position of legs). In yearner and log postures, legs are stretched down together (the second position) and in the soldier, freefaller, and starfish postures legs are spread and stretched down separately (the third position).



**Fig. 2** Most popular sleeping postures (Sleep Position Gives Personality Clue 2003)

To extract the features from the depth signals, fast Fourier transformation (FFT) is performed on selected 2D scan planes. Cross-sectional 2D scans of the 3D point cloud are performed on instances of the captured depth map in various positions.

In this section, some preliminary steps in preparing the raw depth data are described follow up by a discussion on the feature selection approach based on 2D scans and their FFT analysis.

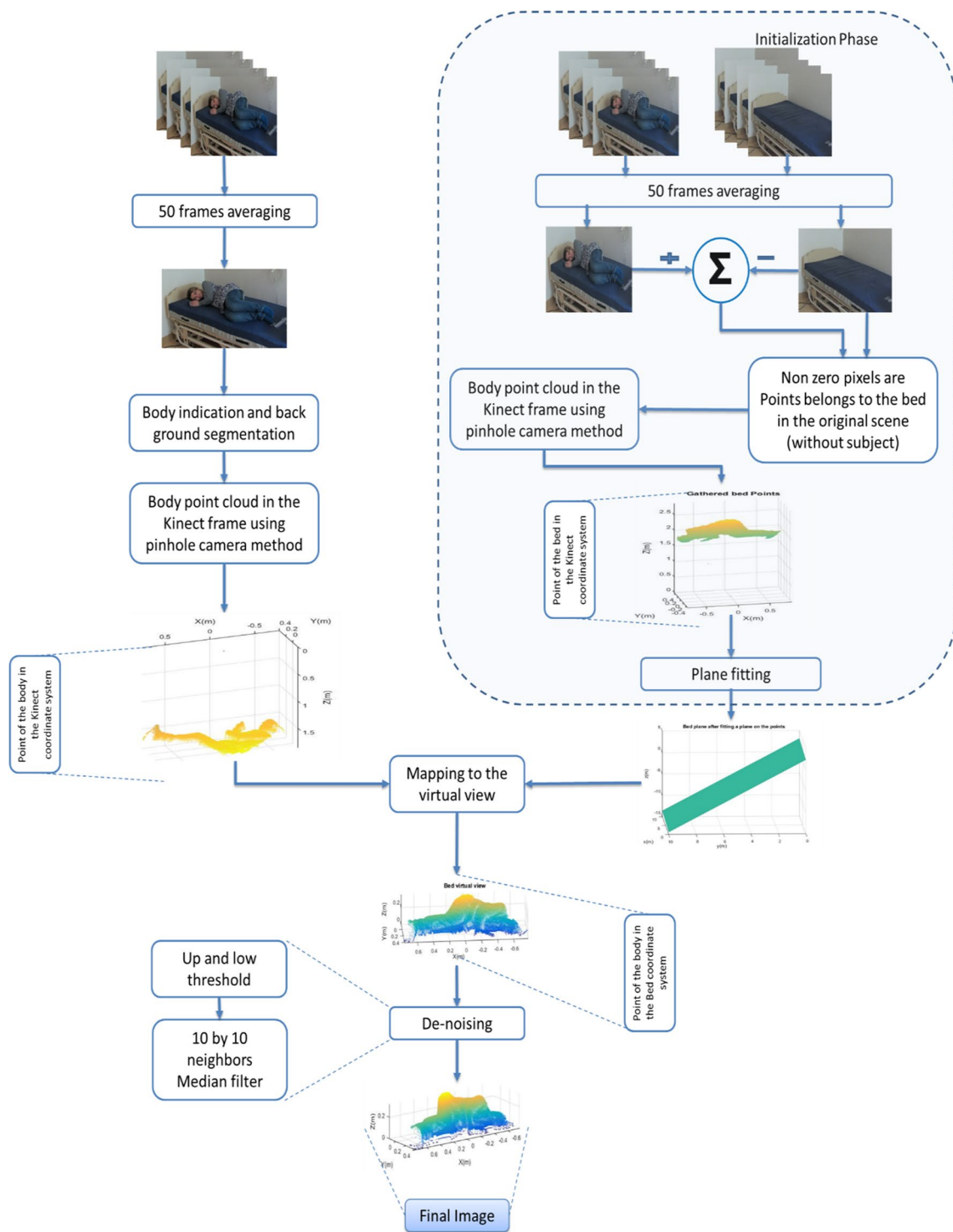
### 2.1 Depth data pre-processing

As it is shown in Fig. 1a, the depth sensor can be installed in the monitoring room on any of the accessible walls. For our experimental studies, the sensor has been placed on a tripod positioned beside the bed as shown in Fig. 1a. Each depth frame (depth map) contains a measurement of the distance of objects to the sensor's image plane based on optical signal processing (Dal Mutto et al. 2012) (which uses a projector and a receiver to implement a stereo vision algorithm resulting in noisy images). The dependency on the sensor's plane and noise should be resolved first, in order to provide a better estimation.

Figure 3 shows the overall procedure in preparing the data to capture various 2D scans, in order to provide a unique body measurement data regardless of the position of the sensor and its inclination angle. As it can be seen in this Figure, after averaging over frames, median filters are utilized for noise removal. Then, the bed plane is extracted, and the body has been aligned to the bed plane to make the sample ready to extract features for sleep posture estimation. In the following details associated with each of these steps are presented.

One way to reduce the effect of noise is averaging over depth map (i.e. use the average of obtained depth value over 50 successive depth map). In Lachat et al. (2015), the advantages of averaging over the different numbers of successive depth map evaluated. It also shows that the accuracy improves by increasing the number of frames. However, it is also shown that using more than 50 frames for averaging results has no effect in further removing the unwanted noise.

The Kinect Body Indicator (Bodyframe Class 2014) is utilized to distinguish the human body from the background. This indicator first determines whether each pixel belongs to a human body and then gives the depth measurement of the body with respect to the origin of the sensor which changes depending on the position and orientation of the sensor with respect to the body. First, the 2D image plane coordinates are mapped to that of the 3D sensor frame using a pinhole camera model (Dal Mutto et al. 2012). Figure 1c shows the sensor coordinate system. where each pixel in the depth map image ( $I(u, v)$ ,  $0 \leq u \leq 512$ ,  $0 \leq v \leq 424$ ) is mapped to a real point cloud  $(X, Y, Z)$  (all in meter) which shows the



**Fig. 3** The preprocessing of the raw depth data. In the initialization phase the plane of the bed is first estimated. Each 50 frames are then averaged and mapped to the virtual view. Finally, median filter and signal thresholding are applied to reduce the unwanted noise

actual position of a pixel  $I(u, v)$  with respect to the coordinate frame of the Kinect sensor.

The calculated values of  $X$ ,  $Y$ , and  $Z$  show the position of the body with respect to the frame of the Kinect sensor.

