Unobtrusive Sleep Monitoring Using Movement Activity by Video Analysis

Yuan-Kai Wang [1,2,*] **, Hung-Yu Chen** [1] **and Jian-Ru Chen** [1]

[1] Graduate Institute of Applied Science and Engineering, Fu Jen Catholic University,
    New Taipei 24205, Taiwan
[2] Electrical Engineering, Fu Jen Catholic University, New Taipei 24205, Taiwan
* Correspondence: ykwang@fju.edu.tw

check for
updates

Abstract: Sleep healthcare at home is a new research topic that needs to develop new sensors, hardware and algorithms with the consideration of convenience, portability and accuracy. Monitoring sleep behaviors by visual sensors represents one new unobtrusive approach to facilitating sleep monitoring and benefits sleep quality. The challenge of video surveillance for sleep behavior analysis is that we have to tackle bad image illumination issue and large pose variations during sleeping. This paper proposes a robust method for sleep pose analysis with human joints model. The method first tackles the illumination variation issue of infrared videos to improve the image quality and help better feature extraction. Image matching by keypoint features is proposed to detect and track the positions of human joints and build a human model robust to occlusion. Sleep poses are then inferred from joint positions by probabilistic reasoning in order to tolerate occluded joints. Experiments are conducted on the video polysomnography data recorded in sleep laboratory. Sleep pose experiments are given to examine the accuracy of joint detection and tacking, and the accuracy of sleep poses. High accuracy of the experiments demonstrates the validity of the proposed method.

Keywords: sleep pose recognition; keypoints feature matching; Bayesian inference; near-infrared images; scale invariant feature transform

1. Introduction

Sleep disorders induce irregular sleeping patterns and sleep deprivation that have serious impacts on health. Obstructive sleep apnea (OSA) [1] is one of the most well recognized sleep disorders. OSA is characterized by repetitive obstruction of the upper airways during sleep, resulting in oxygen de-saturation and frequent brain arousal. It is a symptom that not only decreases sleep quality by sleep disturbance, but also has severe influence which may be life-threatening. Reduction in cognitive function, cardiovascular diseases, stroke, driver fatigue and excessive day time sleepiness are common among OSA patients.

Sleep monitoring systems [2] are an important objective diagnosis method to assess sleep quality and identify sleep disorders. They provide quantitative data about irregularity of brain and body behaviors in sleeping periods and duration. This information helps the analysis of sleep-wake state, diagnosis of the severity of disorders, and prompt treatment of sleep-related diseases.

Polysomnography (PSG) is a standard diagnostic tool in sleep medicine [3] that measures a wide range of biosignals, including blood oxygen, airflow, electroencephalography (EEG), electrocardiography (ECG), electromyography (EMG) and electro-oculography during sleep time. The subject monitored by PSG has to sleep by carefully wearing a lot of sensors with numerous electrodes attached to his whole body, which increases sleep disturbance and discomfort.

A sleep technician has to read these overnight data and mark sleep status manually according to the standardized scoring rule. It is a complex, costly and labor-intensive instrument that is absolutely not adequate for sleep care at home.

Alternative approach with minimized contact-based sensors has been developed [3,4]. Unconstrained portable biosensors, accelerometer, RFID (Radio Frequency IDentification), pressure sensors and smartphone are individually-applied or combined to detect sleep behaviors such as limb movement, body movement and sleep position. Detected sleep behaviors are further analyzed to inference sleep-wake pattern and assess sleep disorder. Actigraphy is a good representative that is commonly used with a watch-like accelerometer device attached typically to the wrist. These alternatives apply contact-based sensors and are disadvantageous to the data acquisition and sleep quality.

The noncontact approach usually employs imaging sensors such as microwave, thermal imaging and near-infrared (NIR) camera to non-intrusively detect sleep behaviors. Among these imaging modalities, NIR camera is more desired for home sleep analysis because it is low-cost, easily accessible and highly portable. Sivan et al. (in 1996) and Schwichtenberg et al. (in 2018) [5,6] shows that NIR video alone is enough for effective screening of OSA. Manual analysis from NIR videos can result in high correlation to PSG based diagnosis. Automatic video analysis is then proposed for sleep-wake detection [7–10] and sleep behavior analysis [11–16]. The NIR video is analyzed to extract subject's motion information by the methods such as background subtraction, optical flow, image difference and edge filter. Classification methods are then employed to inference sleep states and body parts. However, robust methods for stable analysis remains a necessary concern.

Challenges of robust sleep video analysis come from the characteristics of NIR videos and sleep pose. Capturing sleep behavior using NIR camera in dark environment has the problems of non-uniform illuminance and bad image quality. NIR camera needs to actively project a near-infrared light source on objects. An unevenly illuminated image is usually obtained because the projected area is over-exposure but the other area is under-exposure. Moreover, NIR images have poor imaging quality because they has low resolution, low signal-to-noise ratio (SNR) and low contrast. In addition, noise is inevitably introduced to NIR images because of the low-light situation for sleep. Low contrast, high noise and uneven illumination degrade the distinctiveness of motion and edge features that are usually applied in existing studies. Therefore, extracting body movement from IR video suffers from the image degradation problem.

Sleep pose recognition from videos is highly challenging for the nonrigid characteristics of the human body. Deformable and partially occluded postures plus with irregular movements induce high variation and inconsistency of appearance and shape features of human body, and make it difficult to classify postures.

This paper proposes a nonintrusive method to recognize sleep body positions. The joint-based approach is adopted for our sleep pose recognition, including joint detection and pose recognition. The joint detection step finds body joints using keypoints extraction and matching, and builds human models from these joints. The tracking step updates the human model by an online learning approach to consider the constraint of sequential change of viewpoint of joints. The recognition step includes a Bayesian network and statistical inference algorithm to deal with the problems of self-occlusion and pose variations. A special scheme is developed for the detection of joints to conquer poor image quality and self-occlusion problems. A special design, called IR-sensitive pajamas, which attaches visual markers on human joints is proposed. Joint markers contain visual patterns distinguished with each other and each marker corresponds only to one specific joint. A human joint model is constructed from the detected joint markers within a sleep image and is recognized to the sleep pose being supine or lateral. An earlier version of this paper appeared in [11]. Compared to the prior paper, this study contains (1) a substantial number of additional survey, explanations and analysis, and (2) various additional experiments to investigate the impact an accuracy of infrared image enhancement for sleep analysis.

The rest of this paper is organized as follows. Section 2 reviews state-of-the-art sleep behavior analysis with contact- and noncontact-based approaches. Section 3 describes our method that includes three components: NIR image enhancement, joint detection and tracking, and sleep pose recognition. Section 4 presents the experimental evaluation compared with different methods. Finally, Section 5 concludes the characteristics and advantages of the proposed method.

2. Related Works

The medical foundation of sleep analysis from videos is the physiological interpretation of body movement during sleep, that is first reviewed and explained. Existing contact and noncontact methods for sleep pose recognition are also discussed. The final subsection introduces general posture and action recognition methods been developed in computer vision, and identify the importance of feature representation for detection and recognition problems.

2.1. The Relation between Sleep Behavior and Sleep Disorder

Several studies of sleep medicine [5,17] show that body movements are an important behavioral aspect during sleep. They can be associated to sleep states and connected to the sleep state transitions. In particular, major body movements have been described in non-rapid eye movement (NREM) sleep, whereas small movements predominantly occur during REM phases. Therefore, the frequency and duration of body movements are important characteristics for sleep analysis. Moreover, some study [18] has found correlations between sleep body positions and sleep apnea. There is also some clinical evidence that the analysis of body pose is relevant to the diagnosis of OSA. A sleep monitoring system that detects body positions during sleep would help the automatic diagnosis of OSA.

Body positions in sleep posture can be classified into supine, lateral and prone positions. Supine and lateral positions are dominant postures in adults, but children and infants may have prone positions [19]. Many patients with OSA exhibit worsening event indices while supine. Since supine OSA has been recognized to be the dominant phenotype of the OSA syndrome [20], objective and automatic position monitoring becomes very important because of the unreliability of patient's self-report of positions [21,22].

