1-1-2013

Human Motion Analysis Based on Sequential Modeling of Radar Signal and Stereo Image Features

Meng Wu
University of Denver

Human Motion Analysis Based on Sequential Modeling of Radar Signal and Stereo Image Features

---

A Thesis

Presented to

the Faculty of the Daniel Felix Ritchie School of Engineering and Computer Science

University of Denver

---

in Partial Fulfillment
of the Requirements for the Degree
of Master of Science

---

by

Meng Wu

August, 2013

Advisor: Jun Zhang

Author: Meng Wu
Title: Human Motion Analysis Based on Sequential Modeling of Radar Signal and Stereo Image Features
Advisor: Jun Zhang
Degree Date: August, 2013

Abstract

Falls are one of the greatest threats to elderly health in their daily living routines and activities. Therefore, it is very important to detect falls of an elderly in a timely and accurate manner, so that immediate response and proper care can be provided, by sending fall alarms to caregivers.

Radar is an effective non-intrusive sensing modality which is well suited for this purpose, which can detect human motions in all types of environments, penetrate walls and fabrics, preserve privacy, and is insensitive to lighting conditions. Micro-Doppler features are utilized in radar signal corresponding to human body motions and gait to detect falls using a narrowband pulse-Doppler radar. Human motions cause time-varying Doppler signatures, which are analyzed using time-frequency representations and matching pursuit decomposition (MPD) for feature extraction and fall detection. The extracted features include MPD features and the principal components of the time-frequency signal representations. To analyze the sequential characteristics of typical falls, the extracted features are used for training and testing hidden Markov models (HMM) in different falling scenarios. Experimental results demonstrate that the proposed algorithm and method achieve fast and accurate fall detections.

The risk of falls increases sharply when the elderly or patients try to exit beds. Thus, if a bed exit can be detected at an early stage of this motion, the related injuries can be prevented with a high probability. To detect bed exit for fall prevention, the trajectory of head movements is used for recognize such human motion. A head detector is trained using the histogram of oriented gradient (HOG) features

of the head and shoulder areas from recorded bed exit images. A data association algorithm is applied on the head detection results to eliminate head detection false alarms. Then the three dimensional (3D) head trajectories are constructed by matching scale-invariant feature transform (SIFT) keypoints in the detected head areas from both the left and right stereo images. The extracted 3D head trajectories are used for training and testing an HMM based classifier for recognizing bed exit activities. The results of the classifier are presented and discussed in the thesis, which demonstrates the effectiveness of the proposed stereo vision based bed exit detection approach.

Acknowledgements

This thesis work is supported in part by the Knoebel Institute for Longevity and Health Awards of the University of Denver (No. 89317-142152) , OKT Enterprises, LLC, and National Science Foundation Industry/University Cooperative Research Center for Safety, Security and Rescue Research.

The radar data used in this thesis was collected in the Radar Imaging Lab at the Center for Advanced Communications, Villanova University, USA.

Contents

List of Tables

List of Figures

Chapter 1. Introduction

1.1 Motivation and background

Population aging has already become a global issue, and health care issues of elderly people attract many researchers' attention. Falls of the elderly cause major public health concerns as they often result in injuries, disability, and even death [19,22]. Almost 30 percent of elderly older than 65 years fall each year and about 10 percent of falls can lead to serious injuries when falls happen to those elderly people. Immediate assistance after a fall can reduce related complications [18]. Therefore, it is very important to detect elderly falls in a timely and accurate manner, so that immediate responses and proper care can be provided. When the elderly or patients try to get out of bed in the nursing homes or wards, the likelihood of falls increases sharply [3]. Bed exit alarms warn caregivers when movements related to bed exit are detected. The bed exit detection can be implemented in various ways with the common goal of informing caregivers to ensure the safety of a person who is trying to get out of the bed, and thus to reduce the risk of falls. To achieve this goal, it is necessary and urgent to develop reliable approaches to detect and predict related human motion and issue early warnings to reduce or prevent fall injuries.

1.2 Contribution

The contribution of the research work is summarized as following.

- **Radar signal based fall detection.** The matching pursuit decomposition (MPD) algorithm and the short time Fourier transform (STFT) are used for extract signal features, respectively. The principle component analysis (PCA) and time sequence decimation are used for dimension reduction of STFT features, which preserves information for classification and also reduce computational complexity. Hidden Markov model (HMM) and support vector machine (SVM) based classification algorithms are utilized for classification of fall and non-fall activities and both achieve satisfying classification performance.

- **Stereo image based bed exit detection.** The three-dimensional (3D) head tracks are used for recognizing bed exit activities. Specifically, first a head-shoulder area image model is trained using through histogram of oriented gradients (HOG) features, which can offer effective information for shape or edge detection. The head detection results are then processed by a data association algorithm to generate correct recognition of human head areas. Based on the resulted head detection areas, the 3D head trajectories are constructed using frames recorded by stereo camera through matching scale invariant feature transform (SIFT) keypoints in head areas of both left and right stereo frames. The resulted 3D head trajectories are used for recognizing bed exit activities and the performance is investigated in the thesis.

Chapter 2. Literature review

The existing fall detection approaches fall into three categories, the radar based approaches, vision based approaches and physical sensor based approaches. The literature review for each of the categories is provided as following.

