A Vision-Based System for In-Bed Posture Tracking

Shuangjun Liu and Sarah Ostadabbas
Augmented Cognition Lab (ACLab)
Electrical and Computer Engineering Department
Northeastern University, Boston, MA 02115
liu.shu@husky.neu.edu, ostadabbas@ece.neu.edu

Abstract

Tracking human sleeping postures over time provides critical information to biomedical research including studies on sleeping behaviors and bedsore prevention. In this paper, we introduce a vision-based tracking system for pervasive yet unobtrusive long-term monitoring of in-bed postures in different environments. Once trained, our system generates an in-bed posture tracking history (iPoTH) report by applying a hierarchical inference model on the top view videos collected from any regular off-the-shelf camera. Although being based on a supervised learning structure, our model is person-independent and can be trained off-line and applied to new users without additional training. Experiments were conducted in both a simulated hospital environment and a home-like setting. In the hospital setting, posture detection accuracy using several mannequins was up to 91.0%, while the test with actual human participants in a home-like setting showed an accuracy of 93.6%.

1. Introduction

1.1. Motivation

Human sleeping posture's track record has shown to be an important medical indicator for many healthcare complications including sleep apnea [16], pressure ulcers [4], and even carpal tunnel syndrome [22, 23]. Sleeping posture has also an effect on dream experience [34]. Moreover, patients are usually required to maintain specific postures after certain surgeries to get a better recovery result [1]. Among these applications, automatic in-bed posture tracking to prevent pressure ulcers (bedsores) has received a lot of attentions lately. Pressure ulcers appear commonly in hospitals and nursing homes, particularly in patients who lack the ability of repositioning themselves or those who cannot feel the pain of being in the same posture for an extended period of time [4]. In hospital settings, caregivers need to be attentive to the patients who are more susceptible to this condition, and take action to relieve pressure by changing their lying postures every two hours. The United State healthcare system takes on a serious monetary burden in order to prevent and treat pressure ulcers, putting a strain on all hospital resources [15]. This is largely due to the difficulty of treating developed ulcers when the price of managing a single full-thickness pressure ulcer can be as high as $70,000 [8, 30]. Furthermore, methods employed to reduce the incidents of hospital-acquired ulcers requires already overworked nursing staff to come to patients on a regular basis and manually reposition them [32]. Some alternative solutions have been previously proposed to assist in preventing bedsores, the most common of which is automatic posture tracking of the bed-bound patients and personalizing the care for patients based on their need for repositioning [25, 26].

1.2. Related Works

Currently, in-bed posture detection methods mainly rely on the use of pressure mapping systems. Authors in [28] extracted binary signatures from pressure images obtained from a commercial pressure sensing mat and used a binary pattern matching technique for posture classification. The same group also introduced a Gaussian mixture model (GMM)-based clustering approach for concurrent posture classification and limb identification using pressure data [24]. Pictorial structure model of the body based on both appearance and spatial information was employed to localize the body parts within pressure images in [17]. Although pressure mapping-based methods are effective at localizing areas of increased pressure and even automatically classifying overall posture [24], the pressure sensor mats are expensive and require frequent maintenance. These obstacles have prevented pressure sensing solutions from achieving large-scale popularity.

By contrast, vision-based methods show great advantages such as low cost and ease of maintenance. Vision-based human pose tracking for surveillance applications has been an active area of research during the past few decades and many generic human detection and posture tracking methods have already been developed [27]. Bourdev et al. proposed a new notion, "poselet" to be used for human detection and pose estimation as well as localization of body components [6]. It was defined with a set of examples that were tightly clustered in the configuration space of the body, which was parametrized by 3D coordinates of the joints. In [5], the same group provided a better way to define and use poselets, which made it applicable to a wider range of object categories. Another extension to this model used a deformable part model with several poselets, each aligned to a specific configuration of key points [14].

Vision-based methods have also been employed in several healthcare scenarios. Sathyanarayana et al. in [31] gave a comprehensive review of vision-based technique for fall detection, action and activity monitoring, sleep monitoring, etc. For in-bed status monitoring, some groups only focus on the action detection such as leaving or getting into a bed [11]. A comprehensive critical care platform is introduced in [21]. Martinez et al. proposed "BAM" descriptor based on depth information collected from a Microsoft Kinect, which could monitor the sleeping posture and movement data [20]. His additional work further addressed high level activities such as removing bed covers [19]. Yu et al. also successfully employed the depth data to localize the head and body parts [35].

In spite of the related body of work, 2-dimensional vision-based in-bed posture analysis has not been explicitly addressed in the literature. Though generic vision-based method exists, they either focus on a street scenario for pedestrian detection or for the purpose of daily activity detection and none of them is specifically optimized and configured for in-bed posture detection. Normally, generic methods are designed to cope with versatile cases, which are sophisticated and computationally intensive and not suitable for realtime applications. Nevertheless, in medical domain, time constraints and strict accuracy requirements are both needed in majority of the applications, including in-hospital patient monitoring. For a bed-bounded patient who has lost the repositioning ability, misrecognition can last for a long time and can lead to serious consequences. As proposed in [31], utilizing the application context of patient monitoring can be a useful way to develop novel techniques that are accurate and robust and yet cost-effective.