2.2. Contact Sensors for Sleep Analysis

The screening of sleep is typically obtrusive using contact sensors, like biosensors, accelerometers, wrist watches, and headbands [5]. Contact sensors have to be worn on user's body, have possible skin irritations and may disturb sleep. Saturation signal of oxygen in the arteries obtained through oximetry measurement or pulse oximetry is very effective for OSA diagnosis. Another attempt has been carried out by considering thoracic and abdominal signals, e.g., the respiration waveform. The use of EEG and ECG signals for the detection of sleep-wake and breathing apnea is a well-known standard. Actigraphy is a commonly used technique for sleep monitoring that uses a watch-like accelerometer based device attached typically to the wrist. The device monitors activities and later labels periods of low activity as sleep. It is a headband that users need to wear each night so that it can detect sleep patterns through the electrical signals naturally produced by the brain. The pressure sensitive mattress is an interestingly alternative of contact-based approach to identify occurrence of sleep movements [23]. It can monitor change in body pressure on the pad to detect movements. The main advantage of the pressure sensitive mattress is that users do not need to wear any device. But, it is a high-cost device and in some cases it may be uncomfortable to sleep on the pad and thus, they can affect sleep quality.

Sleep pose can be recognized by a contact-based approach [24]. Accelerometer [25], RFID [26], and pressure sensors [27] are used to acquire raw motion data of human, and inference sleep state from the motion data. While these methods are appropriate for sleep healthcare, their raw motion data is insufficient for accurate classification of sleep pose. More abundant data of human motions

acquired by imaging sensors with both spatial and temporal information should greatly benefit sleep pose recognition.

2.3. Noncontact Sensors for Sleep Behavior Analysis

Noncontact approach usually employs imaging sensors to noninvasively detect sleep behaviors. The methods with microwave [28] and thermal imaging system [29] are advantageous to see through the bed covering to detect body movement. However, these imaging modalities are expensive, not portable, and is unable to perform long-time nocturnal video recording. More works adopted NIR camera [30,31] and computer vision techniques to extract body movement and chest respiration. The use of a NIR video camera to analyze movement patterns of a sleeping subject promises an attractive alternative.

Sleep posture has been analyzed in some studies. The system [12] detects posture change of the subject in bed by observing chest or blanket movement, and uses optical flow to evaluate the frequency of posture change. Wang and Hunter [14,15] addressed the problem of detecting and segmenting the covered human body using infrared vision. Their aim was to specifically tailor their algorithms for sleep monitoring scenarios where head, torso and upper legs may be occluded. No high-level posture is automatically classified. Yang et al. [13] proposed a neural network method to recognize the sleep postures. However, only edges detected by linear filters are applied as movement features of human body for the neural classifier, and recognition results were not satisfied. Liao and Kuo [16] proposed a video-based sleep monitoring system with background modeling approaches for extracting movement data. Since a NIR camera is often employed to monitor night-time activities in low illumination environment without disturbing the sleeper, the image noise can be quite prominent especially in regions containing smooth textures. They investigated and modeled the lighting changes by proposing a local ternary pattern based background subtraction method. However, none of these methods recognize positions of sleep body, especially supine and lateral positions that are strongly related to OSA. High-level postures can be robustly recognized by the modeling of articulated human joints.

Motion and texture features used for sleep pose recognition are sensitive to image degradation. Enhancing images before feature extraction [32] by recovering illumination, reducing noise and increasing contrast, can greatly improve the distinctness of features and increase detection and recognition accuracy [33]. Nonlinear filters, for example Retinex, have been successfully applied on many computer vision tasks [34] instead of vision based sleep analysis.

2.4. Pose Recognition by Computer Vision

Computer vision-based methods have been proposed for human pose recognition as a non-intrusive approach to capture human behavior with broad applications ranging from human computer interfaces, video data mining, automated surveillance, and sport training to fall detection. Appearance-based approach that directly adopts color, edge and silhouettes features for the classification of human poses has been widely applied [35]. However, nonrigid variations produced by limbs movement and body deformation make it difficult to the classification by appearance features. Joint-based approach [36] on the other way builds a structured human model by detected joints and/or body parts from low-level features, and then recognize poses by the structured human model. This approach tolerates not only high deformation but also self-occlusion issues in human poses.

Sleep poses are also highly deformable and partial occlusion often occurs. They should be tackled by joint-based approach [14]. Apart from partial occlusion, it is likely that limited motion information is available from partial and irregular movements, which seriously affects the usability of traditional feature extraction methods.

Local features such as scale invariant feature transform (SIFT) [37] and Speeded Up Robust Features (SURF) [38] are a new design scheme for salient feature representation of human postures. Its success on numerous computer vision problems has been well demonstrated [39]. Here we give

only some examples of applying SIFT on action recognition. Scovanner et al. [40] proposed a 3D SIFT feature to better represent human actions. Wang et al. and Zhang et al. [41,42] applied SIFT flow to extract keypoint-based movement features between image frames and obtain better classification results than the features such as histogram of oriented gradients and histogram of optical flows. Local features have superior representability not only for visible-band images but also for NIR images. The paper [43] gave a comprehensive study of the robustness of four keypoint detectors for the three spectra: near infrared, far infrared and visible spectra. Robustness are demonstrated that performance of the four detectors is remaining for all the three spectra, although these interest point detectors were originally proposed for visual-band images. While these papers show the advantages of local features, it is still not easy to get articulated structure of human joints by local features.

In summary, sleep behavior is important and can be analyzed by non-contact approaches to achieve unobtrusive methods. NIR video analysis for sleep posture is challenging but innovative. A summary table of the reviews from related works is given in Table 1.

Table 1. Summary of reviews from related works.

Critical Points	Arguments
Sleep behavior is important to OSA diagnosis	• Body movements are an important behavior for sleep diagnosis • Body position is correlated to sleep apnea • Sleep behavior including body movements and body positions are relevant to OSA • Supine position is a critical posture
Non-contact and unobtrusive analysis are advantageous but have challenges	• Contact sensros have been well developed for sleep diagnosis • Non-contact imaging sensors are unobtrusive and better than contact sensors • NIR is a good non-contact sensor with many advantages • Challenges of NIR video analysis for sleep analysis are nonuniform illumination processing and nonrigid body deformation
Motivation of this paper	• Retinex is a nonlinear filter that could be applied to improve the NIR videos with nonuniform illumination • Joint-based methods are an important approach to traditional posture recognition in computer vision • Keypoints methods such as SIFT could be applied to joint-based posture recogntion

3. Sleep Pose Recognition

The proposed method analyzes near-infrared videos to recognize sleep poses by body joint model. The near-infrared images are first enhanced by an illumination compensation algorithm to improve the quality of feature extraction. SIFT-based local features are employed to perform joint detection of human body. Poses are recognized with a Bayesian inference algorithm to solve the occlusion issue.

A novel idea is developed for the joint detection and modeling in our method. One basic idea is that it becomes usual to delicately customize some bedding materials to facilitate more accurate monitoring, such as the mattress pad sensor. Therefore, we propose an unobtrusive passive way by revamping pajamas with NIR sensitive fabrics around the joint positions of human body. The fabrics that is sensitive to NIR light source can reflect more lights and show high intensity values in NIR images. With the NIR-sensitive fabrics we get more visual information of body joints and are able to detect and recognize sleep poses. Figure 1 gives an illustration of our design. There are ten joints that are common for posture and action recognition. We make ten NIR-sensitive patches with special fabrics sewing on pajamas.
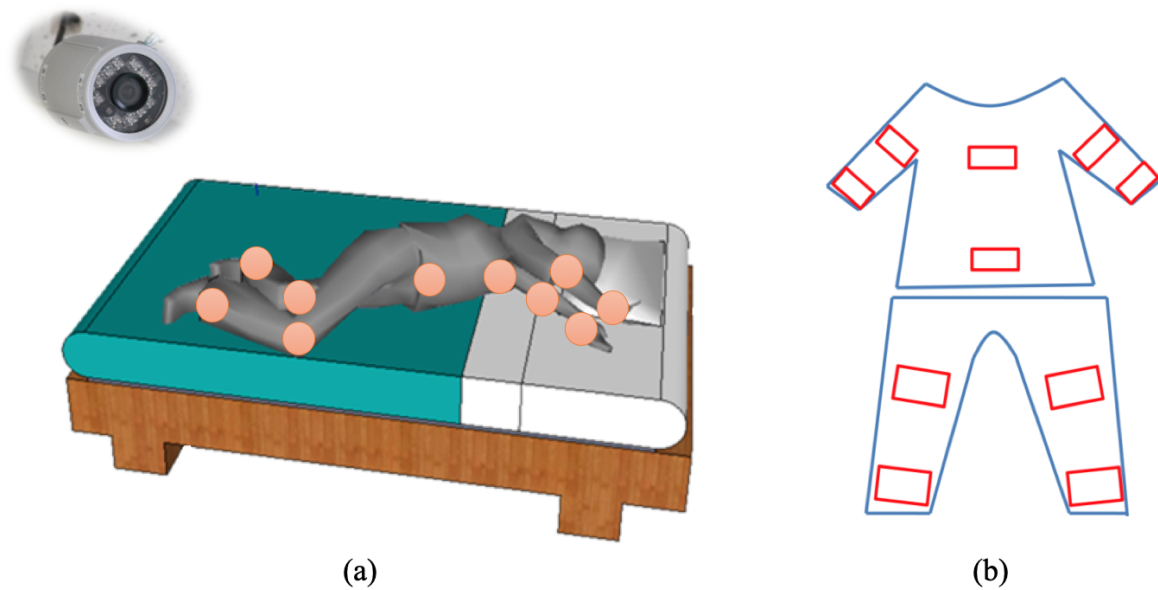
(a)   (b)

Figure 1. The proposed scheme for sleep pose recognition. (**a**) An infrared camera is used to acquire the sleep videos and analyze body joint positions. (**b**) Pajamas with ten near infrared (NIR)-sensitive patches are designed to facilitate the detection of body joints.