2.1 Radar based approaches

Radar signals can provide valuable human body motion information. Radar based gait characterization using various machine learning algorithms shows effective and satisfying performance [10, 14]. [10] collects data of seven different activities, i.e., running, sitting still, walking, crawling, walking while holding a stick, boxing while standing in place and boxing while moving forward, uses six features to represent the micro-Doppler signatures of the radar signals in their short-time Fourier transform (STFT) results, and utilizes those features to train an SVM for the classification of the seven activities,. In [14], the Doppler signatures of human activities are extracted using the mel-frequency cepstral coefficients (MFCC). Based on these features, two different machine learning algorithms, SVM and k-nearest neighbor (KNN), are employed to detect falls. [5] introduces a continuous wave radar to record radar signatures according to the gait of a walking man. The received signal consists of Doppler shifted signals reflected by legs, arms and torso during human body movements. Short-time Fourier transform and chirplet transform are applied to extract parameters of human gait from the signal.

2.2 Vision based approaches

Vision based approaches are widely used in classifying human falls. [21] develops an human fall detection systems using the combination of human shape variation and motion history images (MHI). An MHI is an image calculated from a sequence of motion images, the pixel intensity of MHI images represent the information of recent motion in the images sequence. In [28] the velocity profile features of normal and abnormal activities are identified. The horizontal and vertical velocities of different body components are measured, which exhibits different characteristic patterns for falls. [2] proposes a fall detector monitoring system using image stream of a thermal camera. Measured vertical velocities of the subject is analyzed by the colored segmentation algorithm, and motion features are identified in different pattern of velocities over time. The authors use a neural network based fall detector to identify the characteristic patterns of human body component velocities during a fall.

The information of posture is valuable during fall detection. [9] applies an SVM to classify balance impairments based on minimum foot clearance (MFC) principle. A two-dimensional (2D) motion analysis system collects the foot clearance data, and a high speed camera records the unobstructed walking. The authors apply wavelet transform to extract the features of MFC signals, which is used for SVM models to classify for an elderly subject as being within healthy range or being a balance impaired. In [7] four types of posture, lying ,sitting, standing and bending, are classified by a fuzzy neural network. After extracting human body from background, projection histograms and discrete Fourier transform are applied to extract feature from the silhouette, which are invariant to size and position of a silhouette. The results show that the four postures can be identified with high accuracy.

2.3 Physical sensor based approaches

The physical sensors include accelerometer sensors, acoustic sensors and vibrational sensors. The accelerometer sensor is widely used in wearable devices for fall detection, the essence of wearable device based methods is that garments with embedded sensors detect the moving motion and position change of the human body. Acoustic sensors and vibrational sensors are usually applied in environmental devices.

Wearable device: Accelerometry is used to measure the accelerations of different human body parts, which is widely applied for measuring the human movement to detect activity patterns. [17] utilizes a waist mounted accelerometry to detect the change of body movements, during a fall the body will suddenly change from an upright position to a lying position and the sharply changing acceleration will be detected by the sensor, and the system will issue a fall alarm. In [11], fall detection system is constituted of a belt-worn sensor composed of three accelerometers to detect human body movement denoted by an acceleration vector in a 3D space, and the vector is classified into different types according to different body activities. [23] introduces a fall detection system consisted of a wearable angular velocity monitor and an airbag. When the values of acceleration and angular velocity exceed a threshold, the system trigger the airbag inflation for protecting people by reducing fall related injuries. In [27] an accelerometer sensor is placed on the head. Comparing a reference velocity, this system is able to distinguish falls form normal daily activities. Changes of body postures and physical activities can lead to physiological fluctuations such as heart rate variation or muscle contracting and stretching. [6] demonstrates a system consists of a motion sensor and several Electromyogram (EMG) sensors. The core device of the motion sensor is a triaxial

accelerometer, and the EMG sensors are used for recording the electric information of muscle during its contractions for performing the motion.

Environmental device: Human activities in daily life also lead to changes of physical environments, such as acoustic and vibrational variations. Those varying physical quantities contain the information for recognizing human activities, based on which many methods are developed for recording and analyzing those physical quantities to classify falls and non-falls. [12] proposes an acoustic-FADE system consist of a circular microphone array and data processing software. The software is firstly applied to confirm the position of the sound source, and then if the height of the location of the sound source is larger than a predefined threshold, it is considered as caused by a non-fall, otherwise the sound signal will be enhanced and extracted its MFCC features, and the a nearest neighbor classifiers will confirm whether it is caused by a fall. In [1] a piezoelectric sensor is attached to the floor surface, which is used for transforming the vibration signals into electrical signals. The different floor vibration pattern will be produced by fall and normal activities, which generate different electrical signal patterns for classification. [30] introduces a system consisting of an accelerometer and a microphone, which are attached to the floor and used for detecting for vibration and sound signals. Using a microphone along with a accelerometer is able to reduce the false alarm rate. [24] introduces a system fusing the multitude of vibration, sound and passive infrared sensors. The features of the sensor measurements are extracted by wavelet analysis and used for hidden Markov Models based classification.

2.4 Discussion

The advantages and disadvantages of existing fall detection technologies are discussed as following.

Radars can acquire data under various weather, lighting and ambience conditions, e.g., in a cloudy day, during day and night, in messy living conditions with obstructing furniture existing, or even through a wall. The radars can collect returns at a longer distance from potential targets, and can cover a large surveillance area. Another unparalleled ability of radar is the high frame rate measuring of velocities without intrusive device to human bodies with a reasonable equipment cost. However, the main disadvantage of most main stream radars is that the radar returns contain information from different human body components, and sophisticated radar signal processing and machine learning techniques are required to extract meaningful inferences from the radar signal returns.