1.3. Our Contribution

In this paper, we introduce a robust 2-dimensional vision-based in-bed posture tracking system, which can pervasively be employed in different environments using an inexpensive regular webcam combined with our tracking algorithm. After applying a series of preprocessing steps on the recorded videos from individuals while being in different in-bed positions, each video frame is fed to our posture recognition module. During posture recognition, histogram of oriented gradients (HOG) feature descriptors of each video frame are extracted as the luminance-invariant attributes to be used in a pre-trained posture classifier. A feature space dimensionality reduction using principle component analysis (PCA) is also applied to avoid overfitting issue when training the classifier. A latent parameter is introduced in the context of narrow field of view in our application during posture recognition, which improves the classification accuracy and also acts in the role of occupation detection. After in-bed postures are recognized and labeled as supine, left or right side, our system generates a person-specific in-bed posture tracking history (iPoTH) report for each user.

2. Methodology

2.1. Problem Statement

Our in-bed posture detection algorithm can be described as a supervised classification problem such that given an input RGB image frame I_i of size $M_1 \times N_1$ at time frame t_i, the inference model predicts the current sleeping posture $P(t_i)$ as one of the K postures in the predefined in-bed posture set, $\{P_1, P_2, ...P_K\}$.

2.2. Posture Categories

Human bodies can be represented by deformable templates [13, 29]. Limbs and torso can be deemed as articulated together from a kinematic point of view. It is evident that any two arbitrary sleeping postures can be hardly exactly the same, however, considering a high degree of granularity in posture categorization may result in a great deal of aliasing in classification outcome. Therefore, we built the sleeping posture categories based on the $K = 3$ major stable in-bed postures of supine, $P_1 = S$, right side $P_2 = R$, and left side $P_3 = L$, in which we simply ignore the rare case of prone posture. For system applicability, we also considered another category as the *unoccupied* case, U, where there is no lying person in the detection window.

2.3. Preprocessing Steps

Before feeding each video frame to our posture detection algorithm, a series of preprocessing steps needs to be performed.

2.3.1 Camera self-calibration consideration

First, to make the system work with various RGB video capturing devices and under different lighting conditions, the algorithm should be hardware/condition independent. After capturing several images with a regular webcam under different lighting, it turned out that an automatic self-calibration process runs in the camera software. This process can be corrected by setting fixed exposure and gain, but in the context of our application, we would rather keep self-calibration, since the lighting condition differs tremendously between day and night when the system should work consistently. Regardless of the reason, the histogram adjustment may result in adversary effects in the analysis outcome, including the accuracy of the background subtraction.

To deal with different cameras' self-calibration and lighting conditions, the image feature descriptors used for classification should be luminance-invariant. There are some feature descriptors which meet this requirement such as multi-scale oriented patches (MOPS) [7], scale invariant feature transform (SIFT) [18], and speeded-up robust features (SURF) [3]. However, these descriptors are usually based on the area around interest points. In our application, it is not practical to establish such universal interest points for different persons with variable appearances and sleeping poses. Therefore, we chose histogram of oriented gradients (HOG) as our feature descriptor, which is a window-based descriptor that slides through the image. HOG is similar to SIFT, but is computed on a dense grid of uniformly spaced blocks consisted of cells [9]. Another reason we chose HOG as our descriptor is that it solely relies on gradient information instead of color. Therefore, it can easily be extended to infrared imaging version for low light (e.g. at night) monitoring conditions.

2.3.2 Image downsampling

Sampled on dense grid, HOG method will result in a feature vector with extra large dimension. To avoid the unnecessary large dimensionality, the image can be downsampled. HOG is gradient based and the posture information is only given by the person's profile which generally will not be affected by the downsampling process. Most webcams feedback a 16:9 or 4:3 video stream and for sleeping postures, it is intuitive to observe them from a portrait view. Therefore, we rotate the input image by 90° clockwise with bicubic interpolation downsampling process. We choose and fix image width at N_2, so the output image height is determined by $M_2 = [\frac{N_1 \times N_2}{M_1}]$, where $M_1 \times N_1$ is the size of the input image and $M_2 \times N_2$ is the size of the output image. Note that the output image is already rotated by 90° in our pipeline.

2.4. Feature Extraction

The HOG features are constructed as follows: the downsampled $M_2 \times N_2$ image frame is first divided into small evenly distributed rectangular cells. The unsigned gradient orientations are evenly divided into b orientation bins. The gradients over the pixels of a cell are accumulated into their corresponding bins according to their directions, to form a 1-D histogram. Adjacent cells are grouped into larger spatial regions called blocks and the local histogram "energy" is calculated to normalize the contrast of the local cells. The dense gridded overlapped blocks form the HOG descriptors of the detection window.

HOG features are controlled by several parameters. Assuming that we employ only square cell and block shape, then the parameters of the HOG features are as follows: the cell size of $m \times m$, the block size of $n \times n$, the overlapped cell of size o_l, and the orientation bin's number of b. The final feature vector dimension is also determined by the input image size of $M_2 \times N_2$. The length of the feature vector, F, then is a function of all parameters as $|\mathrm{F}| = \left\lfloor (\frac{M_2}{m} - n)/(n - o_l) + 1 \right\rfloor \times \left\lfloor (\frac{N_2}{m} - n)/(n - o_l) + 1 \right\rfloor$.