Another issue of the proposed method is to distinguish different joints. We apply the concept in augmented reality to design fiducial markers in order to not only reliably detect each joint, but also easily distinguish all joints. A fiducial marker supplements with image features originating in salient points of the image. However, we need to design a fiducial marker system that consists of a set of distinguishable markers to encode different joints. Desirable properties of the marker system are low false positive and false negative rates, as well as low inter-marker confusion rates. We adopt the markers designed in the paper [44] that derives the optimal design of fiducial markers specially for SIFT and SURF. Its markers are highly detectable even under dramatically varying imaging conditions.

3.1. Near-Infrared Image Enhancement

An illumination compensation algorithm [45] that includes single-scale Retinex and alpha-trimming histogram stretching is applied to enhance the NIR video. SSR is a nonlinear filter to improve lightness rendition of the images without uniform illumination. However, SSR does not improve contrast. The histogram stretching with alpha-trimming is then followed to improve contrast. Let I_t be an NIR image at time t and I'_t is the enhanced result, our image enhancement algorithm is a composition function of three successive steps as follows:

$$I' = m(f(R(I))), \ where \ R(I) = \log(I) - \log(G(c) \otimes I). \tag{1}$$

$R(\cdot)$ is the SSR, $f(\cdot)$ is the alpha-trimming histogram stretching, and $m(\cdot)$ is a median filter for denoising. The SSR function $R(\cdot)$ uses a Gaussian convolution kernel $G(c)$ with size c to compute the scale of illuminant, and enhance the image by log-domain processing. The alpha-trimming histogram stretching applies a gray-level mapping that extends the range of gray levels into the whole dynamic range. Median filter is applied to eliminate the shot noise in the NIR images. The effect of the enhancement is majorly influenced by the size of Gaussian kernel size c. Figure 2 illustrates the influence with different kernel sizes. The original image has uneven illumination. The image in Figure 2b is still dark and the contrast of human body is low. It also has white artifact around image boundary. Figure 2f has better illumination compensation effect, better contrast on human body, and less artifact.
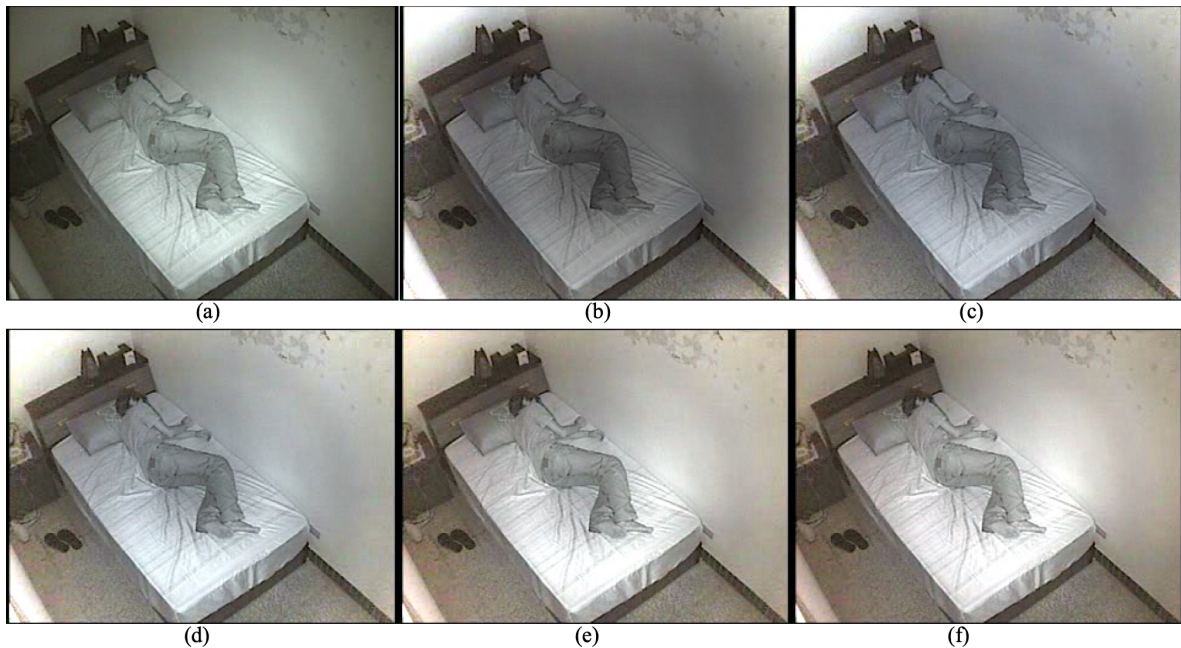
Figure 2. Enhanced NIR images with different Gaussian kernel sizes from 50 to 250. (**a**) Original image. (**b**) $c = 50$. (**c**) $c = 100$. (**d**) $c = 150$. (**e**) $c = 200$. (**f**) $c = 250$.

3.2. Detection and Tracking of Human Joints by Distinctive Invariant Feature

We propose a SIFT-based joint detection algorithm to detect joints in the first image, and a structured online learning to track those detected joints in the video. The SIFT analyzes an input image at multiple scales in order to repeatedly find characteristic blob-like structures independently of their actual size in an image. It first applies multiscale detection operators to analyze the so called scale space representation of an image. In a second step, detected features are assigned a rotation-invariant descriptor computed from the surrounding pixel neighborhood. The detail of applying SIFT to detect human joints is described in the following.

Given an enhanced image I'_t at time t and a set of joint markers M_i, $i = 1{\sim}10$, we apply SIFT to extract keypoint features of I'_t and M_i, and compute the correspondence set C_i to find all joints. The keypoints of I'_t is represented as a sparse set $X_t = \{x_t^j\}$ that is called image descriptor. The keypoints of M_i is a set $Y_i = \{y_i^k\}$ that is called a joint model descriptor. A correspondence set $C_i = \{(x_t^j, y_i^k) | s(x_t^j, y_i^k) > \tau\}$ is obtained for each joint marker M_i, where $s(\cdot, \cdot)$ is the matching score and τ is a matching threshold. The matching between X_t and Y_i is done by a best-bin-first search that returns the closest neighbor with the highest probability with pairwise geometric constraints. The set of matched joints $\widehat{J}$ is a set of joint coordinates J_i defined as follows:

$$\widehat{J} = \{J_i \mid J_i = median(C_i) \; and \; |C_i| > \alpha\} \tag{2}$$

where σ is a threshold of minimum number of matched keypoints. A matched joint requires enough matched keypoints, i.e., $|C_i| > \sigma$. The coordinates of the ith are calculated as the median coordinates from $x_t^j \in C_i$ to screen outlier of keypoint-based joint detection, which can reduce the distance error of joint. While the keypoint-based matching is robust to illumination, rotation and scale variance, some joints may not be detected because of occlusion. That is, some joints are not detected if $|\widehat{J}| < 10$.