Camera vision based approaches analyses human motion features using recorded images or videos, which contains abundant and micromesh information on the human body motion. With the flourishing of image processing techniques, the camera vision based approaches can be applied to learn, recognize and distinguish the motions with subtle differences. The shortcomings of vision based approaches include privacy issues that the subject under study would perceive the existence of camera and feel uncomfortable when being watched, lighting issues as cameras may require of lighted environments, and also cost issues caused by high price of high frame rate cameras.

Wearable devices are easy to operate and install in general, and the cost efficiency of these devices is obviously resulting in wide application. However, the most significant disadvantage is that these devices are intrusive and need to be fixed physically with the users, which also easily causes disconnection of the device. Also the wearable devices lead to inconvenience of the users in daily life. Environmental sensor based approaches do not need the users to wear the devices so that these approaches can provide a regular and convenient lifestyle. However those methods limit user activities into the areas under monitoring, and also the environmental

devices often generate results with high false alarm rate due to a large number of ambient interference sources.

Based on the characteristic of human fall and bed exit motions, we determine to use radar and stereo imagery based approaches for investigating the motion analysis application, respectively. As we discussed above, radars possesses high sensitivity and a high frame rate in sensing , which is ideal for detecting objects with dramatic velocity changing during a very short time period, also radars can decrease privacy concerns of the patients. For the above reasons, a radar based approach is proposed and implemented for fall detection in this thesis. For bed exit detection, we notice that subtle movements of different human body components need to be analyzed for effective detection and prediction such human motion, which leads us to explorer image based techniques. We also aim to investigate 3D trajectories of human body components which is invariant to view angles, which results in the use of 3D image sensors. We utilize stereo image sensors also with the considerations that stereo image sensors can be used outdoor in direct sunshine. In this thesis, we propose to use both 2D and 3D trajectories of head and shoulder movements to identify bed exit motions based on stereo camera. We also have the considerations for future work to fuse the two modalities of radar and image based sensing techniques, aiming to provide high accuracy in both spatial and time domains with a reasonable device cost so that such equipment can be used for enhancing elderly or patients life quality in their living environments.

Chapter 3. Fall detection based on sequential modeling of radar signal time-frequency features

In this chapter, we utilize the HMM to characterize the time-frequency features of radar signals for the purpose of fall detection, and the feasibility of early warning is investigated. We separately use two time-frequency analysis methods, i.e., matching pursuit decomposition (MPD) and Short-time Fourier transform (STFT), to extract sequential feature vectors from radar signals. The vector sequences are then transformed to symbol sequences by k-means clustering for HMM training and testing, we also train SVM model and classify the experimental data.

This chapter consists of four sections. In Section 3.1, we provide radar signal processing and sequential feature extraction using MPD and STFT. The training and testing of HMMs are described in Section 3.2. Section 3.3 presents SVM based classification and corresponding results. In Section 3.4, we discuss our experimental results.

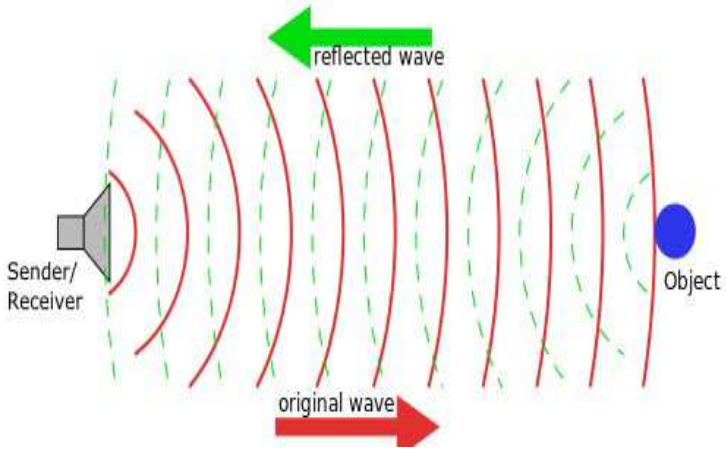


Figure 3.1: Radar application scenario.

3.1 Radar signal processing and feature extraction

We denote the received narrowband radar signal in Figure 3.1 in the continuous-time domain as

$$r(t) = a(t)e^{i\theta(t)}, \tag{3.1.1}$$

where $a(t)$ is the amplitude and $\theta(t)$ is the phase of the received signal.

In our experimental studies, the radar data are collected at the Radar Imaging Laboratory of the Center for Advanced Communications, Villanova University. The record time is 10 second for each experiment. Background subtraction was performed before processing the data. In this case, we apply an elementary exponential averaging, which is a background subtraction technique that subtracts the time invariant background radar signals reflected from surrounding objects, so that only the radar signal reflected by human body is preserved. An example of the radar signals is shown in Figure 3.2. The periodic sampling results in discrete-time observations $r[k] = r(kT)$, where $T = 10^{-3}$ s is the sampling interval.