These values change when the sensor changes since they are dependent on the position and inclination of the sensor. To overcome this dependency and still maintaining the global depth profile of the sensed data, a bed virtual view

is defined which maps a depth image from the bed coordinate in the reference frame to this new view. Mapping the depth data to the virtual view facilitates the signal classification method by making the view-invariant with respect to the sensor placement and its orientation with respect to the bed. This virtual view provides a unique representation of the body regardless of the Kinect inclination angle (e.g. tilt angle) or its position (the height of the Kinect). Figure 1b shows the bed virtual view and real Kinect sensor view and Fig. 1c shows the bed coordinate system and the corresponding Kinect coordinate system. In order to map the current Kinect view of the body to the bed view, the first step is to obtain the bed's position in the Kinect view.

The main idea is to first find the bed plane in the depth sensor view using few sample points which are located in the physical position of the bed. In order to obtain these points, two different depth map frames are employed. The first frame is captured when there is no subject sleeping on the bed and second frame is captured when the subject is sleeping on the bed. We refer to these two frames as  $I_{pr}(u, v)$  and  $I_{ORG}(u, v)$ , respectively. Then, we define Eq. (1).

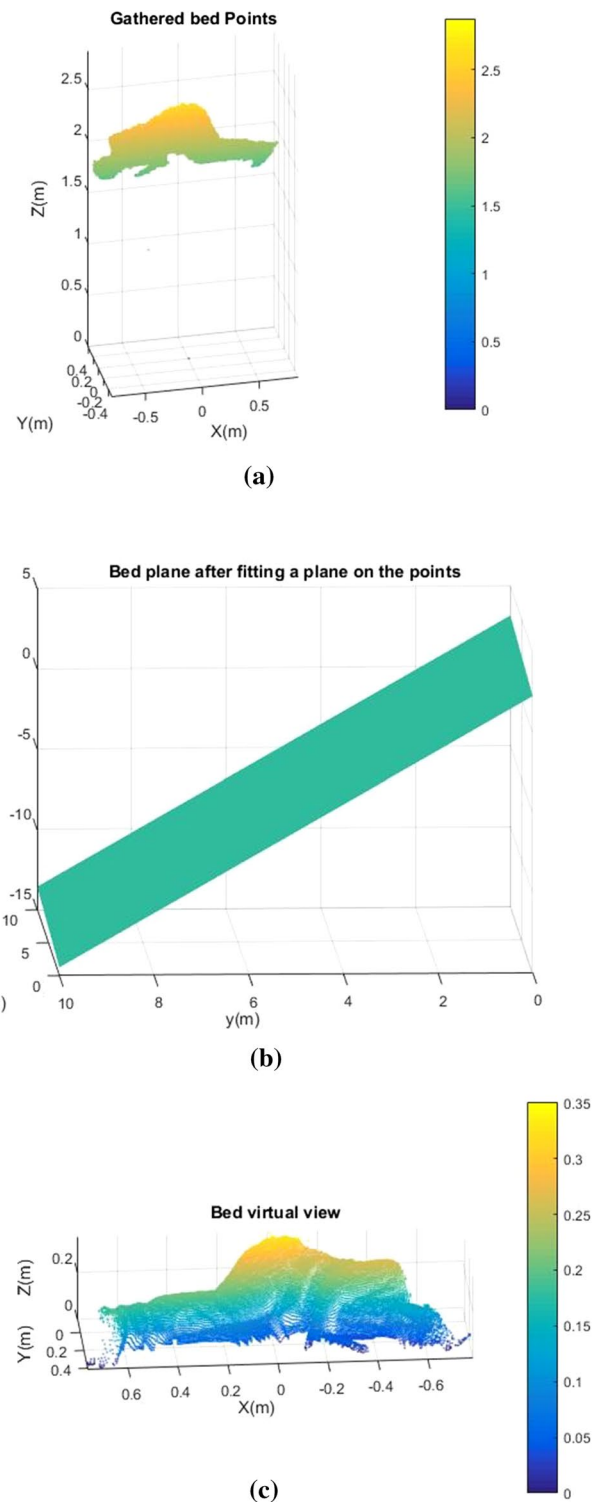
$$I_{bed}(u, v) = \begin{cases} I_{pr}(u, v), & I_{pr}(u, v) - I_{ORG}(u, v) \neq 0 \\ 0, & O.w \end{cases} \quad (1)$$

The  $I_{bed}(u, v)$  contains the value of the distance between bed points and the sensor origin and it contains zero for other points which don't belong to the bed. Let's define points  $(X_b, Y_b, Z_b) \in S_{bed}$  where  $S_{bed}$  is the set of bed points (points which  $I_{bed}(u, v) \neq 0$ ) transformed to the real point cloud. Figure 4a shows the point cloud of the bed ( $S_{bed}$ ) in the Kinect sensor view.

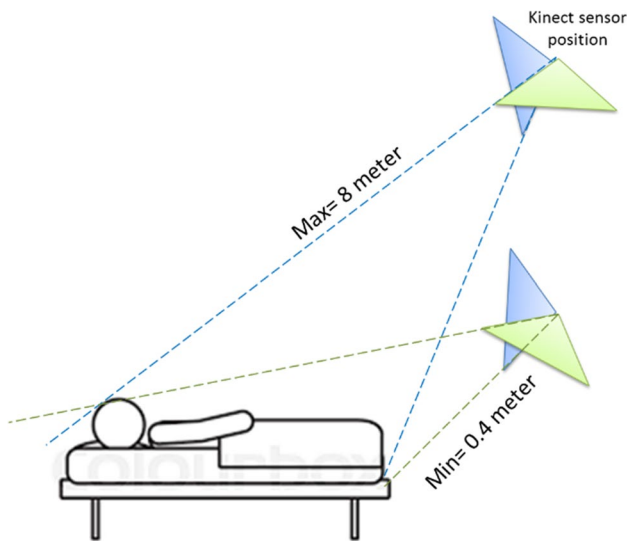
The description of the plane, representing the bed in the sensor coordinate frame, can be defined as:  $P_{bed} = A x + B y + C z + D = 0$ . This representation can be computed through a collection of measured sample points belonging to the bed, or  $(x_0, y_0, z_0) \in S_{bed}$  by minimizing  $\sum \left( \frac{Ax_b + By_b + Cz_b + D}{\sqrt{A^2 + B^2 + C^2}} \right)^2$ . Figure 4a shows an example of collected data (i.e.  $(X_b, Y_b, Z_b) \in S_{bed}$ ) and Fig. 4b shows the final plane using this approach.

In order to obtain the virtual view of the sleeping body, we compute the distance between each body's point and the bed plane. Figure 4c shows the resultant segmented body in the bed virtual view. It should be mentioned that the sensor should be able to capture the whole body in each position and due to limited field of view (FOV) and feasible range of Kinect, not all Kinect position can record the whole-body data. Figure 5 illustrates the impact of Kinect limitation on the final possible Kinect's position.

Figure 6 illustrates the image of a sleep posture under two different tilt orientations of the depth sensor. Figure 6a shows the bed plane in these two different sensor's tilts and Fig. 6b shows the final posture under both camera positions.