An example of keypoint detection is shown in Figure 3a. Most detected keypoints are clustered together around joint markers, that demonstrates the image descriptor X_t is a salient representation of the set of joint markers. Figure 3b shows an example of joint detection. The joint marker is shown at the left top of the image. These lines represent matched keypoints between the joint marker and the visual marker in the pajamas.
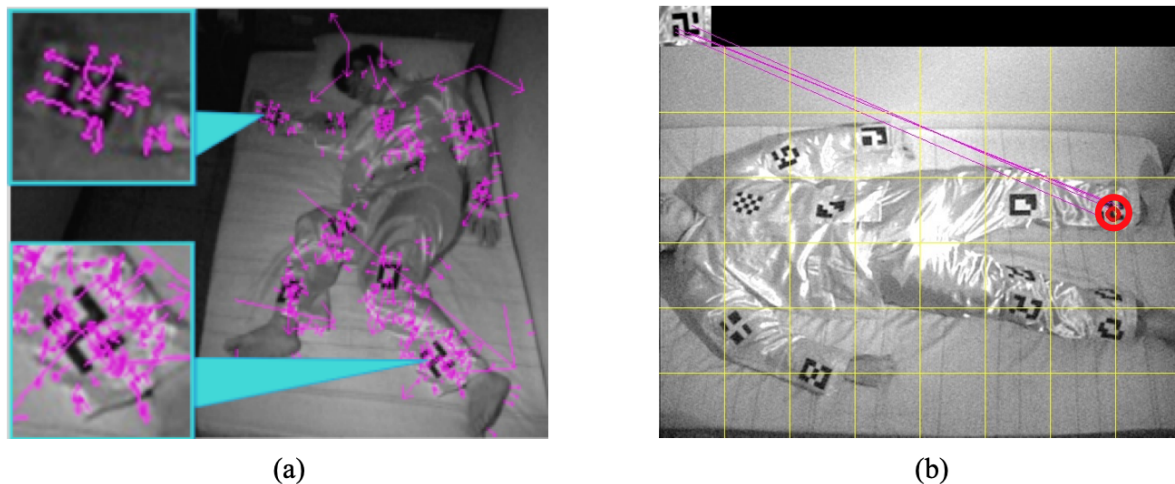
Figure 3. Keypoint detection and joint detection. (**a**) Image descriptor of an original lateral-pose image overlaid with keypoints of length and orientation information. Each arrow represents one detected keypoint. (**b**) A joint detection example for left ankle. It is detected by matching a specific joint marker representing the left ankle with a visual marker in the sleep image.

A structured output prediction with online learning method [46] is applied to perform adaptive keypoint tracking of the human joints. It is used for homography estimation of correspondence between frames and binary approximation of model. Structured output prediction is handled by applying RANSAC to find the transformation of the correspondence. To speed up the detection and description process, we adopt SURF instead of SIFT. This is achieved by not only relying on integral images for image convolutions but applying a Hessian matrix-based measurement for the detector and a distribution-based descriptor.

3.3. Sleep Pose Estimation by Bayesian Inference

The sleep pose model is mathematically formulated as a Bayesian network, and the estimation of pose class is achieved by probabilistically inferencing the posterior distribution of the network.

Bayesian network combining probability theory and graphical model for statistical inference is employed in this paper because of its great robustness capability with missing information. That is, pose class can be inferred even when there are undetected joints.

Our sleep pose model is a Bayesian network G with one root node representing the state of pose and ten child nodes J_i, $1 \leq i \leq 10$, corresponding to the states of joints. This network gives a probabilistic model to represent the causal relationship between sleep pose and joint positions: a given pose affects the positions of joints, and thus the pose can be inferred by Bayesian theory from a given set of joint positions. The property of conditional independence exists in this naïve model and is helpful for the inference of poses.

Let the undetected and detected joint sets be individually represented by U and $\widehat{J}$. The conditional posterior distribution of p given $\widehat{J}$ can be derived as follows:

$$P(p|J_1, \cdots, J_{10}) = \pi \sum_{J_i \in U} \prod_{J_i \in \widehat{J}} P(p|J_i), \tag{3}$$

where π is the prior probability $P(p, J_1, ..., J_{10})$ that is in the form of full joint probability. The estimated sleep pose $\widehat{p}$ is the maximum a posterior (MAP) estimate given by the following:

$$\widehat{p} = arg\ max\ P(p|J_1, \cdots, J_{10}), \tag{4}$$

Both approximate and exact inference algorithms can be applied for the MAP calculation of our sleep pose model. While approximate inference algorithms are usually more efficient than exact inference algorithms, our sleep pose model could be solved more efficiently with exact inference

approach. We employ clique-tree propagation algorithm [47] that first transforms the Bayesian model into undirected graph, then use query-driven message passing to update the statistical distribution of each node.

An example of sleep pose estimation is shown in Figure 4, which is a lateral pose with fully detected joints. The reconstructed human model is depicted with a cardboard representation.
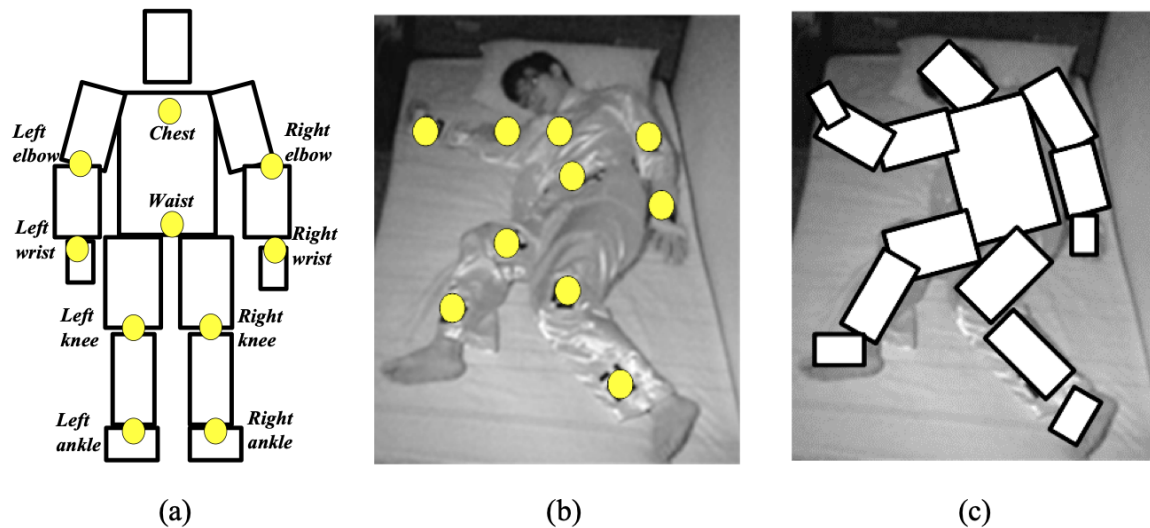

(a)   (b)   (c)

Figure 4. An example of reconstructed human model by fully detected joints. (**a**) A standard model with a standing pose. (**b**) A sleep image overlaid with the detected/tracked joints found by keypoint match and structured learning. (**c**) Reconstructed model of the lateral pose.

4. Experimental Results

We evaluated these components in our method: near-infrared image enhancement, joint detection and pose recognition. Accuracy and performance compared with existing methods are conducted and discussed. Two individual setups of experiments were established. The first setup was established to evaluate the efficacy of near-infrared image enhancement, whose accuracy was assessed by total sleep time (TST) measured from the sleep–wake detection results obtained after the enhancing of images. The second setup is built for joint detection and pose recognition. The detection of ten joints is assessed by pixel difference of joint positions, and the recognition of supine/lateral poses are assessed by classification precision.

The experimental setups were conducted in a sleep laboratory equipped with a PSG, near-infrared cameras and related instruments. The PSG was adopted to obtain ground truth. A full-channel PSG (E-Series, Compumedics Ltd., Melbourne, Australia) was employed to record overnight brain wave activity (electroencephalography from C3-M2, C4-M1), eye movements (left and right electrooculogram), electrocardiography, leg movements (left and right piezoelectric leg paddles), and airflow (oral/nasal thermister). The near-infrared video has the resolution of 640 * 480 pixels with the frame rate of 20 fps.

4.1. Effectiveness of Near-Infrared Image Enhancement

In this subsection, the assessment of image enhancement is carefully created from sleep experiments, because there is still no objective criteria to evaluate the quality of near-infrared images. Since our goal is to build a computer-vision based sleep evaluation system, it is reasonable to adopt a sleep quality criteria, TST, obtained in a sleep experiment to evaluate our image enhancement component. TST is the amount of actual sleep time including REM and NREM phases in a sleep episode; this time is equal to the total sleep episode less the awake time.