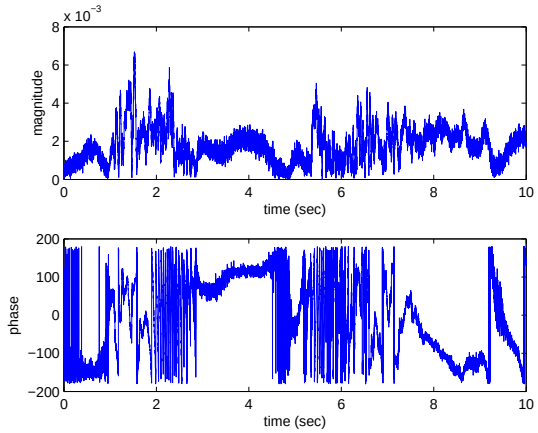


Figure 3.2: Example of the collected radar signal. The upper plot is the amplitude $a[k]$ and the under plot is the phase $\theta[k]$ of a received signal.

3.1.1 Radar signal feature extraction using Matching Pursuit Decomposition

In this section, we use the MPD algorithm to obtain the features of the radar signal $u[k]$ which is the real part of $r[k]$, i.e., $u[k] = \Re\{r[k]\} = a[k]\cos(\theta[k])$, where $\Re\{\cdot\}$ denotes the real part.

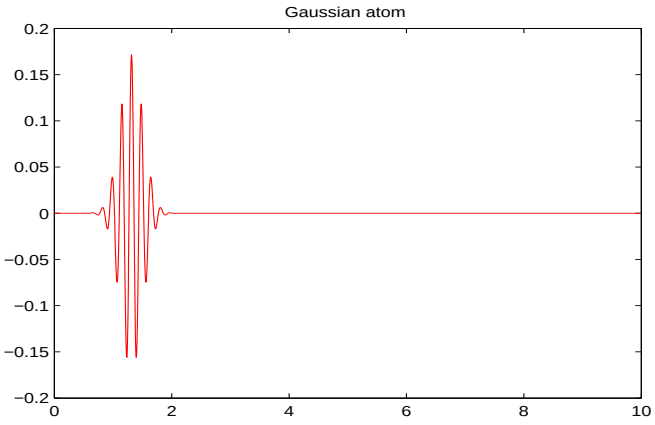
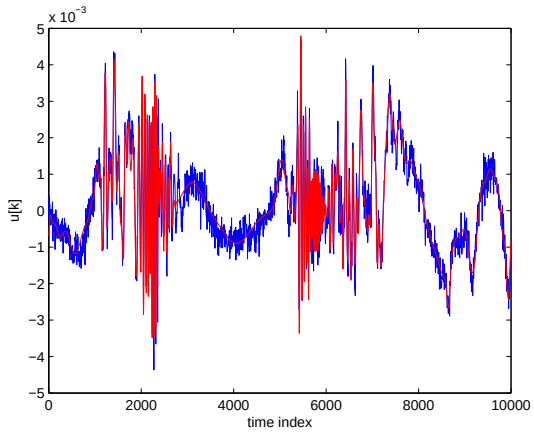
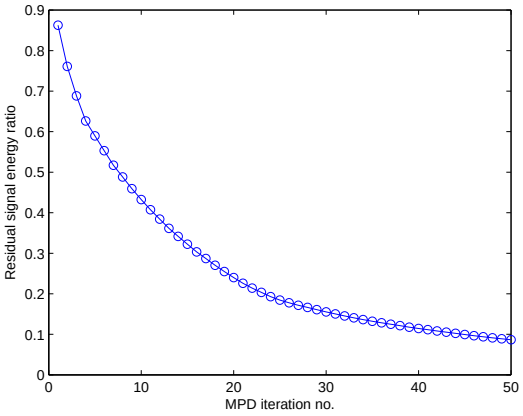


Figure 3.3: A Gaussian atom.



(a)



(b)

Figure 3.4: Signal analysis using MPD. (a) The original signal (blue) vs. the synthesized signal (red) using 50 MPD iterations, and (b) The residual signal energy ratio after each iteration.

The MPD algorithm decomposes the signal into a series of signal atoms which belong to an atom dictionary denoted by $\mathcal{D}$ [16]. Specifically, a Gaussian atom in Figure 3.3 is described as

$$g_{\varphi}[k] = \xi_{\varphi} e^{-\sigma^2(kT-\tau)^2} \cos(2\pi \upsilon kT), \tag{3.1.2}$$

where ξ_{φ} is a normalizing constant for unit energy, $\varphi = \{\tau, \upsilon, \sigma\}$ represents the time-shift, frequency-shift and scaling parameters. The signal $u[k]$ can be decomposed iteratively as

$$u[k] = \sum_{j=0}^{J-1} \omega_j g_{\varphi_j}[k] + r_J[k], \tag{3.1.3}$$

where $r_J[k]$ is the residue after J MPD iterations, and $r_0[k] \equiv u[k]$. The atom used in the jth iteration, $g_{\varphi_j}[k]$, is described by parameter set $\varphi_j = \{\tau_j, \upsilon_j, \sigma_j\}$. The expansion coefficient ω_j is calculated as following:

$$\omega_j = \sum_k r_j[k] g_{\varphi_j}[k], \quad j = 0, ..., J-1. \tag{3.1.4}$$

Combining the atom parameter set φ_j and atom amplitude parameter ω_j results in an expanded parameter vector $\mathbf{e}_j = [\tau_j, \upsilon_j, \sigma_j, \omega_j]^T$ for $1 \leq j \leq J$, where $(\cdot)^T$ denotes matrix or vector transpose. Stacking $\mathbf{e}_j$ for all the J iterations yields the feature sequence, denoted as $\mathbf{E} = [\mathbf{e}_1\ \mathbf{e}_2\ \ldots\ \mathbf{e}_J]$.