**Fig. 4** a Bed points in Kinect sensor view  $I_{bed}$ . b Fitted plane to the bed in the Kinect sensor view. c The final body point cloud in bed virtual view. Vertical color bar shows the distance of the point cloud from the X–Y plane in each view, the vertical color map shows the color of each distance in meter. (Colour figure online)



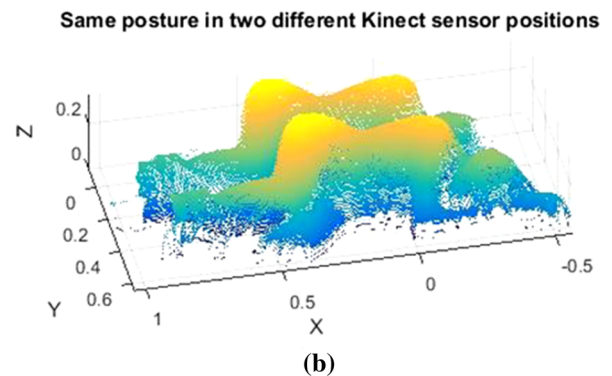
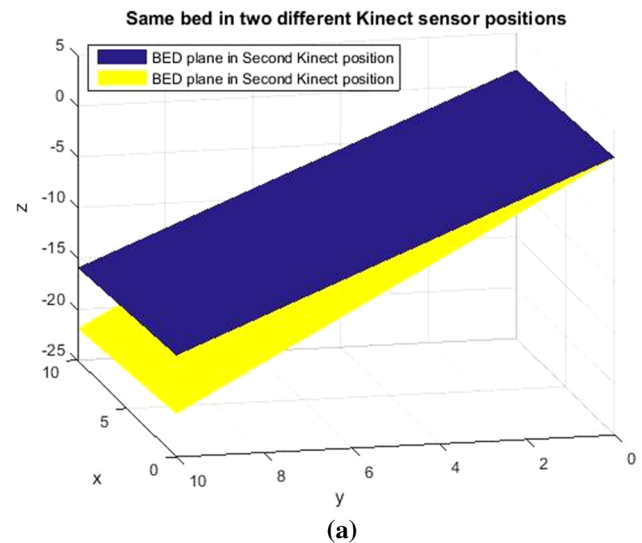
**Fig. 5** The range of the Kinect position (including tilt and height of the sensor). The limitation in the Kinect field of view and recording range will restrict the final possible position of the Kinect

As you can see, although, the subject's position in  $X$ - $Y$  plane is slightly different,  $Z$  values on which our proposed feature extraction method is based on, are almost the same. The difference in  $X$ - $Y$  plane is due to different sensor tilts.

Finally, a 2-D median filter is applied to the resulting image to smooth it and remove the remaining noise. Figure 7 shows the impact of 2D-median filter on the final image.

## 2.2 2D cross-sectional

Figure 8 shows the main idea of cross-sectional scans along  $X$  and  $Y$  axes, i.e., horizontal and vertical scans, respectively. Horizontal scans are along the height of the bed and vertical scans are along the width of the bed. In Fig. 8, vertical scans are shown in blue and vertical scans are shown in red. As it can be seen in the figure, each scan reveals the relative information of a particular part of the body. The location of the scan plane is very important and contains other spatial information regarding the local position with respect to the bed frame and the expected body position. Designing a method which can sequentially define an order of scan planes can also be used as a part of the efficient signal classification method. For example, a vertical scan from the upper part of the body reveals some signal information about the possible position of the head or a vertical scan of the lower part of the body can contain the data pertaining to the position of the legs. The 2D scans of the depth data are first normalized within the range of  $[0, 255]$ . The fast Fourier transformation (FFT) can be performed on the 2D cross sectional scans signals. This will result



**Fig. 6** Comparison of two different Kinect sensor tilts. **a** The bed planes. As it can be seen the bed plane is different in each sensor position since the angle between the Kinect sensor and bed is different for each tilt. **b** The final body posture. The same posture in two camera tilts as it can be seen there is no angle between these two postures since both are mapped to the bed view

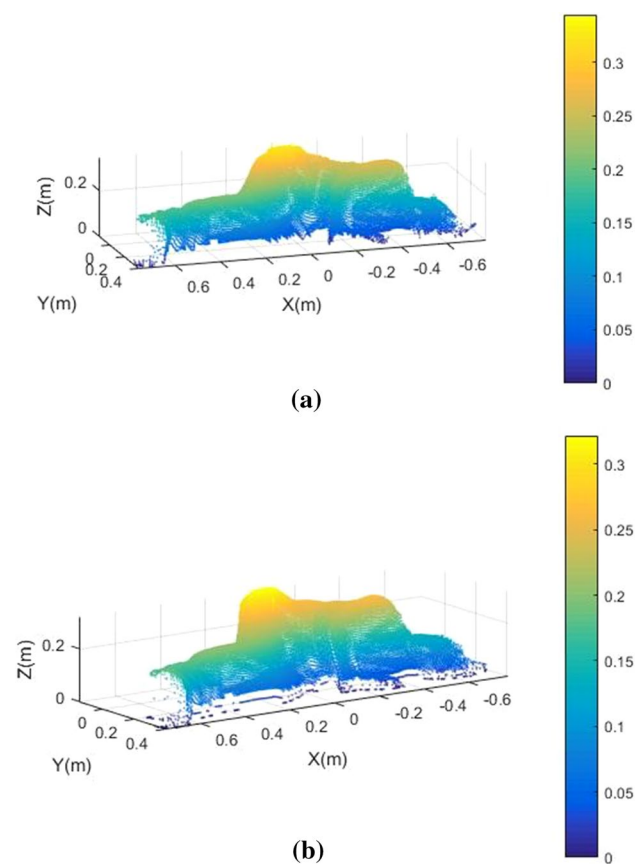
in a set of metrics which can be used as a feature vector in the follow-up pattern classification for sleep posture detection.

## 3 Feature selection

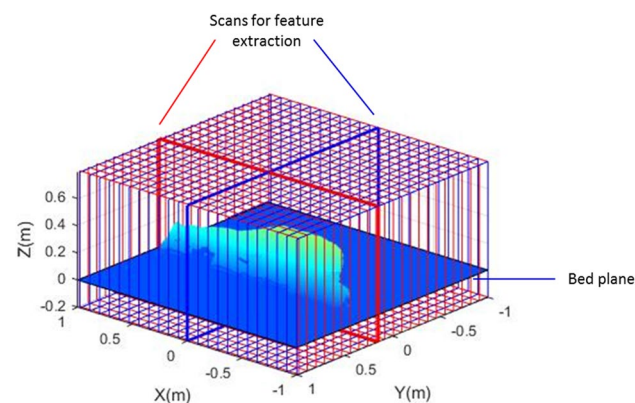
Inspired by the fact that the position of different parts of the body and the relationship among them can assist in recognizing the sleep posture, the features from the horizontal and vertical 2D cross-sectional scans are extracted from the depth map. In the following, we dig into the details of the features.

### 3.1 Features selection for main postures classification

In general, there exists potentially a large number of locations for defining the scan planes. For two main postures i.e. supine



**Fig. 7** The effect of the threshold and median filter on the final image. **a** After averaging of frames and applying threshold and changing the viewpoints. **b** Final body point cloud after applying median filter



**Fig. 8** 2D cross sectional scans. Red and blue lines represent horizontal and vertical scans, respectively, the highlighted red and blue lines show the sample scans which are taken to extract the feature of each posture. (Colour figure online)

and side postures the most relevant scans with higher information about the sleep posture should be selected. The scans with the most body pixels are selected as the final scan planes.