2.5. In-Bed Posture Classification

Compared to a pedestrian detection, lying subjects will not always show a straight up posture in detection window. This can be caused by a misaligned hospital bed or subject self relocation. Therefore, we introduce a latent variable in classification to handle this issue. To generalize this model, we employ $\mathrm{F}(I)$ to represent the HOG features extracted from a given image frame I. Since the dimension of an HOG feature vector is extremely large, we employed the principle component analysis (PCA) technique to reduce the dimensionality, which is acting as our mapping kernel of feature vector as $\phi(\mathrm{F})$. Furthermore, we added a latent variable z to address the uncertainty in target alignment, so the feature vector will be a function of both image frame and latent variable, as $\phi(\mathrm{F}(z, I))$. To define a practical function, we observed the following practical considerations and addressed them by introducing z in the classification problem.

2.5.1 Hospital bed misalignment

In most real-world cases, hospital beds are wheeled and moved in and out frequently. Also, it is probable that the bed will be wheeled back in opposite orientation. We assume that the hospital staff will generally re-position the bed to the original location with either near vertical or inverse vertical orientations. Thus, the latent variable is defined as $z = (x, y, \theta)$, where x and y represent the shift, and θ represents the rotation variation. Based on prior knowledge, the latent variable can be constrained to a narrower domain. For this study, we have assumed $p(z)$, prior model for latent variable in both uniform and Gaussian distributions. For example, a Gaussian $p(z)$ with zero mean is assumed to have the rotation prior with standard deviation of 20°, and two kernels located at 90° and -90°, and shift prior with standard deviation of 3 pixels.

2.5.2 Narrow field of view

Sliding window across image plane is the most commonly strategy for for object detection, which is computationally intensive [12]. For applications with large field of view, such as pedestrian detection in a street this seems a required step. However, for in-bed posture detection, the distance between the person in bed and the camera is limited by the room ceiling height (roughly between 8-9 feet for most bedrooms) and the bed height. This will result in a narrow field of view, where the person tightly occupies the view field. In our design, we take this into consideration and simplify detection process considering only limited variations. This may reduce the flexibility of the algorithm to handle

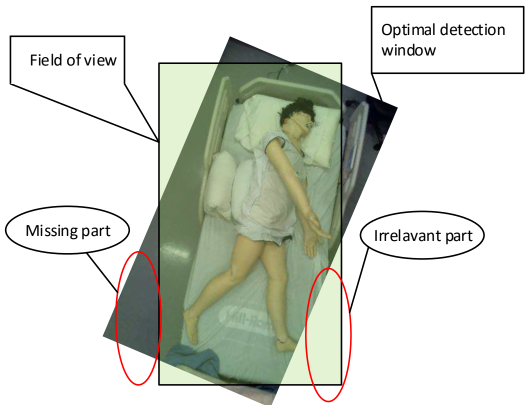


Figure 1: The optimal detection window misaligned with the field of view.

versatile scenarios, however this is sufficient for our specific application, where a camera can be easily setup only once during installation to provide overview perspective. In many computer vision datasets for object detection such as PASCAL [12], a bounding box is used to indicate the target, which gives a tight bound around the target area. In our case, subjects are located in the center of the vision field and almost fully occupy it. So we can simply employ the acquired images directly for training purpose. Latent variables are introduced in some work, where they can employ the surrounding information of the target object [13]. This is not the case in our dataset as the surrounding information sometime is missing as shown in Fig. 1. To tackle this problem, we adopted a "symmetric" wrapping method to pad the missing areas to generate the kernel function $\phi(\mathrm{F}(z,I))$ [2].

2.5.3 Introducing a latent variable in posture classification problem

Employing a support vector machine (SVM) as the classifier, score of a certain category is represented as:

$$s(\mathrm{F},z) = P(t_i)(w^T\phi(\mathrm{F}(z,I_i))+b) \tag{1}$$

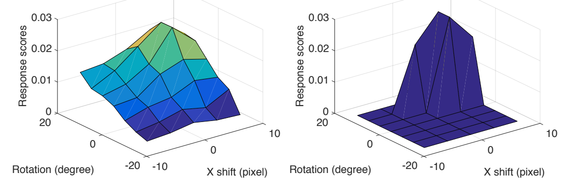
where $P(t_i)$ is the predicted in-bed posture label at time t_i, and w and b are the coefficients trained for the SVM. Defining $\gamma = (w,b)$, the posterior of a positive detection can be achieved from a sigmoid mapping function:

$$P(P^*|\mathrm{F},\gamma,z) = \frac{1}{1+e^{\alpha s(\mathrm{F},z)+\beta}} \tag{2}$$

where P^* means a positive detection. The coefficient α and β will be achieved from fitting of the training dataset. Alternatively, we can also employ a simple linear mapping function to map the score to range (0,1) as a pseudo posterior. Both models are tested in our experiments and results are reported in Section 3. The posterior of latent variable z can be achieved from Bayesian rule by assuming a uniform distribution of bed occupation:

$$P(z|P^*,\mathrm{F},\gamma) \propto P(P^*|\mathrm{F},\gamma,z)P(z) \tag{3}$$

(a) (b)



(c) (d)

Figure 2: Latent variable optimization: (a) original image, (b) rectified image, (c) posterior response of latent variable with linear mapping method, (d) posterior response of latent variable with sigmoid function.