In Figure 3.4(a), we compare an original radar signal $u[k]$ (blue curve) with its MPD representation obtained after $J = 50$ iterations of decomposition (red curve), which shows that the MPD can synthesize the radar signal well. Figure 3.4(b) shows that, after 50 iterations, the residual signal energy is below 10% of the original signal energy. The generated atoms are sorted in accordance with time-shift τ from small to large to form the radar signal feature vector sequence $\mathbf{E}$.

3.1.2 Radar signal feature extraction using STFT

During daily activities, velocities of different human body components assume unique patterns. For a monostatic radar, the Doppler principle states that the frequency shift of a radar signal reflected from a target is related with the target velocity as following

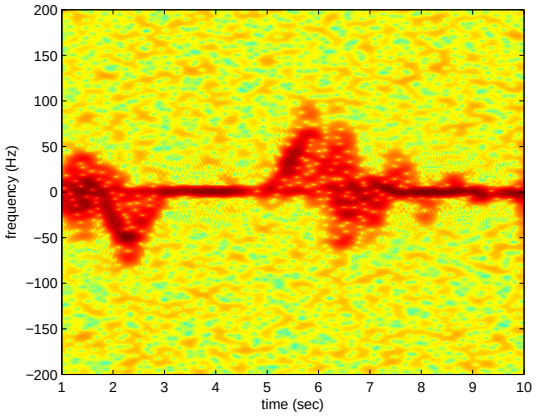
$$\Delta f = \frac{2 f_c \nu}{c} \cos(\alpha), \tag{3.1.5}$$

where Δf is the frequency shift, ν is the velocity of target, c is the speed of light, f_c is the carrier frequency, and α is the angle between the target moving direction and the radar line-of-sight direction. Because falls usually cause a high acceleration, they will generate different Doppler frequency patterns in received signals from those corresponding to non-fall activities.

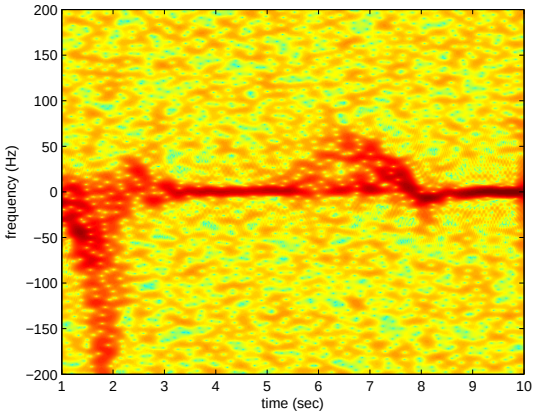
The radar signal $r[k]$ is converted into the time-frequency domain by STFT, given by

$$R(m,n) = \sum_{k=1}^{K} r[k] w^*(kT - mT) e^{-j2\pi nFk}, \tag{3.1.6}$$

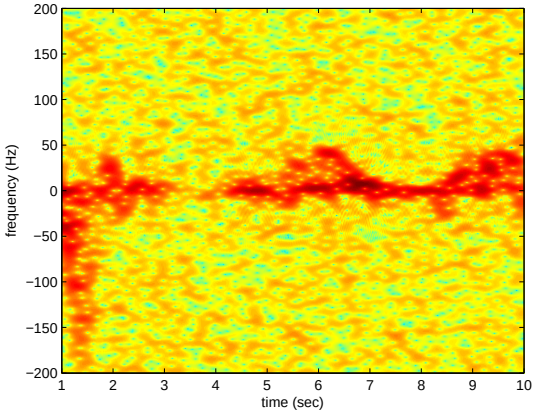
where both m and n are integers, m is the time index, $n = -200, ..., 0, ..., 200$ is the frequency index, $T = 10^{-3}$ s is sampling period, F is frequency step size, $w(\cdot)$ is a window function, and $(\cdot)^*$ denotes complex conjugate. Figure 3.5 show the STFT results of the three human movements, Figure 3.5(a) is the result of sit down and stand up, Figure 3.5(b) is fall backward and Figure 3.5(c). There are two peaks in the Figure 3.5(a), the left peak is overall below the time axe which is caused by human body moving away from radar during sit down, and the right peak is overall positive which is due to human body moving forward to radar during stand up. We can easily identify what are differences between Figure 3.5(a) and Figure 3.5(b) 3.5(c). In both Figure 3.5(b) and 3.5(c) there is a main peak separately which is caused by fall backward, and the amplitude of the peak is larger than



(a)



(b)



(c)

Figure 3.5: STFT Spectrograms of three human movements. (a) sit down and stand up ; (b) fall backward; and (c) fall backward with 45°.

the peak caused by sitting in Figure 3.5(a). This phenomenon verifies the velocity of human body components moving away from radar during fall backward is larger than the velocity of human body components moving away from radar during sitting. Carefully comparing the main peaks in Figure 3.5(b) with Figure 3.5(c), we can see the peak in Figure 3.5(b) is stronger than the peak in Figure 3.5(c), which is cause by the smaller angle of the fall to the line of sight direction of radar. If this angle is smaller, the peak is stronger, which is consistent with Doppler principle.