In this first sleep setup, eighteen subjects involved in the experiments are divided into two groups, including a normal group without a sleep disorder and an OSA group with sleep disorder. Table 2 gives statistics of the subjects in the two groups. From the data we confirmed that these subjects have only OSA symptom and have no PLM problems, because the means of body mass index and PLM are statistically identical, and the OSA group has higher RDI and lower sleep efficiency.

Table 2. Statistics of the two groups in the sleep-wake detection experiment. SD means standard deviation. RDI (respiratory disturbance index) is the number of abnormal breathing events per hour of sleep. PLM (periodic leg movements) is the number during nocturnal sleep and wakefulness. Sleep efficiency is the proportion of sleep in the episode, i.e., the ratio of total sleep time (TST) to the time in bed.

		Age	Body Mass Index	RDI	PLM	Sleep Efficiency
Normal group	Mean	42.63	24.39	4.23	1.28	91.44
	SD	14.37	4.18	2.44	1.61	4.7
OSA group	Mean	51.1	24.99	35.85	1.43	84.22
	SD	15.25	2.69	21.68	3.03	12.3

Both PSG data and IR videos were recorded overnight for each subject. Sleep stages were scored using the standard AASM criteria with thirty-second epochs. Obstructive apneas were identified if the airflow amplitude was absent or nearly absent for the duration at least eight seconds and two breaths cycles. A well-trained sleep technician was responsible for configuring the PSG for each subject. The ground truth of sleep state is manually labeled by the sleep experts working in hospital. The ground truth of the TST is obtained by the total sleep time calculated from the manual annotation.

Estimated TST is obtained from near-infrared videos by a process including the proposed image enhancement, followed by background subtraction and sleep-wake detection. We extract body movement information from the motion feature obtained from background subtraction. A statistical differencing operator is applied to measure the distance d_t between the enhanced infrared image I'_t and the previous background model B_{t-1} as follows:

$$d_t = |I'_t - B_{t-1}| \tag{5}$$

The background model B_{t-1} was recursively updated by the Gaussian mixture background subtraction method [48], and can successfully approximate the background of a given image which has a time-evolving model. The body movement BM_t at time t is obtained by accumulating the total motion pixels in d_t:

$$BM_t = \sum_x \sum_y M_t$$
$$where \ M_t(x,y) = \begin{cases} 1, \ if \ d_t(x,y) > threshold \\ 0, \ otherwise \end{cases} \tag{6}$$

Sleep activity feature was then calculated from BM_t by descriptive statistics such as mean, standard deviation and maximum, defined over an epoch of 30 s. Sleep activity is obtained by the sleep-wake detection algorithm that is modeled by the linear regression [49] of the sleep activity feature.

The proposed image enhancement algorithm is first qualitatively compared with four classical methods: histogram stretching, histogram equlization, gamma correction and single-scale Retinex, with regard to illumination uniformity and contrast. Figure 5b–d show the results of three global enhancement methods with raised image contrast, but the middle of these images is over-exposed. Figure 5e is the result of SSR filtering with uniform illumination but low image contrast. Figure 5f show the result of the proposed method that can solve both non-uniform illumination and low contrast issues.
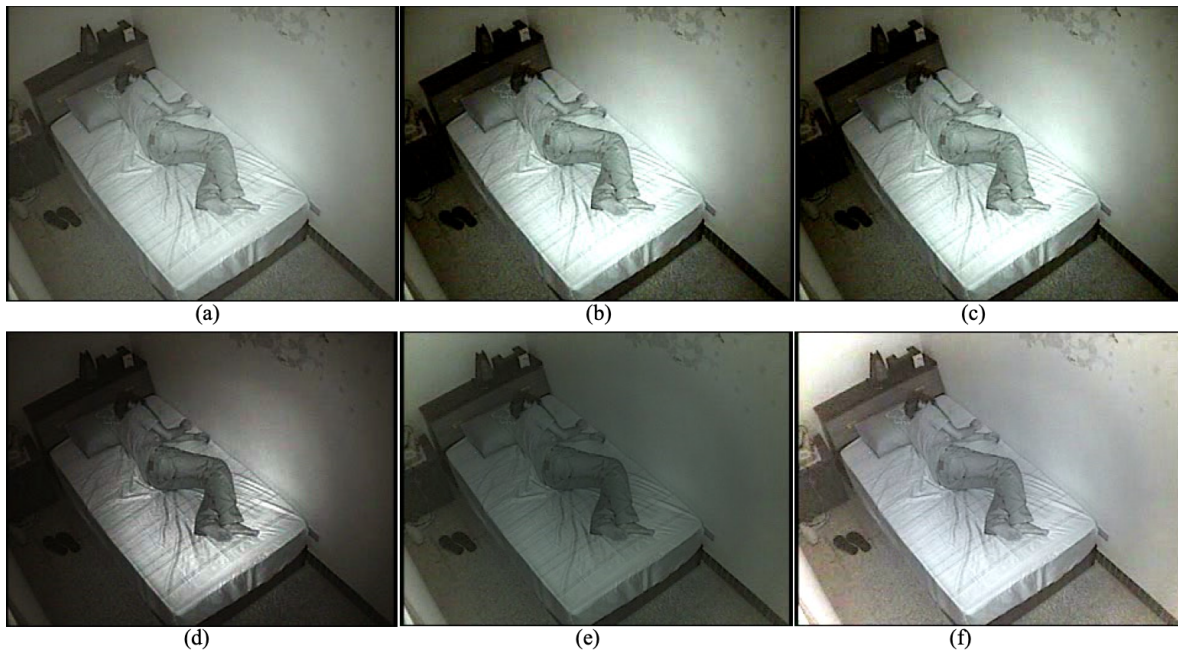
Figure 5. Enhancement of NIR images. (**a**) Original image. (**b**) Histogram stretching. (**c**) Histogram equalization. (**d**) Gamma correction. (**e**) Single-scale Retinex. (**f**) The proposed illumination compensation method.

Quantitative assessment of the proposed method was achieved by an index called TST error E_{TST}, that is defined as the normalized difference between estimated TST and ground truth of TST.

$$E_{TST} = \frac{|Estimated\ TST - Ground\ truth\ TST|}{Ground\ truth\ TST} \tag{7}$$

Its value is between [0,1]. Lower E_{TST} means better performance of our method.

The performance for a normal group with respect to three different sleep activity features, MA (moving average), MS (moving standard deviation), and MM (moving maximum), is shown in Figure 6. The three performances without image enhancement are not statistically differential. However, the performances with image enhancement are consistently lower than those without enhancement, and MA with image enhancement has the best performance.
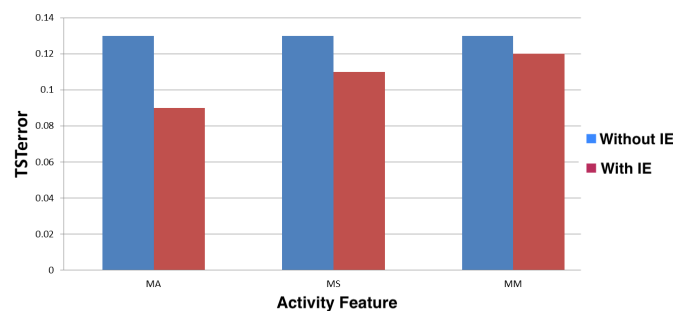

Figure 6. Effect of image enhancement (IE) for the performance improvement of the error of total sleep time.

The performance comparison of using image enhancement with respect to normal and OSA groups is shown in Table 3. MA is adopted as the only sleep activity feature in this comparison. Specificity (SPC) and negative predictive value (NPC) are also calculated for performance comparison. We can observe that while normal group has better performances with less standard deviation, OSA group still has good performances.

Table 3. Effect of image enhancement (IE) for the performance improvement of the error of total sleep time.

		E_{TST}	SPC	NPV
Normal group	Mean	0.09	0.95	0.91
	STD	0.16	0.04	0.10
OSA group	Mean	0.15	0.93	0.84
	STD	0.18	0.19	0.21

4.2. Evaluation of Pose Recognition

The second experiment has five subjects of various gender, height and weight wearing the custom pajamas and sleeping with free postures. Ten video clips were recorded with respect to the episodes of supine and lateral poses of various limb angles and occlusion. Half clips are randomly chosen for training of the classifier, and the remaining half for test. Ground truth of body positions and joint positions are manually labeled. The pose recognition in this experiment will not detect sleep and wake states. Therefore, background subtraction and sleep-wake detection algorithms are excluded, and the method in this experiment will include the three components in Section 3: image enhancement, joint detection and pose recognition. However, only the effectiveness of joint detection and pose recognition is validated here.