The optimal latent variable, z_{opt} is obtained by applying maximum a posteriori (MAP) probability estimate algorithm on Equation (3). An example of this process is demonstrated in Fig. 2. An original image of Fig. 2a has been realigned in Fig. 2b with a slight rotation. It is noticeable that the bed was not realigned to a perfect vertical position, but latent variable searching process focuses on the subject instead of the bed. Note that subject in the rectified image has been brought back to a centered vertical position as we anticipated. Fig. 2c and Fig. 2d employ two different mapping method for posterior, (c) is simple linear mapping from the classification scores and (d) employs the sigmoid function.

After we obtain an z_{opt}, data are further classified into a specific in-bed posture as L,R or S. Three classifiers are trained based on the error-correcting output codes (ECOC) [10], assigned as shown in Table 1 . During detection session, only the first classifier employs the latent variable after localizing the target to an optimal position. It is not necessary to tune the latent variable again in the posture recognition stage. The benefit of this design is to reduce the computational cost and improve the system's realtime performance.

Table 1: ECOC for posture classification.

Posture	SVM1	SVM2	SVM3
P_1	1	1	0
P_2	-1	0	1
P_3	0	-1	-1

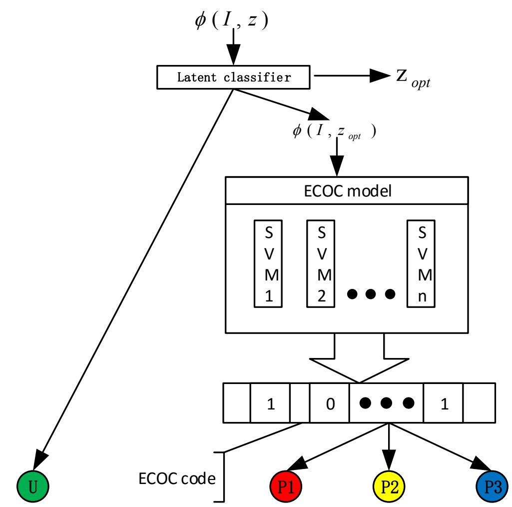


Figure 3: Latent classifier with ECOC mechanism.

2.5.4 Details on classification implementation

Latent variable method is introduced in several computer vision applications, which aims at specific class detection [13]. Instead of employing latent variable for single category classification, it turned out that the occupied samples even with different postures of L, R and S are more similar to each other when compared to the unoccupied ones. Therefore, we employed latent variable for a combined category instead of individual ones in order to reduce the computational intensity of our algorithm. Furthermore, well-aligned samples with well-centered bed setting are easier to be collected in our case to form the training set and model parameters can be trained directly from them. This is an example of a privilege learning approach using privileged training information [33]. The detection process algorithm is shown in Algorithm. 1. To alleviate the computation cost, we only search the latent variable in a searching space S_{sch} which only contains limited points in the latent variable space. The final structure of latent variable search is shown in Fig. 3.

Algorithm 1: Latent variable optimization and occupation detection.

Data: I
Result: z_{opt}, occupied or not
initialization;
Find z_{opt} to maximize $P(z|P^*, \mathrm{F}, \gamma)$
if $P(z_{opt}|P^*, \mathrm{F}, \gamma) > 0$ **then**
  return *true*;
else
  return *false*;

2.5.5 Filtering transitioning states

During posture changes in bed, transitioning states can introduce misleading classification results to the system. To

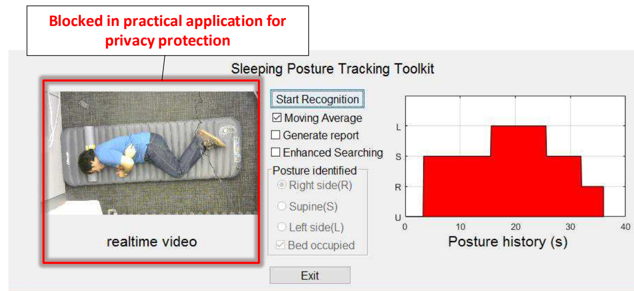


Figure 4: A snap shot of sleeping tracking system GUI.

avoid unreliable detection during these transition, we employed a first input first output (FIFO) buffer to accommodate recent posture results and perform a sliding median filter on them. The buffer can be represented as $[P(t_{i-L}), P(t_{i-L+1}), ..., P(t_i)]$, where L stands for median filter window width. Having $i > L$ holds only backward filter window, which will result in a bit of delay; however, it suppresses the jitters during posture transition. We extended the posture set with unoccupied case to be a state set $\{U, S, R, L\}$ to cover all the possible states of the system output. To perform the median filtering, we translated the state set to its enumeration integer indexes as $\{0, 1, 2, 3\}$. The reason behind choosing median filter is having a discrete state set. In our system design, this function is optional if delay has to be avoided.