Because the STFT transforms one-dimensional time-domain data to a two-dimensional representation in the time-frequency domain, the size of the resulting time-frequency matrix is very larger. Figure 3.5 shows the spectrogram matrix $\mathbf{S} = [s_{n,m}]$ whose (n, m)th element $s_{n,m} = |R(m, n)|^2$ has a dimension of 400×10000, where $|\cdot|$ denotes the module operation. To facilitate HMM modeling and classification, we consider below the reduction of the matrix dimension in both time and frequency domains.

Dimension reduction in frequency domain

To choose a frequency range for analysis, we calculate the relative energy burst curves as [14]

$$W_1(m) = \sum_{n=-200}^{-1} s_{n,m}^2 + \sum_{n=1}^{200} s_{n,m}^2, \tag{3.1.7}$$

$$W_2(m) = \sum_{n=-200}^{-25} s_{n,m}^2 + \sum_{n=25}^{200} s_{n,m}^2. \tag{3.1.8}$$

Comparing Figure 3.6(a), 3.6(b) and 3.6(c), we find that the pattern of $W_2(m)$ in Fig. 3(c) is more representative for the spectrogram result depicted in Figure 3.6(a). It means that the Doppler frequency ranges of -200 to -25 Hz and 25 to 200 Hz contain more useful information. As such, spectrogram entries corresponding to Doppler frequency from -24 to 24 Hz are excluded from processing as they contain

high clutter energy. The spectrogram matrix $\mathbf{S}$ of dimension 400×10000 is then converted into matrix $\mathbf{P} = [p_{n,m}]$ of dimension 352×10000.

After the effective frequency range is determined, we further use the PCA for feature dimension reduction. The essence of PCA is that, through an orthogonal linear transformation, data are converted to new orthogonal coordinates which ensure the variance of the projection of the data to the first coordinate to be the greatest, then the second greatest variance to the second coordinate, and so on. Those coordinates with major variances are called the principal components.

We use Ω to denote the frequency range used in the analysis, and Φ to represent the new coordinate space spanned by the principal components of $\mathbf{P}$. The PCA is performed in the following steps.

1) Calculate the mean value z_m of the mth column of $\mathbf{P}$, and subtract it from the mth column. Repeating this procedure for all the columns results in a new matrix $\mathbf{Q} = [q_{i,m}]$ with dimension 352×10000.

2) Calculate the sample covariance matrix $\mathbf{H} = \mathbf{Q}\mathbf{Q}^T$.

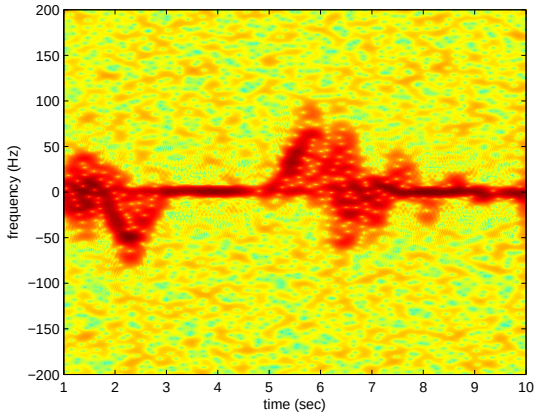
3) Calculate the eigenvalues and the corresponding eigenvectors of $\mathbf{H}$.

4) Select the first n largest eigenvalues sorted in a descending order. The corresponding n eigenvectors constitute a new principal component matrix $\mathbf{Y}$ with dimension $352 \times n$.

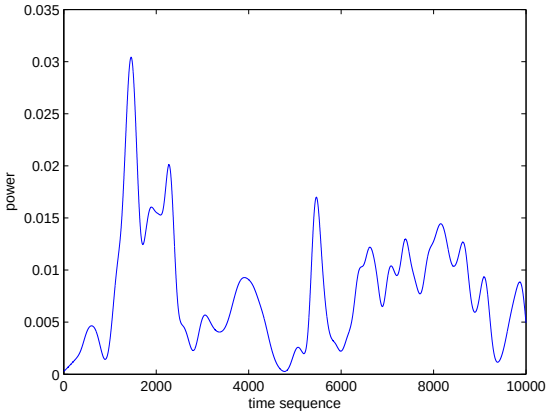
5) Calculate the projection matrix as $\mathbf{B} = \mathbf{Y}^T\mathbf{P}$, where the dimension of matrix $\mathbf{B}$ is $n \times 10^4$.

As a result, we are able to use principal components in space Φ to preserve the information in frequency space Ω.

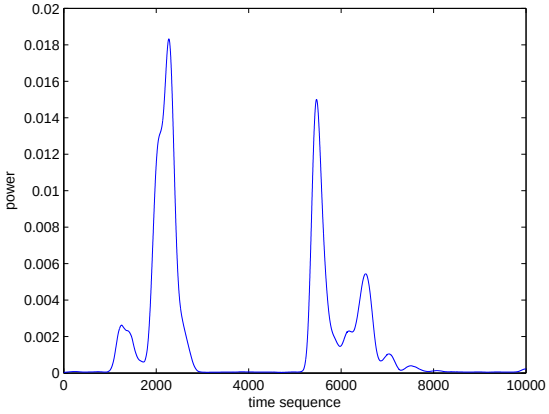
In Figure 3.7, the energy bursts constructed from the first 10 principal components Φ (red curve) are compared with those of the original signal in Ω (blue curve).