Figure 9 shows two sample scans by performing this method. Figure 9a, b show the view of the body in supine and side postures, respectively. Figure 9c, d illustrate selected horizontal (highlighted red line) scan and their FFT results of their above postures and similarly Fig. 9e, f show the vertical (highlighted blue line) scan and their FFT results of their above postures. The FFT of both vertical and horizontal scans are utilized to extract features to estimate the main postures.

### 3.2 Features selection for classification of hands and legs

Various scans in different locations can also result in useful information about the sleep posture and the details of the body parts. Exploring the most relevant scan planes are required to efficiently find details of hands and legs configurations during sleep in order to estimate the details of each sleep posture.

Generally, with the assumption that the position of the head is located on the top of the bed, the point cloud representing the sleeping body can be divided into two equal regions, which the first region belongs to the upper part of the body and the second one is the foot region.

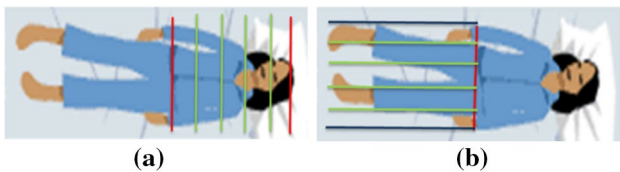
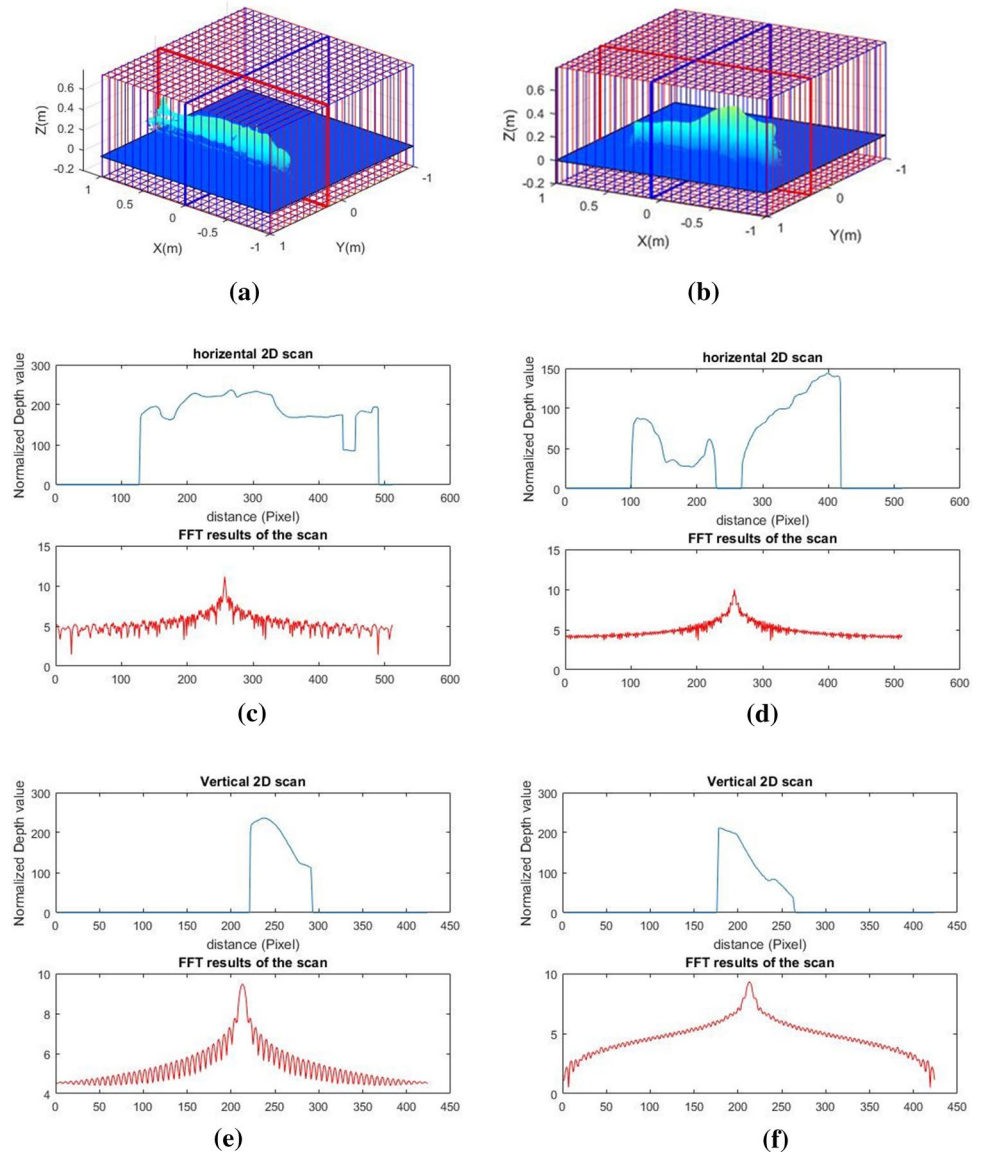
Figure 10 shows the location of the scans which should be performed to estimate the position of hands and legs. To examine the positions of hands, the first region of the body (upper part of the body) is considered and four scans are performed. Each scan is located in 12.5% of whole pixels of the body in the upper part. Similarly, to estimate the position of legs, four horizontal scans are performed in the second region of the body (foot region). Each scan is located at 25% of the whole width of the body. Figure 11 shows the horizontal scans to estimate the position of legs. The corresponding posture is shown on the right side, and the signal of the selected scans with its FFT result is shown in the left hand of the figure. The blue lines show the horizontal boundary scans of the body and the green lines are selected scans. Selected scans signals are shown in the next column, starting from the far-right scan. Similarly, Fig. 12 illustrates the obtained vertical scans to estimate the position of hands. The corresponding posture is shown on the right side and its corresponding scan and the FFT result of each scan is shown in the left hand of the figure. The red lines show the middle and the lowest scans of the body and the green lines are selected scans to estimate the position of hands. The first scan, in the middle column, is pertaining to the head of the subject and others are shown in the same order.

## 4 Pattern classification and experimental results

In order to evaluate the proposed approach, a dataset including different sleep postures under a variety of conditions is constructed. Fourteen subjects participated in the dataset



**Fig. 9** Sample horizontal and vertical scans. **a** The supine posture red lines are possible horizontal scans and the highlighted red line is the selected horizontal scan and blue lines are the vertical scans and the highlighted blue line is selected vertical scan. **b** The side postures with the same colored line as **a**. **c** Corresponding selected horizontal scan of the supine posture in **a** and its FFT result. **d** Corresponding selected horizontal scan of the side posture in **b** and its FFT result. **e** Corresponding selected vertical scan of the supine posture in **a** and its FFT result. **f** Corresponding selected vertical scan of the side posture in **b** and its FFT result. (Colour figure online)



**Fig. 10** Scans to estimate the position of hands and legs. **a** Four vertical scans to extract the features for estimating hands positions. **b** Four horizontal scans to extract the features for estimating legs positions

preparation sessions. None of them knew about the final purpose of the experiments. In each session, the subject was requested to sleep with different postures as they usually do in their bed. In addition, in several sessions, the subject is covered with a thin blanket (Fig. 1d). Table 1 summarizes