2.6. In-Bed Posture Tracking History (iPoTH)

For long-term posture monitoring purposes such as sleeping behavior studies, we implemented a history report generation function in this system. When enabled, it will generate a history report for each monitoring interval. The in-bed posture tracking history (iPoTH) report will be named with exact date and time of data collection. It follows the format of 'iPoTH_month_date_year_hour_minute_second'. The report content is a 2-column array of posture record with the time stamps.

2.7. GUI Design

A graphical user interface (GUI) has been designed as shown in Fig. 4. The GUI has two major graphic displays, one for realtime video display and another one for a iPoTH report visualization. The grey "Posture identified" panel gives the very recent recognized posture result alongside a check box for occupation detection. The moving average median filter, report generation, and enhanced searching functions are all optional to the user listed as check boxes in the GUI, where the enhanced searching will introduce the latent variable during prediction. This GUI is for demo purpose and in practical application the realtime video could be blocked to protect the patient's privacy.

(a) (b) (c) (d) (e) (f)

Figure 5: Sleeping mannequin images captured at hospital setting.

(a) (b) (c) (d) (e) (f)

Figure 6: Artificial unoccupied dataset at hospital setting.

3. Experimental Results

Currently, there is no public in-bed posture dataset. Therefore, in order to evaluate our posture tracking performance, we setup our system in two different scenarios, a simulated hospital environment and a home-like setting. We collected data and applied our in-bed posture classification algorithms on datasets from both scenarios as well as on realtime videos.

3.1. Mannequin Test at Hospital Setting

First, we setup our system in a healthcare practice simulation lab with real-life clinical settings.

3.1.1 Data recording hardware

Off-the-shelf materials were used to prototype the proposed system and demonstrate the potential of our system to be built easily anywhere. In our setup, the system hardware includes: (1) HP envy x360 laptop, running Windows 10 Home Edition i5 CPU@1.7GHz and memory of 8GB, and (2) Logitech C525 HD webcam with the maximum frame rate of 15fps. We attached the camera to the ceiling tile of the simulation lab over a hospital bed. We chose a 16:9 video format with 1280×720 resolution. For the downsampling process, we chose the driven size of $N_2 = 64$.

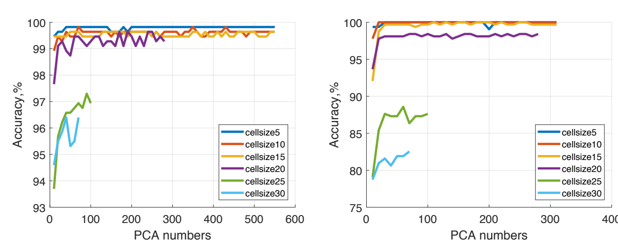
3.1.2 Dataset design

In our system design, we not only evaluate in-bed postures, but also detect the occupation state. The two general categories, are labeled as $\tilde{U}$ for occupied ones and as U for unoccupied ones. The occupied ones are further classified into categories from posture set $\{S, R, L\}$. To form the occupied dataset, we employed one male and one female life-size mannequins to present random postures for each posture category. We used the hardware described in Section 3.1.1 and Windows10 built-in camera app for video acquisition. To form the training dataset, we recorded approximately 10 frames from each mannequin for each posture category. Limited by the number of the mannequins, we dressed the mannequins with different hospital gowns to account for different subjects and clothing. We noticed some ambiguous postures during dataset forming, such as the body sometimes stably stayed in between supine and side lying posture. To clarify this ambiguity, we established a rule to determine whether or not a posture belongs to S class, which mainly depends on the upper body. S refers to postures, where the angle between the body plane and the mattress is less than 45 degrees. In several cases, the mannequins are positioned with a pillow support. In total, we collected 315 samples for occupied dataset with 102, 102, and 111 samples for L, R, and S categories, respectively. Some sample frames are shown in Fig. 5. To form the negative dataset with unoccupied bed images, we simply manipulated the bed, pillows, and blankets on it to form several unoccupied cases, in which some samples are show in Fig. 6.

3.1.3 Posture classification performance evaluation

To find the model parameters used for latent variable searching, we employed the unoccupied bed dataset, D_U with 113 samples and occupied bed dataset, $D_{\tilde{U}}$ with 315 samples. PCA analysis was applied on the $D_U \bigcup D_{\tilde{U}}$ of the training set and the same PCA coefficients were employed on test set. An SVM binary classifier was applied and a 10-fold cross validation technique was conducted to evaluate the posture recognition accuracy. To choose the optimal number of principle components and the HOG feature cell size, we evaluated the occupation detection and also the posture estimation performance with varying PCA numbers with step 10 and also varying cell size. From the result shown in Fig. 7, we can see cell size ranging from 5 to 15 yields a good result. To choose optimal PCA number, we chose the one that yields the best performance with cell size ranging from 5 to 15. In our test, we set PCA number as 110.