(a)



(b)



(c)

Figure 3.6: Spectrogram and energy burst results of the radar data. (a) Spectrogram; (b) Energy burst curve of $W_1(m)$; and (c) Energy burst curve of $W_2(m)$.

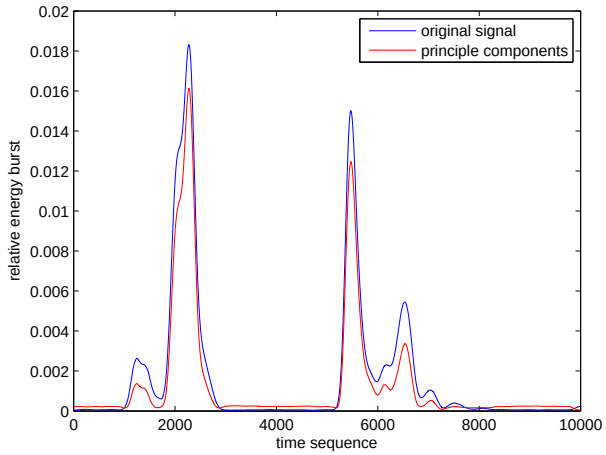


Figure 3.7: Comparison of the energy bust curves.

It is evident that space Φ consisting of the first 10 principal components preserves the information very well. As a result, matrix $\mathbf{P}$ of dimension 352×10000 is reduced to matrix $\mathbf{B} = [b_{k,m}]$ whose dimension is only 10×10000.

Dimension reduction in time-domain

Because the spectrogram transforms the one-dimensional data into a two-dimensional representation, it has a higher correlation in the time-frequency domain. We decimate the spectrogram in the time domain without noticeable change by compromising the quality of time-frequency representations. In the following, we choose to retain 50 samples per second, $\mathbf{B}$ is converted into a new matrix $\mathbf{U} = [\mathbf{u}_j] = [u_{k,j}] = [b_{k,20j}]_{k=1,\ldots,10,j=1,\ldots,500}$ with dimensions 10×500.

After feature reduction in the frequency and time domains as described above, the spectrogram matrix $\mathbf{S}$ with dimensions 400×10000 is transformed to a feature matrix $\mathbf{U}$ with dimensions 10×500, whose columns contain the radar signal time-frequency sequences.

3.2 Hidden Markov model based classification and results

In our study we use discrete HMMs to classify falls and non-falls. Applications of HMMs can recognize various sequential patterns by automatically learning from the data sequence and optimizing model parameters [20,29]. Through the two processing methods described before, we obtain two types of feature sequence, $\mathbf{E}^{\alpha,\beta}$ by MPD and $\mathbf{U}^{\alpha,\beta}$ by spectrogram, where $\alpha = 1, 2, 3$ is the model index, and $\beta = 1, ..., 10$ is the sequence index. We use the k-means clustering [8] to convert a feature vector sequence into a symbol sequence for HMMs.

All the HMM training and testing are accomplished based on the leave-one-subject-out cross validation in this chapter.

3.2.1 Radar data collection

In our field experiment, two subjects perform three types of activities facing a radar system, sit-stand, fall backward to floor along the radar radiation direction, and fall backward to floor with a 45° angle to the radar radiation direction. We collected 5 data sets for each of the three motion models from each subject, 10 data sets for each motion type, and 30 data sets in total. In each data set, 10 s of radar signal returns are collected.

3.2.2 Model training

An HMM λ is described by the following parameters as

$$\lambda = \{\mathbf{A}, \mathbf{B}, \boldsymbol{\pi}\}, \tag{3.2.1}$$

where $\mathbf{A}$ is the state transition matrix, $\mathbf{B}$ is the observation probability matrix, and $\boldsymbol{\pi}$ is the initial state probability distribution vector.

Model training based on MPD features: We cluster the observation vectors generated by the MPD into 25 clusters. A feature vector, $\mathbf{e}_j^{\alpha,\beta}$, is transformed into a feature symbol as following:

$$\begin{aligned} c_j^{\alpha,\beta} &= \arg\min_i \left|\mathbf{e}_j^{\alpha,\beta} - \mathbf{c}_i\right|^2, \\ j &= 1, ..., 50, \ i = 1, ..., 25, \end{aligned} \tag{3.2.2}$$

where $c_j^{\alpha,\beta}$ is the feature symbol corresponding to the MPD feature sequence $\mathbf{e}_j^{\alpha,\beta}$, and $\mathbf{c}_i$ is the centroid of the ith cluster generated by the k-means algorithm. The feature symbol $c_j^{\alpha,\beta}$ is concatenated into a symbol sequence as $\mathbf{c}^{\alpha,\beta} = [c_1^{\alpha,\beta}, ..., c_{50}^{\alpha,\beta}]^T$. We use the symbol sequence to train the corresponding model as

$$\hat{\lambda}_\alpha^{\text{MPD}} = \arg\max_{\lambda_\alpha^{\text{MPD}}} \Pr(\mathbf{c}^{\alpha,\beta}|\lambda_\alpha^{\text{MPD}}), \tag{3.2.3}$$

where $\lambda_\alpha^{\text{MPD}}$ is the HMM corresponding to the αth motion class. After training the HMM with the symbol sequences, we obtain the trained model parameter set $\lambda^{\text{MPD}} = \{\lambda_1^{\text{MPD}}, \lambda_2^{\text{MPD}}, \lambda_3^{\text{MPD}}\}$.