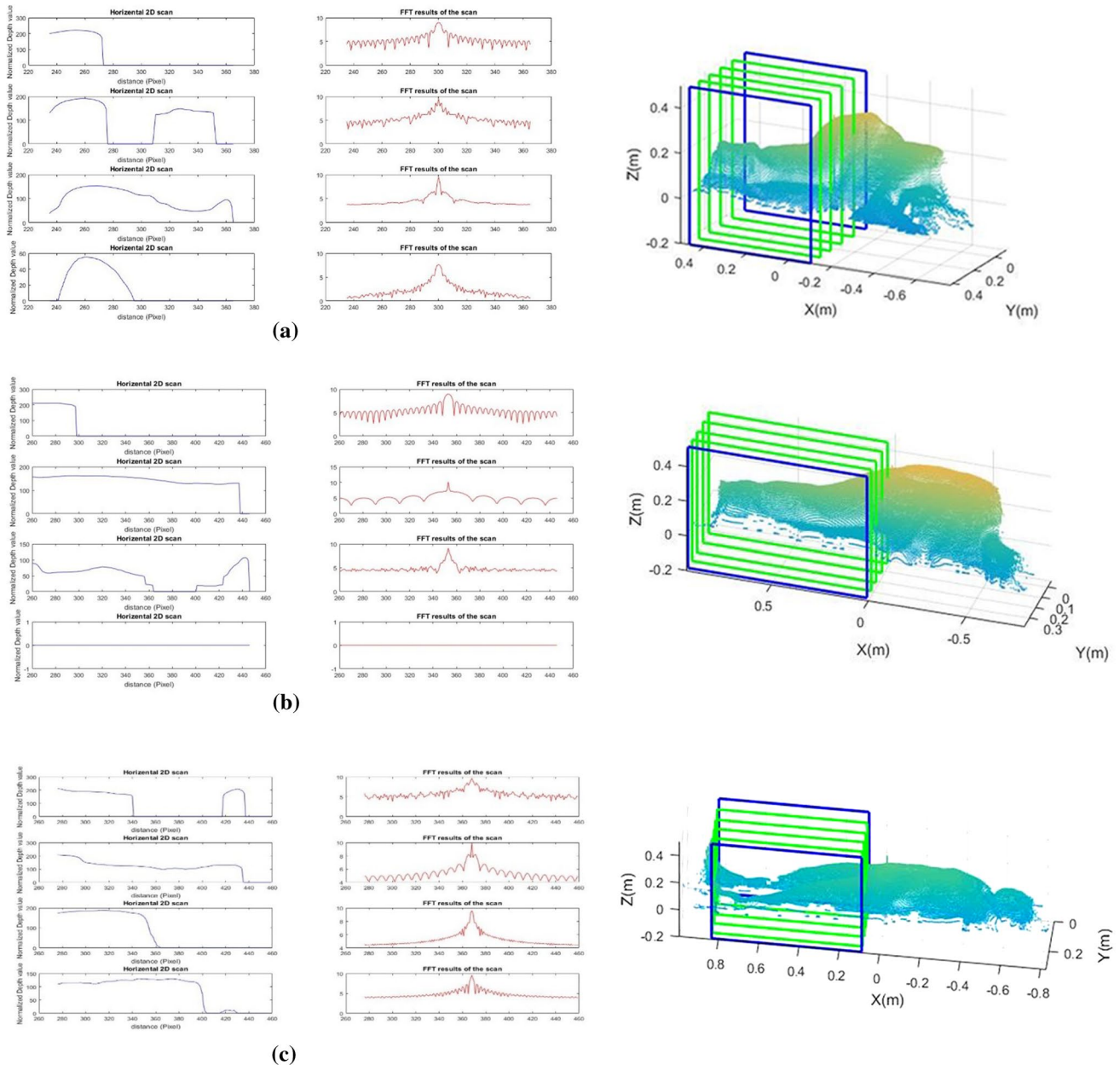
the samples in the dataset. The height and weight of subjects are reported as an indicator of the body mass. Also, men and women have different body mass. The number of data collected for main postures is also reported in this table.

### 4.1 Main postures estimation

The FFT result of each scan reveals the frequency component of the scan signal. In general, it was observed that a scan belongs to smoother postures (i.e. supine) contains lower frequencies components compare to a posture such as a side posture. The 2D cross-sectional scans provide a smaller number of features about the posture in comparison with the FFT of the whole image. Still, each 2D scan contains many data with some redundant information. For example, to estimate the main postures, 10 vertical and 10

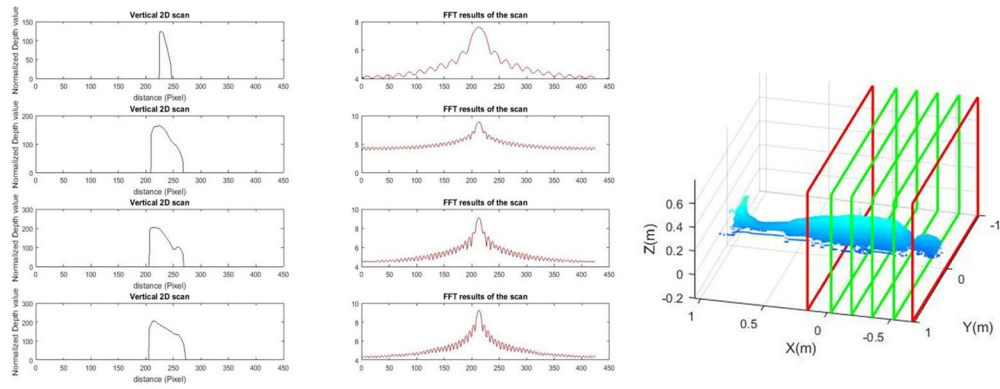
horizontal scans are utilized providing 9342 features for each depth map in the dataset. In order to extract more dominant data, T-test ranking method is employed which is a statistical method deployed, in order to rank features for training a classifier (Theodoridis and Koutroumbas 1999). Figure 13 illustrates ranking of the features associated with FFT result of 2D cross-sectional scans. As it can be seen, the first 3000 features (i.e. frequency components) are the most dominant features which were then used to train the SVM model.