To further validate the recognition robustness, we intentionally shifted and rotated the bed in order to extend the original dataset to contain alternative position and rotation cases. Based on this, we synthesized the inverse direction



(a) (b)

Figure 7: Bed occupation and posture classification accuracy based on the number of principle components with different HOG cell size using mannequin data: (a) occupation detection performance, (b) posture estimation performance.

edition to simulate when hospital bed relocated with an inverse orientation. For latent variable search, there was only very small margins in y direction around the person, which could be ignored. The variation along x-axis is represented by searching space $S_x = [-8, 8]$ with a step size of 4 and rotation as $S_\theta = [-15^\circ, 15^\circ]$ with a step size of 5°.

To present a fair comparison between result with and without latent variable searching, we first tested only on the samples with upright orientation. We combined Gaussian (gau) and uniform (unif) prior with sigmoid (sig) posterior and linear mapping (lin) posterior separately. The classification results are shown in Table 2. For Gaussian prior, considering the synthesized inverse orientation cases, we used the Gaussian mixtures with two equal weighted kernels located in two opposite direction as orientation prior model. In our experiment, occupation test and posture classification test were performed separately. For posture classification, we assumed all of them are correctly recognized as occupied ones. Table 2 demonstrates that all the latent searching methods no matter what exact prior and posterior model they take, give much better performance in occupation detection with error rate of 3.6% than the one without latent variable. The posterior mapping method does not have a significant effect. We believe this is reasonable as the score and the fitted posterior are positively correlated. Overall, the uniform prior case showed better performance, which may caused by the test set distribution. In our case, misaligned samples have almost the same quantity as well-aligned ones, so the uniform distribution fits better for our test set. However, in real world cases, it would be more like a normal distribution as care givers will be trained to well-aligned the bed. The confusion matrix with uniform prior and posterior posterior mapping is shown in Table 3.

3.2. Human Test at Home-like Setting

To test the feasibility of our posture recognition method in multiple scenarios, we made another setup with an air mattress and a camera hanging above it to perform our method on real human participants.

Table 2: Classification results of different prior and posterior mapping methods with HOG cell size 5 and PCA number 110.

	sig+gau	sig+unif	lin+gau	lin+unif	no latent
Posture Classification	90.3%	91.0%	90.3%	91.0%	60.7%
Occupation Detection Error	3.6%	3.6%	3.6%	3.6%	70.0%

Table 3: Confusion matrix of uniform + sigmoid latent searching.

		Predicted Postures		
		L	R	S
Actual Postures	L	118	1	0
	R	0	57	16
	S	6	2	77

3.2.1 Dataset design

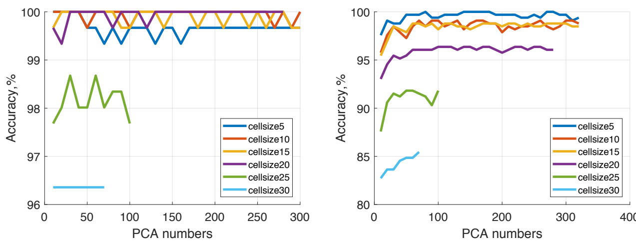
To form the occupied dataset, we invited 12 participants (11 males) to take part in this experiment. The experimental procedure is as follows. An air mattress was well positioned in the center of a webcam field of view. Participants were asked to lie in bed in three general categories of supine (S), left side (L), and right side (R). Within each posture category, participants were encouraged to adjust their poses to feel natural and comfortable. We recorded approximately 10 frames from each participant for each posture category. We used the same hardware described in Section 3.1.1 but this time with Logitech C720 webcam and Windows10 built-in camera app for video acquisition. In total, we collected 358 samples for occupied dataset with 115, 120, and 123 samples for L, R, and S categories, respectively.

For negative dataset, it could be any frame without a person in it. Since our system is targeted for different environments such as hospital or family residence, video frames would have various backgrounds that we can hardly simulate in one fixed lab setup. Unfortunately, there are no publicly available dataset focusing on top view indoor images. However, since HOG features depend highly on the gradient information, we were able to use any image without human in it to provide the random gradient information for our inference model. Therefore, we collected the unoccupied sample images in two ways. First, we simply generated several random scenarios without human by our experimental setup, which shared high similarities in the surrounding area with our occupied dataset. The other way was using the public online dataset INRIA. To generate unoccupied sample images from INRIA, we employed a random window strategy within the original images and cropped a random patch from each of them. In total, we generated 1257 unoccupied samples for U set. Several occupied and unoccupied samples are shown in Fig. 8.

(a) (b) (c) (d) (e) (f)

(g) (h) (i) (j) (k) (l)

Figure 8: Human subjects dataset: (a)-(f) samples with human subjects, (g)-(l) samples without human in it.



(a) (b)

Figure 9: Bed occupation and posture classification accuracy based on the number of principle components with different HOG cell size using real human data: (a) occupation detection performance, (b) posture estimation performance.