Figure 7 gives some example results of joint detection with respect to various poses. The positions of IR-markers are marked on the original image to evaluate the effectiveness of the proposed method. Some quantitative experiments are conducted to evaluate the localization accuracy of the proposed method. Figure 8a shows the effect of the threshold of matched keypoint numbers (σ) on detected joint numbers. High precision can be obtained with $\sigma = 1$. This result indicates that the design of joint markers is distinctive and so the joint localization does not need to match more numbers of keypoints. Figure 8b shows average detection rate of each joint position with mostly achieving high accuracy. The detection of right knee joint is a little not satisfactory because of perspective variations. Euclidean distance errors of each joint position is shown in Figure 8c. The average error of all joints is 6.57 pixels. The result indicates higher error in both ankle joints.
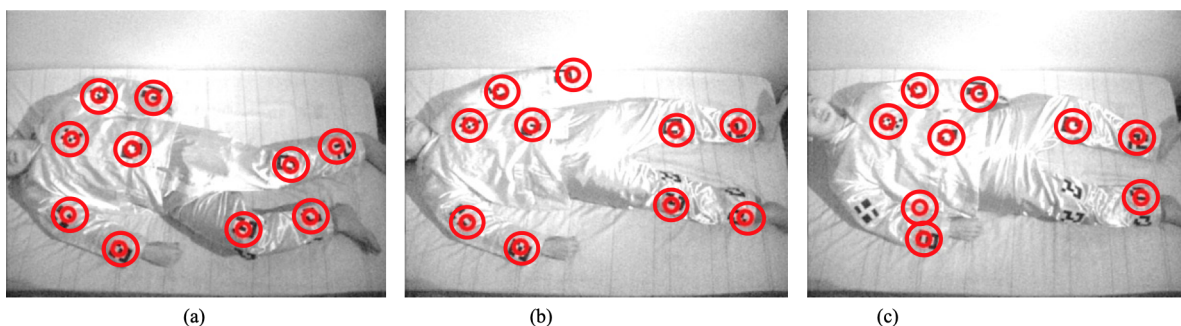

(a)　　　　　　　　　　　　(b)　　　　　　　　　　　　(c)

Figure 7. Examples of joint detection results. Each double-red circle represents a detected joint. (**a**) Lateral pose with ten successful detections. (**b**) Supine pose with ten successful detections. (**c**) Supine pose with eight successful detections. The right-knee and right-elbow joints are missed because of great distortion induced by perspective and cloth's wrinkles.
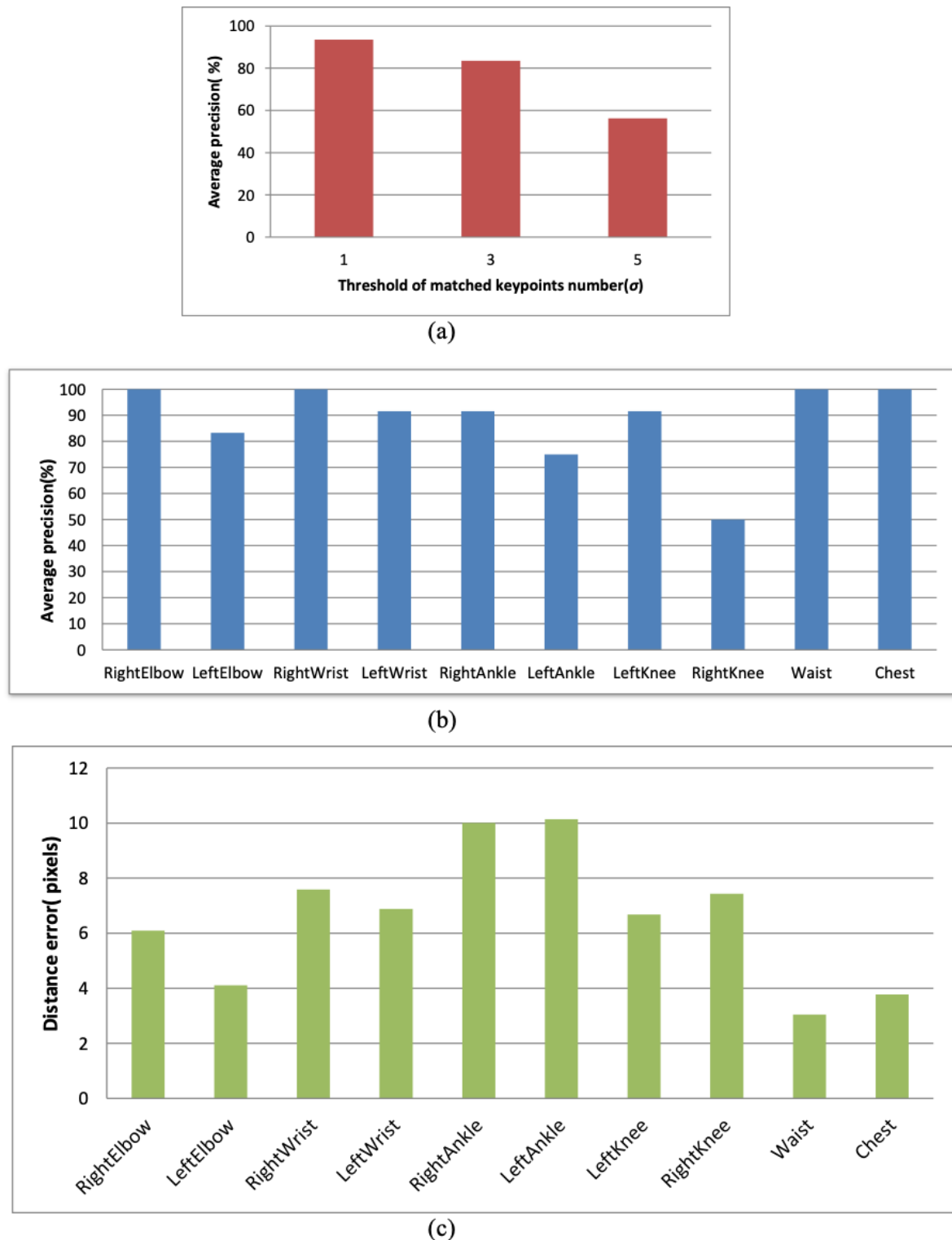
(a)

(b)

(c)

Figure 8. Accuracy evaluation. (**a**) The effect of the threshold of matched keypoint numbers. (**b**) Average precision of joint localization. (**c**) Average distance error of joint localization.

The performance of the proposed sleep pose model with clique-tree propagation inference was evaluated by comparing with five inference algorithms. Four are approximate algorithms: stochastic sampling, likelihood sampling, self-importance sampling, and backward sampling, and one exact algorithm: loop cutset conditioning. Figure 9 show execution and precision of the six algorithms. The result shows that clique-tree propagation algorithm takes the least time to achieve the best precision.