**Fig. 12** Scans to estimate hands positions. **a** First hands position and its corresponding scans. The red lines show the middle and the lowest scans of the body and the green lines are selected scans to estimate the position of hands. Each blue signal in the left column is one of the scans corresponding to one of green scan and the red signal in the middle column is its corresponding FFT result. **b** Second hands position and its corresponding scans. **c** Third hand position and its corresponding scans. **d** Forth hand positions and its corresponding scans. (Colour figure online)

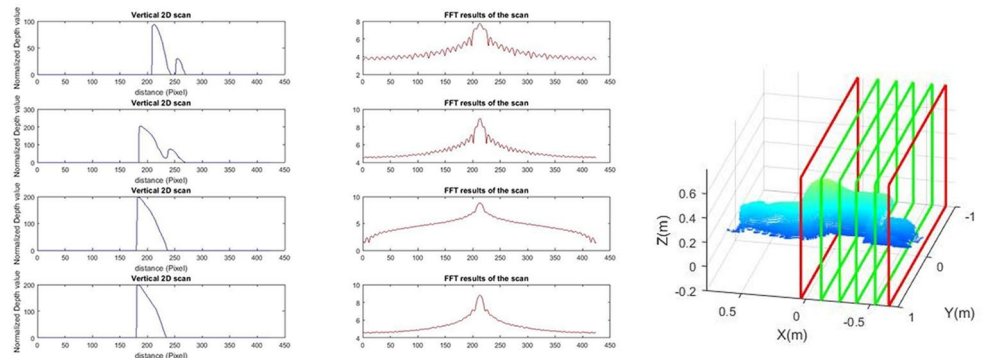


**Fig. 11** scans to estimate position of legs. **a** First leg position and its corresponding scans the blue lines show the horizontal boundary scans of the body the green lines are selected scans to estimate the position of legs. Each blue signal in the left is one of the scans cor-

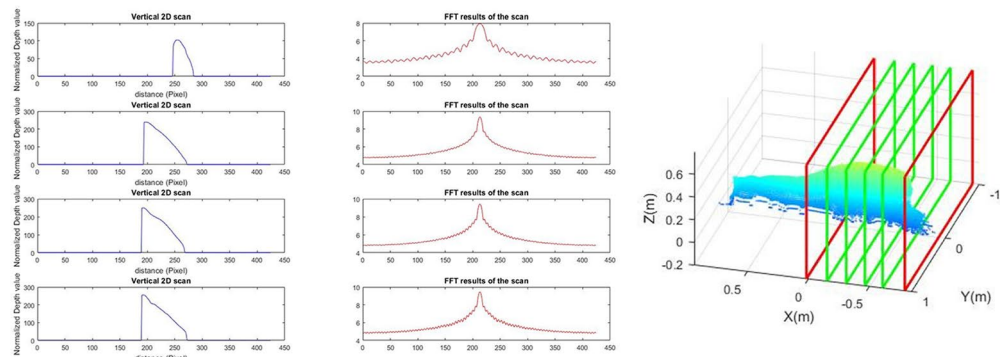
responding to one of green signal and the red chart is its corresponding FFT result. **b** Second leg position and its corresponding scans. **c** Third leg position and its corresponding scans. (Colour figure online)



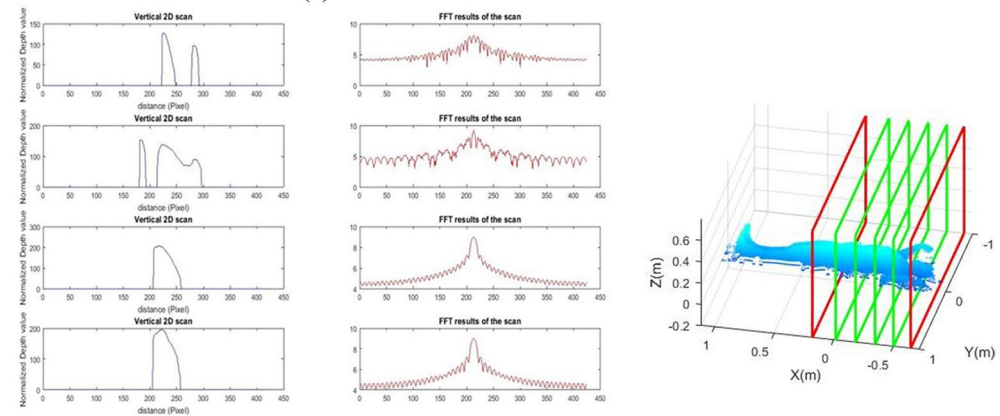
(a)



(b)



(c)

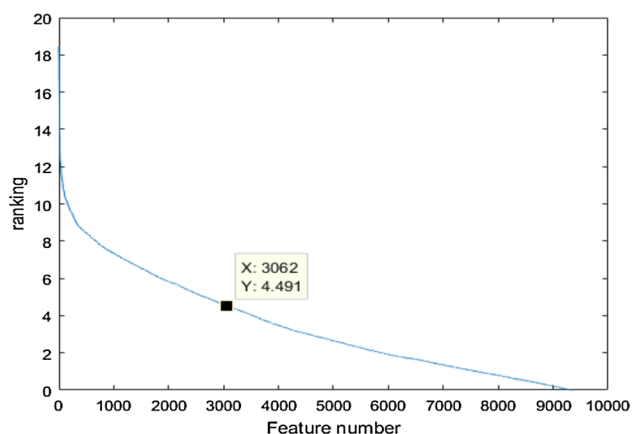


(d)

**Table 1** Details of the employed dataset to evaluate the proposed sleep posture estimation approach

	Height range	Weight range	Gender	Number of data	
With cover	138–185	48–110	Male, female	Side	180
				Supine	265
				Total	445
Without cover	138–185	48–110	Male, female	Side	602
				Supine	569
				Total	1171

The number of collected data in each posture and the characteristics of subjects participated in the experiment are reported in this table

**Fig. 13** The rank of FFT of 2D cross-sectional features (sorted by their rank) utilizing T-test method

The features extracted from skeleton joint provided by Kinect SDK library is employed to compare with our proposed approach. The skeleton joints are widely utilized for posture and sleep posture estimation. For example, in Lee et al. (2015) these skeleton joints are utilized to detect motion and sleep posture. In Mongkolnam et al. (2017), the skeleton joints are utilized to detect the position and situation in the bed along with pressure sensors. In Manzi et al. (2016), the skeleton joints are employed to distinguish different postures including the sleep posture. Therefore, the skeleton joints are utilized to provide a comparison with our method. The similar approach as utilized in Manzi et al. (2016) is employed to normalize the captured skeleton joints by Kinect SDK. Two different experiments are carried out to evaluate the proposed approach and compared with the result of features extracted from skeleton joints. In the first experiment, shown in the Table 2, both training and testing data are captured from one subject. In many cases, the monitoring system is designed to monitor a specific subject in these cases the analysis of the result when the training and Testing dataset is captured from one subject is important.

The result of this experiment is shown in the Table 2. The experiment is performed on four different subjects among male and females with different body mass. The gender, height, and weight of each subject are reported in the table as an indicator of the body mass. In addition, the total number of captured data of each subject is written in the table, for each set of data (with and without cover) 30% of the whole depth map is utilized to train SVM and 70% are utilized to test. The SVM classifier is based on finding the best hyperplane to separate the points of one class from those of other classes. The best hyperplane provides the largest margin between the points of classes. For some sample points, the different kernel for the SVM hyperplane can be employed to improve the classifier separation accuracy. One of the possible kernels is radial basis function (RBF) which uses a Gaussian function. The experiment is conducted over three different datasets for each subject. The first dataset is collected when the subjects don't use any blanket, the second one is captured when the subjects are covered by a thin blanket, and the final dataset is the union of both datasets. The SVM test and train have been run over 1000 times to provide the average accuracy. The selected SVM classifier kernel is RBF with sigma equal to 0.67 and in all cases, the proposed approach is also compared with skeleton joints features. As it can be seen the proposed approach has a better result in most of the cases without the need of extracting the skeleton joints in advanced which is a challenging task. The result in the Table 2 shows the comparable accuracy of two approaches. However, our proposed approach doesn't require any preliminary learning stage in order to detect the joints associated with the skeleton model. In addition, the sensor needs to be located directly above the subject in order to be able to detect all of the points of the human body in skeleton joints approach.