3.2.2 Posture classification performance evaluation

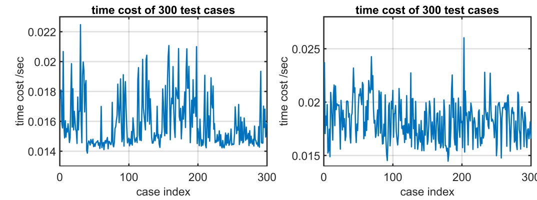
We employed similar test procedure described in Section 3.1.3. The prediction performances for both occupation and poses with different PCA number and cell size are shown in Fig. 9. We chose cell size equal to 10 and PCA number to be 110. Even though perfect accuracy could be achieved with larger component numbers, however it was unstable when components were added or deduced from our dataset. As for human dataset design, we did not intentionally build up the samples with mattresses rotated and shifted. We only synthesized alternative images from one participant's original 15 samples, and our latent classification method found all these variations, correctly. To simulate the application scenario, we trained the model with 11 subjects and left out one person's data which the model has never seen as the test set. Since we aim at posture estimation when subject reaches a stable state instead of transitional ones, static recorded frames can fairly reflect model accuracy instead of all frames from a stream. On our test set, the model detected all the occupations correctly and achieved a posture estimation accuracy of 93.6% with confusing matrix shown in Table 4.

3.3. Time Efficiency Test

We also tested the time efficiency of our algorithm. To rule out the hardware factors, we employed the pre-recorded videos to conduct the run-time test by excluding the acquisition time. We ran 300 test cases to evaluate the time efficiency with and without PCA. The recognition pipeline with PCA showed an average run-time of 15.9ms and the recognition without PCA showed an average run-time of 18.3ms, as shown in Fig. 10(a) and (b), respectively. Therefore, PCA in our method reduces the time cost approximately by 13% compared to the results with original HOG features.

Table 4: Confusion matrix of uniform + sigmoid latent searching

		Predicted Postures		
		L	*R*	*S*
Actual Postures	*L*	9	0	0
	R	0	11	0
	S	0	2	9



(a) (b)

Figure 10: Posture tracking pipeline run-time test with 300 cases: (a) time cost with PCA, (b) time cost with HOG features directly and no PCA.

When the latent variable is introduced, the time cost depends on the granularity, dimension, and range of the searching space. For time performance evaluation of our enhanced searching, we considered a searching space with $S_x = \{-8, -4, 0, 4, 8\}$ and $S_\theta = \{-10°, -5°, 0°, 5°, 10°\}$. The average time cost was 280ms.

4. Conclusion

In this paper, based on a series of experiments, we demonstrated that our proposed vision-based posture tracking system can successfully detect the person's in-bed position over time and generate a person-specific iPoTH report. Our system is implemented based on a regular webcam and a laptop, which makes it easy to be setup in ordinary family/nursing homes or hospitals. Its intuitive GUI provides direct visual feedback to non-expert computer users to operate it. The proposed system not only provides caregivers with a long-term in-bed posture history of each patient, but also could be used in sleeping behavior studies applicable in several psychophysiological domains.

Nonetheless, our current work is still limited under certain contexts. Its effectiveness with night vision systems instead of visible light cameras remains untested, which will be explored in future work. Another major challenge for in-bed posture monitoring is the high probability of subject being covered by a sheet or blanket. In fact, vision-based methods would no longer be functional in this case where other sensing modalities should be explored for information retrieval. This is an important issue for further exploration. Larger dataset should also be built up to provide a more reliable evaluation platform for in-bed posture and pose estimation studies.

References

[1] M. Abouzari, A. Rashidi, J. Rezaii, K. Esfandiari, M. Asadollahi, H. Aleali, and M. Abdollahzadeh. The role of postoperative patient posture in the recurrence of traumatic chronic subdural hematoma after burr-hole surgery. *Neurosurgery*, 61(4):794–797, 2007. 1

[2] T. Akenine-Möller, E. Haines, and N. Hoffman. *Real-time rendering*. CRC Press, 2008. 4