The positive predictive value, negative predictive value, specificity and sensitivity of the proposed method are 80%, 71%, 63% and 86%.
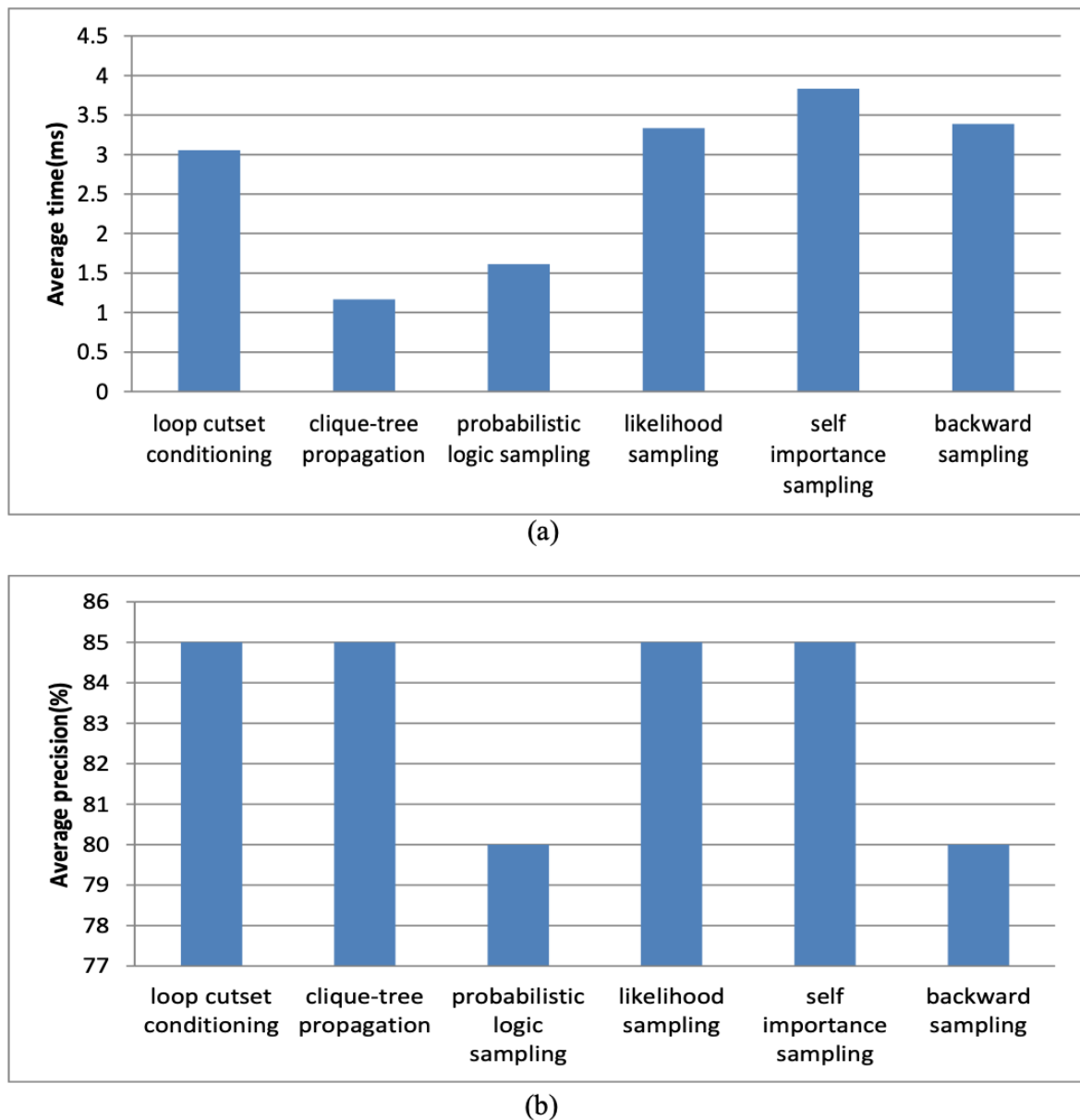

(a)

(b)

Figure 9. Comparison of six inference algorithms with respect to (**a**) execution time, and (**b**) accuracy.

The experimental results are compared with previous results: RTPose [14], MatchPose [15] and Ramanan [50]. The comparison is shown in Table 4. The previous results are obtained from their published publications. The results of our method were always superior than previous results except to right knee. Note that the accuracy of torso and head of our method are obtained from chest and waist, and the accuracy of ankles and knees in the table correspond to the lower legs and upper legs in previous publications.

Table 4. Comparisons with previous results. The numbers represent accuracy in percentage. N/A means the data are not available from their methods.

	Torso	Right Ankle	Left Ankle	Right Knee	Left Knee	Right Elbow	Left Elbow	Right Wrist	Left Wrist	Head
Ramanan	80	60	53	60	37	N/A	N/A	N/A	N/A	53
RTPose	93	N/A	N/A	70	80	N/A	N/A	N/A	N/A	80
MatchPose	97	45	69	75	80	N/A	N/A	N/A	N/A	94
Our	100	92	75	50	92	100	83	100	92	100

5. Conclusions

An automatic sleep pose recognition method from NIR videos is proposed in this paper. The method proposes keypoint-based joint detection and online tracking to persistently locate joint positions. Sleep poses are recognized from located joints by statistical inference with the proposed Bayesian network. Our method detects and tracks the human joints of great variations and occlusions with high precision. Experimental results validate the accuracy of the method for both supine and lateral poses. Further studies could incorporate more non-invasive sensors to develop a cheap and convenient sleep monitoring system for home care with long-term stability. With the reliable method proposed in this paper, an enhanced pajamas for home sleep monitoring can be further developed by costume designers.

Author Contributions: Conceptualization, Y.-K.W.; Funding acquisition, Y.-K.W.; Investigation, H.-Y.C.; Project administration, Y.-K.W.; Resources, Y.-K.W.; Software, J.-R.C.; Supervision, H.-Y.C.; Validation, H.-Y.C.; Visualization, H.-Y.C.; Writing—original draft, J.-R.C. and H.-Y.C.; Writing—review and editing, Y.-K.W.

References

1. Berry, R.B.; Budhiraja, R.; Gottlieb, D.J.; Gozal, D.; Iber, C.; Kapur, V.K.; Marcus, C.L.; Mehra, R.; Parthasarathy, S.; Quan, S.F.; et al. Rules for scoring respiratory events in sleep: Update of the 2007 AASM manual for the scoring of sleep and associated events. *J. Clin. Sleep Med.* **2012**, *8*, 597–619. [CrossRef] [PubMed]
2. Procházka, A. Sleep scoring using polysomnography data features. *Signal Image Video Process.* **2018**, *12*, 1043–1051.
3. Roebuck, A.; Monasterio, V.; Gederi, E.; Osipov, M.; Behar, J.; Malhotra, A.; Penzel, T.; Clifford, G.D. A review of signals used in sleep analysis. *Physiol. Meas.* **2014**, *35*, 1–57. [CrossRef] [PubMed]
4. Park, K.S.; Choi, S.H. Smart Technologies Toward Sleep Monitoring at Home. *Biomed. Eng. Lett.* **2019**, *9*, 73–85. [CrossRef] [PubMed]
5. Sivan, Y.; Kornecki, A.; Schonfeld, T. Screening obstructive sleep apnoea syndrome by home videotape recording in children. *Eur. Respir. J.* **1996**, *9*, 2127–2131. [CrossRef] [PubMed]
6. Schwichtenberg, A.J.; Choe, J.; Kellerman, A.; Abel, E.; Delp, E.J. Pediatric videosomnography: Can signal/video processing distinguish sleep and wake states? *Front. Pediatr.* **2018**, *6*, 158. [CrossRef] [PubMed]
7. Scatena, M. An integrated video-analysis software system de-signed for movement detection and sleep analysis. Validation of a tool for the behavioural study of sleep. *Clin. Neurophysiol.* **2012**, *123*, 318–323. [CrossRef] [PubMed]
8. Kuo, Y.M.; Lee, J.S.; Chung, P.C. A visual context-awareness-based sleeping-respiration measurement system. *IEEE Trans. Inf. Technol. Biomed.* **2010**, *14*, 255–265. [PubMed]

9. Cuppens, K.; Lagae, L.; Ceulemans, B.; Huffel, S.V.; Vanrumste, B. Automatic video detection of body movement during sleep based on optical flow in pediatric patients with epilepsy. *Med. Biol. Eng. Comput.* **2010**, *48*, 923–931. [CrossRef]

10. Choe, J.; Schwichtenberg, A.J.; Delp, E.J. Classification of sleep videos using deep learning. In Proceedings of the IEEE Conference on Multimedia Information Processing and Retrieval, San Jose, CA, USA, 28–30 March 2019; pp. 115–120.

11. Wang, Y.K.; Chen, J.R.; Chen, H.Y. Sleep pose recognition by feature matching and Bayesian inference. In Proceedings of the International Conference on Pattern Recognition, Stockholm, Sweden, 24–28 August 2014; pp. 24–28.

12. Nakajima, K.; Matsumoto, Y.; Tamura, T. Development of real-time image sequence analysis for evaluating posture change and respiratory rate of a subject in bed. *Physiol. Meas.* **2001**, *22*, 21–28. [CrossRef]

13. Yang, F.C.; Kuo, C.H.; Tsai, M.Y.; Huang, S.C. Image-based sleep motion recognition using artificial neural networks. In Proceedings of the International Conference on Machine Learning and Cybernetics, Xi'an, China, 5 November 2003; pp. 2775–2780.