In the second experiment, the depth maps captured from 14 subjects are employed to train and test the SVM classifier. 30% of the collected data is utilized for training and 70% is utilized for testing. Table 3 shows the result of this experiment. Like the first scenario (shown in Table 2), the experiment is conducted over three datasets (data of sleeping with cover, without cover and union of both datasets) over 1000 run. The results are shown in the first row of Table 3. As it can be seen the results of the proposed approach is better than the approach using skeleton joints. In fact, using FFT of the signals is more robust to estimate the sleep posture of various body mass. In addition, the results of SVM training and testing on datasets consisting of samples associated with blanket cover case, shows reduction in the accuracy. This reduction was expected since the blanket cover creates an occlusion and hence reduction in detecting details of the subject's body. The second row of Table 3 shows the time utilized for training in a machine with CPU of core i7, 3.3 GHz and 16 GB of ram, the third row shows the number

**Table 2** Accuracy of the proposed approach when the testing and training dataset are collected from one single subject

Test and training sets are from one subject	Gender	H <sup>1</sup>	W <sup>2</sup>	Posture detection accuracy percentage							
				Without cover			With cover			Combined	
				P <sup>3</sup>	S <sup>4</sup>	# <sup>5</sup>	P <sup>3</sup>	S <sup>4</sup>	# <sup>5</sup>	P <sup>3</sup>	S <sup>4</sup>
Subj1	Female	165	59	100	100	337	99.32	96.66	135	98.86	96.76
Subj2	Male	183	75	100	99.08	36	100	96.89	27	100	97.14
Subj3	Male	175	90	100	99.86	66	98.44	98.38	50	99.01	99.21
Subj4	Female	170	48	100	100	95	98.44	96.28	78	99.21	97.91

The result is shown for different subject with different body mass. The experiment for each subject has been carried out when the subject is covered with blanket, without blanket and both. The first and second subjects are considered as normal body mass among men and women. The third subject has higher body mass and forth subject has a lower body mass

<sup>a</sup>Height (cm)

<sup>b</sup>Weight (kg)

<sup>c</sup>The proposed method

<sup>d</sup>The method using features extracted from Skeleton point (Manzi et al. 2016)

<sup>e</sup>Number of data 70% is utilized for testing and 30% for training over 1000 run

of data utilized for training. It should be mentioned, however, our proposed approach needs more time for training it does not need any training phase for extracting skeleton joints.

In the third experiment, the training datasets (with cover, without cover and union of both) is collected from only one female subject with normal boy mass (subject one in Table 2) while the test data is collected from different subjects (13 subjects). The details of the datasets are reported in Table 4a. Table 4b–d show the confusion matrices of this experiment with using the proposed approach and skeleton joints approach when data is collected with no cover, with cover, and union of both datasets, respectively. As it can be seen the proposed approach shows a significant improvement over skeleton joints approach. This experiment shows that our approach is robust under different conditions including various body mass and cloths, different sensor positions and tilt. As it can be seen, the frequency elements of scans provide better interpretation of depth images during sleep since it extracts the level of smoothness of the posture which is the main factor in estimating the correct sleep posture.

### 4.2 Hands and legs positions

Similarly, SVM is utilized to evaluate the performance of the proposed approach to estimate more detailed posture configurations. These configurations are illustrated in Fig. 2. As mentioned in Sect. 2, the configuration of hands and legs is the distinguished elements to determine these sleep postures. Three different configurations are defined for legs and four different configurations for hands. The collected data belongs to an individual who asked to sleep in different postures without cover. Table 5a shows the number of data collected for each posture. In this part, four classes are considered to estimate hands position and three classes to estimate the legs position. We also take advantage of the T-test method to rank available features.

Table 5b shows the confusion matrix of the proposed approach for estimating the position of hands. In this table, the entry of hand1 refers to the first hands configuration (Fig. 12a), hand2 refers to the second hands configuration (Fig. 12b), hand3 is the third hands configuration (Fig. 12c), and hand4 is the fourth hands configuration (Fig. 12d).

**Table 3** The accuracy of the proposed approach and the skeleton joint approach

Test and training sets are among different subjects	Without cover		With cover		combined	
	Proposed approach	Skeleton joint	Proposed approach	Skeleton joint	Proposed approach	Skeleton joint
	98.75%	96.19%	96.24%	96.14%	97.96%	95.60%
Training time (s)	0.0711	0.0149	0.0284	0.0097	0.0942	0.0174
Number of training data	351	351	133	133	484	484

Results are reported for dataset containing data of 14 different individuals in conditions when no cover is used, when the subject is covered by a thin blanket and combination of both datasets. The second row shows the time for training and third row shows the number of data utilized for training in each scenario

**Table 4** Results of the experiment conducted when the training data belongs to only one subject and the test data is gathered from different subjects (13 more subjects) the confusion matrix for both cases of using the proposed approach and the skeleton joint features is provided

(a)	Postures	Without cover		With cover		
		TRAIN	TEST	TRAIN	TEST	
Number of data for train and test	Side	197	461	56	252	
	Supine	140	372	79	209	
(b) Without cover	Proposed method			Skeleton joint		
	Side	Supine	Result (%)	Side	Supine	Result (%)
Side	310	6	98.1	58	34	63.0
Supine	62	456	88.0	314	427	57.6
Results	83.3%	98.7%	91.5	15.6%	92.6%	58.2
(c) With cover	Proposed method			Skeleton joint		
	Side	Supine	Result (%)	Side	Supine	Result (%)
Side	194	20	90.7	197	134	59.5
Supine	58	189	76.5	54	75	58.1
Results	77.0%	90.4%	83.1	78.5%	35.9%	59.1
(d) Combined	Proposed method			Skeleton joint		
	Side	Supine	Result (%)	Side	Supine	Result (%)
Side	520	58	90.0	189	120	61.2
Supine	104	613	85.5	434	550	55.9
Results	83.3%	91.4%	87.1	30.3%	82.1%	57.2

(a) Numbers of test and train data. (b) The confusion matrix of the sleep posture without cover. (c) The confusion matrix of the sleep posture with cover. (d) The confusion matrix of the sleep posture using the combination of both datasets

Table 5c shows the confusion matrix on estimating the position of legs. Similarly, the first leg position is shown in the table by leg1, second by leg2, and third by leg3 (Fig. 11). Results in this table show that the position of hands and

legs are estimated accurately by the proposed approach. In estimating the position of hands, only one false positive is reported for hand position 2. Results for leg positions show that the main confusion was between leg positions 2 (log

**Table 5** The confusion matrix of the proposed approach to estimate the position of hands and legs

(a) Posture	Foetus	Free faller	Log	Soldier	Starfish	Yearner
Number of data	41	74	141	145	108	103
(b)	Hand1	Hand2	Hand3	Hand4	Result (%)	
Hand1	102	1	0	0	99.0	
Hand2	0	100	0	0	100	
Hand3	0	0	99	0	100	
Hand4	0	0	0	128	100	
Results	100%	99.0%	100%	100%	99.8	
(c)	Leg1	Leg2	Leg3	Result (%)		
Leg1	29	1	0	100.0		
Leg2	0	221	0	100		
Leg3	0	8	171	95.5		
Results	100%	96.5%	100%	98.1		

(a) The number of data in dataset for each posture mentioned in Fig. 2. (b) The confusion matrix of hand positions. (c) The confusion matrix of legs positions

posture, which is a side posture) and 3 (soldier, start fish and free faller which are supine posture). This suggests that such prediction can be improved by combining the results with the main postures estimation.