[3] H. Bay, T. Tuytelaars, and L. Van Gool. Surf: Speeded up robust features. *European conference on computer vision*, pages 404–417, 2006. 3
[4] J. Black, M. M. Baharestani, J. Cuddigan, B. Dorner, L. Edsberg, D. Langemo, M. E. Posthauer, C. Ratliff, G. Taler, et al. National pressure ulcer advisory panel's updated pressure ulcer staging system. *Advances in skin & wound care*, 20(5):269–274, 2007. 1
[5] L. Bourdev, S. Maji, T. Brox, and J. Malik. Detecting people using mutually consistent poselet activations. *European conference on computer vision*, pages 168–181, 2010. 2
[6] L. Bourdev and J. Malik. Poselets: Body part detectors trained using 3d human pose annotations. *2009 IEEE 12th International Conference on Computer Vision*, pages 1365–1372, 2009. 2
[7] M. Brown, R. Szeliski, and S. Winder. Multi-image matching using multi-scale oriented patches. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 1:510–517, 2005. 3
[8] J. Cuddigan, D. R. Berlowitz, and E. A. Ayello. Pressure ulcers in america: prevalence, incidence, and implications for the future. *Advances in Skin & Wound Care*, 14(4):208, 2001. 1
[9] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 1:886–893, 2005. 3
[10] T. G. Dietterich and G. Bakiri. Solving multiclass learning problems via error-correcting output codes. *Journal of artificial intelligence research*, 2:263–286, 1995. 4
[11] J.-R. Ding. Bed status detection for elder-care center. In *2009 16th International Conference on Systems, Signals and Image Processing*, pages 1–4. IEEE, 2009. 2
[12] M. Everingham, S. M. A. Eslami, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The pascal visual object classes challenge: A retrospective. *International Journal of Computer Vision*, 111(1):98–136, Jan. 2015. 3, 4
[13] P. Felzenszwalb, D. McAllester, and D. Ramanan. A discriminatively trained, multiscale, deformable part model. *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8, 2008. 2, 4, 5
[14] G. Gkioxari, B. Hariharan, R. Girshick, and J. Malik. Using k-poselets for detecting people and localizing their keypoints. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3582–3589, 2014. 2
[15] B. Kuh and S. Coulter. Balancing the pressure ulcer cost and quality equation. *Nursing Economics*, 10(5):353–359, 1992. 1
[16] C. H. Lee, D. K. Kim, S. Y. Kim, C.-S. Rhee, and T.-B. Won. Changes in site of obstruction in obstructive sleep apnea patients according to sleep position: a dise study. *The Laryngoscope*, 125(1):248–254, 2015. 1
[17] J. J. Liu, M.-C. Huang, W. Xu, and M. Sarrafzadeh. Bodypart localization for pressure ulcer prevention. *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 766–769, 2014. 1
[18] D. G. Lowe. Object recognition from local scale-invariant features. *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, 2:1150–1157, 1999. 3
[19] M. Martinez, L. Rybok, and R. Stiefelhagen. Action recognition in bed using bams for assisted living and elderly care. In *Machine Vision Applications (MVA), 2015 14th IAPR International Conference on*, pages 329–332. IEEE, 2015. 2
[20] M. Martinez, B. Schauerte, and R. Stiefelhagen. bam! depth-based body analysis in critical care. In *International Conference on Computer Analysis of Images and Patterns*, pages 465–472. Springer, 2013. 2
[21] M. Martinez and R. Stiefelhagen. Automated multi-camera system for long term behavioral monitoring in intensive care units. In *MVA*, pages 97–100, 2013. 2
[22] S. J. McCabe, A. Gupta, D. E. Tate, and J. Myers. Preferred sleep position on the side is associated with carpal tunnel syndrome. *Hand*, 6(2):132–137, 2011. 1
[23] S. J. McCabe and Y. Xue. Evaluation of sleep position as a potential cause of carpal tunnel syndrome: preferred sleep position on the side is associated with age and gender. *Hand*, 5(4):361–363, 2010. 1
[24] S. Ostadabbas, M. Pouyan, M. Nourani, and N. Kehtarnavaz. In-bed posture classification and limb identification. *2014 IEEE Biomedical Circuits and Systems Conference (BioCAS) Proceedings*, pages 133–136, 2014. 1
[25] S. Ostadabbas, R. Yousefi, M. Faezipour, M. Nourani, and M. Pompeo. Pressure ulcer prevention: An efficient turning schedule for bed-bound patients. *Life Science Systems and Applications Workshop (LiSSA), 2011 IEEE/NIH*, pages 159–162, 2011. 1
[26] S. Ostadabbas, R. Yousefi, M. Nourani, M. Faezipour, L. Tamil, and M. Q. Pompeo. A resource-efficient planning for pressure ulcer prevention. *IEEE Transactions on Information Technology in Biomedicine*, 16(6):1265–1273, 2012. 1
[27] R. Poppe. A survey on vision-based human action recognition. *Image and vision computing*, 28(6):976–990, 2010. 1
[28] M. B. Pouyan, S. Ostadabbas, M. Farshbaf, R. Yousefi, M. Nourani, and M. Pompeo. Continuous eight-posture classification for bed-bound patients. *2013 6th International Conference on Biomedical Engineering and Informatics*, pages 121–126, 2013. 1
[29] D. Ramanan and C. Sminchisescu. Training deformable models for localization. *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, 1:206–213, 2006. 2
[30] M. D. Redelings, N. E. Lee, and F. Sorvillo. Pressure ulcers: more lethal than we thought? *Advances in skin & wound care*, 18(7):367–372, 2005. 1
[31] S. Sathyanarayana, R. K. Satzoda, S. Sathyanarayana, and S. Thambipillai. Vision-based patient monitoring: a comprehensive review of algorithms and technologies. *Journal of Ambient Intelligence and Humanized Computing*, pages 1–27, 2015. 2
[32] D. Smith. Pressure ulcers in the nursing home. *Annals of Internal Medicine*, 123(6):433–438, 1995. 1
[33] V. Vapnik and A. Vashist. A new learning paradigm: Learning using privileged information. *Neural networks*, 22(5):544–557, 2009. 5
[34] C. K.-C. Yu. The effect of sleep position on dream experiences. *Dreaming*, 22(3):212, 2012. 1

[35] M.-C. Yu, H. Wu, J.-L. Liou, M.-S. Lee, and Y.-P. Hung. Multiparameter sleep monitoring using a depth camera. In *International Joint Conference on Biomedical Engineering Systems and Technologies*, pages 311–325. Springer, 2012. 2