14. Wang, C.W.; Hunter, A. Robust pose recognition of the obscured human body. *Int. J. Comput. Vis.* **2010**, *90*, 313–330. [CrossRef]

15. Wang, C.W.; Hunter, A.; Gravill, N.; Matusiewicz, S. Real time pose recognition of covered human for diagnosis of sleep apnoea. *Comput. Med. Imaging Graph.* **2010**, *34*, 523–533. [CrossRef] [PubMed]

16. Liao, W.H.; Kuo, J.H. Sleep monitoring system in real bedroom environment using texture-based background modeling approaches. *J. Ambient Intell. Humaniz. Comput.* **2013**, *4*, 57–66. [CrossRef]

17. Okada, S.; Ohno, Y.; Kenmizaki, K.; Tsutsui, A.; Wang, Y. Development of non-restrained sleep-monitoring method by using difference image processing. In Proceedings of the European Conference of the International Federation for Medical and Biological Engineering, Antwerp, Belgium, 23–27 November 2009; pp. 1765–1768.

18. Oksenberg, A.; Silverberg, D.S. The effect of body posture on sleep-related breathing disorders: Facts and therapeutic implications. *Sleep Med. Rev.* **1998**, *2*, 139–162. [CrossRef]

19. Isaiah, A.; Pereira, K.D. The effect of body position on sleep apnea in children. In *Positional Therapy in Obstructive Sleep Apnea*; Springer Nature Switzerland: Basel, Switzerland, 2014; Volume 14, pp. 151–161.

20. Joosten, S.A.; O'Driscoll, D.M.; Berger, P.J.; Hamilton, G.S. Supine position related obstructive sleep apnea in adults: Pathogenesis and treatment. *Sleep Med. Rev.* **2014**, *18*, 7–17. [CrossRef] [PubMed]

21. Russo, K.; Bianchi, M.T. How reliable is self-reported body position during sleep? *Sleep Med.* **2016**, *12*, 127–128. [CrossRef] [PubMed]

22. Ravesloot, M.J.L.; Maanen, J.P.V.; Dun, L.; de Vries, N. The undervalued potential of positional therapy in position-dependent snoring and obstructive sleep apnea—A review of the literature. *Sleep Breath.* **2013**, *17*, 39–49. [CrossRef] [PubMed]

23. Liu, J.J.; Xu, W.; Huang, M.C.; Alshurafa, N.; Sarrafzadeh, M.; Raut, N.; Yadegar, B. Sleep posture analysis using a dense pressure sensitive bedsheet. *Pervasive Mob. Comput.* **2014**, *10*, 34–50. [CrossRef]

24. Hossain, H.M.S.; Ramamurthy, S.R.; Khan, M.A.A.H.; Roy, N. An active sleep monitoring framework using wearables. *ACM Trans. Interact. Intell. Syst.* **2018**, *8*, 22. [CrossRef]

25. Foerster, F.; Smeja, M.; Fahrenberg, J. Detection of posture and motion by accelerometry: A validation study in ambulatory monitoring. *Comput. Hum. Behav.* **1999**, *15*, 571–583. [CrossRef]

26. Hoque, E.; Dickerson, R.F.; Stankovic, J.A. Monitoring body positions and movements during sleep using WISPs. In *Wireless Health*; ACM: New York, NY, USA, 2010; pp. 44–53.

27. van Der Loos, H.; Kobayashi, H.; Liu, G. Unobtrusive vital signs monitoring from a multisensor bed sheet. In Proceedings of the RESNA Conference, Reno, NV, USA, 22–26 June 2001.

28. Xiao, Y.; Lin, J.; Boric-Lubecke, O.; Lubecke, V.M. A Ka-band low power doppler radar system for remote de-tection of cardiopulmonary motion. In Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference, Shanghai, China, 17–18 January 2005; pp. 7151–7154.

29. Bak, J.U.; Giakoumidis, N.; Kim, G.; Dong, H.; Mavridis, N. An intelligent sensing system for sleep motion and stage analysis. *Procedia Eng.* **2012**, *41*, 1128–1134. [CrossRef]

30. Sadeh, A. Sleep assessment methods. *Monogr. Soc. Res. Child Dev.* **2015**, *80*, 33–48. [CrossRef] [PubMed]

31. Deng, F.; Dong, J.; Wang, X.; Fang, Y.; Liu, Y.; Yu, Z.; Liu, J.; Chen, F. Design and implementation of a noncontact sleep monitoring system using infrared cameras and motion sensor. *IEEE Trans. Instrum. Meas.* **2018**, *67*, 1555–1563. [CrossRef]

32. Gao, Z.; Ma, Z.; Chen, X.; Liu, H. Enhancement and de-noising of near-infrared Image with multiscale Morphology. In Proceedings of the 2011 5th International Conference on Bioinformatics and Biomedical Engineering, Wuhan, China, 10–12 May 2011; pp. 1–4.

33. Holtzhausen, P.J.; Crnojevic, V.; Herbst, B.M. An illumina-tion invariant framework for real-time foreground detection. *J. Real-Time Image Process.* **2015**, *10*, 423–433. [CrossRef]

34. Park, Y.K.; Park, S.L.; Kim, J.K. Retinex method based on adaptive smoothing for illumination invariant face recognition. *Signal Process.* **2008**, *88*, 1929–1945. [CrossRef]

35. Maik, V.; Paik, D.T.; Lim, J.; Park, K.; Paik, J. Hierarchical pose classification based on human physiology for behaviour analysis. *IET Comput. Vis.* **2010**, *4*, 12–24. [CrossRef]

36. Wang, Y.K.; Cheng, K.Y. A two-stage Bayesian network method for 3D human pose sstimation from monocular image sequences. *EURASIP J. Adv. Signal Process.* **2010**, *2010*, 761460. [CrossRef]

37. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [CrossRef]

38. Bay, H.; Ess, A.; Tuytelaars, T.; Gool, L.V. Speeded-up robust features (SURF). *Comput. Vis. Image Underst.* **2008**, *110*, 346–359. [CrossRef]

39. Ouyang, W.; Tombari, F.; Mattoccia, S.; Stefano, L.D.; Cham, W.K. Performance evaluation of full search equivalent pattern matching algorithms. *IEEE Trans. Circuits Syst. II Analog Digit. Signal Process.* **2012**, *34*, 127–143.

40. Scovanner, P.; Ali, S.; Shah, M. A 3-dimensional SIFT descriptor and its application to action recognition. In Proceedings of the 15th ACM international conference on Multimedia, Augsburg, Germany, 25–29 September 2007; pp. 357–360.

41. Wang, H.; Yi, Y. Tracking salient keypoints for human action recognition. In Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, Kowloon, China, 9–12 October 2015; pp. 3048–3053.

42. Zhang, J.T.; Tsoi, A.C.; Lo, S.L. Scale Invariant feature transform flow trajectory approach with applications to human action recognition. In Proceedings of the International Joint Conference on Neural Networks, Beijing, China, 6–11 July 2014; pp. 1197–1204.

43. Molina, A.; Ramirez, T.; Diaz, G.M. Robustness of interest point detectors in near infrared, far infrared and visible spectral images. In Proceedings of the 2016 XXI Symposium on Signal Processing, Images and Artificial Vision (STSIVA), Bucaramanga, Colombia, 31 August–2 September 2016; pp. 1–6.

44. Schweiger, F.; Zeisl, B.; Georgel, P.F.; Schroth, G. Maximum detector response markers for SIFT and SURF. In Proceedings of the Workshop on Vision, Modeling and Visualization, Braunschweig, Germany, 16 November 2009; pp. 145–154.

45. Wang, Y.K.; Huang, W.B. A CUDA-enabled parallel algorithm for accelerating retinex. *J. Real-Time Image Process.* **2012**, *9*, 407–425. [CrossRef]

46. Hare, S.; Saffari, A.; Torr, P.H.S. Efficient online structured output learning for keypoint-based object tracking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 1894–1901.

47. Huang, C.; Darwiche, A. Inference in belief networks: A procedural guide. *Int. J. Approx. Reason.* **1996**, *15*, 225–263. [CrossRef]

48. Wang, Y.K.; Su, C.H. Illuminant-invariant Bayesian detection of moving video objects. In Proceedings of the International Conference on Signal and Image Processing, Honolulu, HI, USA, 14–16 August 2006; pp. 57–62.

49. Cole, R.J.; Kripke, D.F.; Gruen, W.; Mullaney, D.J.; Gillin, J.C. Automatic sleep/wake identification from wrist activity. *Sleep* **1992**, *15*, 461–469. [CrossRef] [PubMed]

50. Ramanan, D.; Forsyth, D.A.; Zisserman, A. Tracking people by learning their appearance. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 65–81. [CrossRef] [PubMed]