## 5 Conclusions

This paper proposes a novel approach to address the sleep posture estimation problem. The most informative features of the sleep postures are extracted by performing a fast Fourier transformation to the selected 2D cross-section depth map of the body. Then the extracted features are ranked by T-test method to train a SVM to classify postures into two main groups of side and supine with and without the presence of a blanket. The performance of the proposed approach to estimate more detailed postures is shown. The more details are provided by using more cross-sectional postures to determine the position of hands and legs.

Employing 2D cross-sectional scans can provide complementary features in other applications too. As it was shown in this work, they can produce distinctive features to recognize the different parts of the human body which makes it a reasonable approach to estimate a wider range of human postures in a daily life.

## References

- Barbara P, Ancoli-Israel S (2001) Sleep disorders in the elderly. *Sleep Med* 2:99–114
- Bixler EO, Vgontzas AN, Have TT, Tyson K, Kales A (1998) Effects of age on sleep apnea in men: I. Prevalence and severity. *Am J Respir Crit Care Med* 157:144–148
- Bodyframe Class (2014) Microsoft. <https://docs.microsoft.com/en-us/previous-versions/windows/kinect/dn772770%28v%3dieb.10%29>. Accessed 2 Dec 2016
- Boulay B, Brémond F, Thonnat M (2006) Applying 3d human model in a posture recognition system. *Pattern Recognit Lett* 27:1788–1796
- Dal Mutto C, Zanuttigh P, Cortelazzo GM (2012) Time-of-flight cameras and Microsoft Kinect™. Springer Science & Business Media, Berlin
- Falie D, Ichim M, David L (2008) Respiratory motion visualization and the sleep apnea diagnosis with the time of flight (ToF) camera. In: *International Conference on Visualization, Imaging and Simulation*, WSEAS, pp 179–184
- Fujiyoshi H, Lipton AJ, Kanade T (2004) Real-time human motion analysis by image skeletonization. *IEICE Trans Inf Syst* 87:113–120
- Gordon S, Grimmer K, Patricia T (2007) Understanding sleep quality and waking cervico-thoracic symptoms. *Internet J Allied Health Sci Pract* 5:1–12
- Hao T, Xing G, Zhou G (2013) iSleep: unobtrusive sleep quality monitoring using smartphones. In: *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems*. ACM, 2013
- Karlsen K, Herlofson K, Larsen JP, Tandberg M (1999) Influence of clinical and demographic variables on quality of life in patients with Parkinson's disease. *J Neurol Neurosurg Psychiatry* 66:431–435
- Lachat E, Macher H, Mittet M, Landes T, Grussenmeyer P (2015) First experiences with Kinect v2 sensor for close range 3D modelling. *Int Arch Photogramm Remote Sens Spatial Inf Sci* 40:93
- Lee J, Hong M, Ryu S (2015) Sleep monitoring system using kinect sensor. *Int J Distrib Sens Netw* 11:205
- Liu JJ, Xu W, Huang M-C, Alshurafa N, Sarrafzadeh M, Raut N, Yadegar B (2013) A dense pressure sensitive bedsheet design for unobtrusive sleep posture monitoring. In: *2013 IEEE international conference on pervasive computing and communications (PerCom)*. IEEE, pp 207–215
- Luštrek M, Kaluža B (2009) Fall detection and activity recognition with machine learning. *Informatica* 33:197–204
- Manzi A, Cavallo F, Dario PA (2016) Neural network approach to human posture classification and fall detection using RGB-D camera. In: *Italian Forum of Ambient Assisted Living*
- Menon A, Kumar M (2013) Influence of body position on severity of obstructive sleep apnea: a systematic review. *ISRN Otolaryngol* 2013:670381
- Metsis V, Kosmopoulos D, Athitsos V, Makedon F (2014) Non-invasive analysis of sleep patterns via multimodal sensor input. *Pers Ubiquitous Comput* 18:19–26
- Mongkolnam P, Booranrom Y, Watanapa B, Visutarrom T, Chan JH, Nukoolkit C (2017) Smart bedroom for the elderly with gesture and posture analyses using Kinect. *Maejo Int J Sci Technol* 11:1
- Oksenberg A, Silverberg DS (1998) The effect of body posture on sleep-related breathing disorders: facts and therapeutic implications. *Sleep Med Rev* 2:139–162
- Oksenberg A, Dynia A, Nasser K, Gadoth N (2012) Obstructive sleep apnoea in adults: body postures and weight changes interactions. *J Sleep Res* 21:402–409
- Panini L, Cucchiara R (2003) A machine learning approach for human posture detection in domotics applications. In: *Proceedings of the 12th international conference on image analysis and processing*, 2003. IEEE, pp 103–108
- Pellegrini S, Iocchi L (2007) Human posture tracking and classification through stereo vision and 3D model matching. *EURASIP J Image Video Process* 2008:476151
- Shotton J et al (2013) Real-time human pose recognition in parts from single depth images. *Commun ACM* 56:116–124
- Sleep Position Gives Personality Clue (2003) BBC. <http://news.bbc.co.uk/2/hi/health/3112170.stm>. Accessed 20 Sept 2016
- Sutherland K, Cistulli PA (2015) Recent advances in obstructive sleep apnea pathophysiology and treatment. *Sleep Biol Rhythms* 13:26–40
- Theodoridis S, Koutroumbas K (1999) *Pattern recognition*. Academic Press, New York
- Torres C, Hammond SD, Fried JC, Manjunath B (2015) Sleep pose recognition in an ICU using multimodal data and environmental feedback. In: *International conference on computer vision systems*. Springer, pp 56–66
- Torres C, Fragoso V, Hammond SD (2016) Eye-CU: sleep pose classification for healthcare using multimodal multiview data. In: *2016 IEEE Winter conference on applications of computer vision (WACV)*
- Wachs J, Goshorn D, Kölsch M (2009) Recognizing human postures and poses in monocular still Images. *IPCV* 9:665–671
- Xiao Z, Mengyin F, Yi Y, Ningyi L (2012) 3D human postures recognition using kinect. In: *2012 4th international conference on intelligent human-machine systems and cybernetics (IHMSC)*. IEEE, pp 344–347

- Yang C, Mao Y, Cheung G, Stankovic V, Chan K (2014) Graph-based depth video denoising and event detection for sleep monitoring. In: 2014 IEEE 16th international workshop on multimedia signal processing (MMSP). IEEE, pp 1–6
- Yao K-W, Cheng C (2008) Relationships between personal, depression and social network factors and sleep quality in community-dwelling older adults. *J Nurs Res* 16:131–139
- Yoshino A, Nishimura H (2016) Study of posture estimation system using infrared camera. In: International conference on human–computer interaction. Springer, pp 553–558

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.