Model training based on STFT: We cluster the observation vectors generated by the spectrogram into 50 clusters. A feature vector, $\mathbf{u}_j^{\alpha,\beta}$, is transformed into a feature symbol as following:

$$\begin{aligned} z_j^{\alpha,\beta} &= \arg\min_i \left|\mathbf{u}_j^{\alpha,\beta} - \mathbf{z}_i\right|^2, \\ j &= 1, ..., 500, \ i = 1, ..., 50, \end{aligned} \tag{3.2.4}$$

where $z_j^{\alpha,\beta}$ is the feature symbol corresponding to the processed STFT feature sequence $\mathbf{u}_j^{\alpha,\beta}$, and $\mathbf{z}_i$ is the centroid of the ith cluster generated by the k-means

algorithm. The feature symbol $z_j^{\alpha,\beta}$ is concatenated into a symbol sequence as $\mathbf{z}^{\alpha,\beta} = [z_1^{\alpha,\beta}, ..., z_{500}^{\alpha,\beta}]^T$. We use the symbol sequences to train corresponding model as

$$\hat{\lambda}_\alpha^{\text{STFT}} = \arg \max_{\lambda_\alpha^{\text{STFT}}} \Pr(\mathbf{z}^{\alpha,\beta} | \lambda_\alpha^{\text{STFT}}), \tag{3.2.5}$$

where $\lambda_\alpha^{\text{STFT}}$ is the HMM corresponding to the αth motion class. After training the HMM with the symbol sequences, we obtain the trained model parameter set $\lambda^{\text{STFT}} = \{\lambda_1^{\text{STFT}}, \lambda_2^{\text{STFT}}, \lambda_3^{\text{STFT}}\}$.

3.2.3 HMM based classification

Classification based on MPD features: The classification of the motion classes based on the MPD feature vectors is given as

$$\hat{\alpha} = \arg \max_{\alpha} \Pr(\mathbf{c} | \lambda_\alpha^{\text{MPD}}), \tag{3.2.6}$$

where $\mathbf{c}$ is a feature sequence from an unknown motion class.

Classification based on STFT: the classification of the motion classes based on the STFT feature vectors is given as

$$\hat{\alpha} = \arg \max_{\alpha} \Pr(\mathbf{z} | \lambda_\alpha^{\text{STFT}}), \tag{3.2.7}$$

where $\mathbf{z}$ is a feature sequence from an unknown motion class.

In this paper, we also present results of early fall detection, where we use the first sections of the feature sequence with various lengths to predict and detect falls.

3.2.4 HMM based classification results

The radar signal data are collected using a narrowband pulse-Doppler radar implemented using a network analyzer, which is operated at a carrier frequency of

8 GHz. The network analyzer is triggered at 1 kHz sampling rate, and the record time of each experiment is 10 seconds. We build and test the HMM algorithm based on the following human motion classes: 1) repeated sitting and standing, 2) falling backward to floor along the radar radiation direction, and 3) falling backward to floor with a 45° angle to the radar radiation direction. Each category of motion class has 10 experimental data sets.

Tables 3.1 and 3.2 demonstrate the confusion matrix of the HMM based classification by respectively exploiting the MPD and STFT features. In Table 3.1, the element of the confusion matrix corresponding to $\mathbf{F}^{i,\beta}$ and λ_j^{MPD} represents the ratio of the feature vectors from motion class i being classified into class j. The entries in Table 3.2 are similarly defined. Tables 3.1 and 3.2 clearly show that both MPD and STFT extract the features sequences for the HMM-based methods to yield reliable classification results.

Table 3.1: HMM recognition results based on MPD features

	λ_1^{MPD}	λ_2^{MPD}	λ_3^{MPD}
$\mathbf{F}^{1,\beta}$	0.9	0	0.1
$\mathbf{F}^{2,\beta}$	0	1	0
$\mathbf{F}^{3,\beta}$	0	0.1	0.9

Table 3.2: HMM recognition results based on STFT features

	$\lambda_1^{\mathrm{STFT}}$	$\lambda_1^{\mathrm{STFT}}$	$\lambda_1^{\mathrm{STFT}}$
$\mathbf{U}^{1,\beta}$	1	0	0
$\mathbf{U}^{2,\beta}$	0	1	0
$\mathbf{U}^{3,\beta}$	0.1	0	0.9

Next, we explore the early fall detection with the observation sequences $\mathbf{U}^{\alpha,\beta}$ extracted from the spectrogram. From the 10-second data in each collected data set, we separately use the first k seconds to study the classification performance.

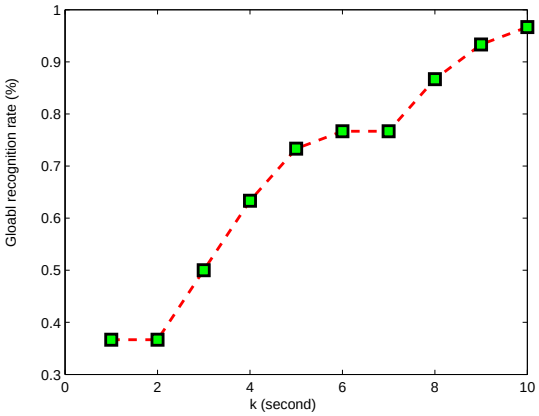


Figure 3.8: The global recognition rate using initial segments of the feature symbol sequence.

The global recognition rate is defined as

$$c(k) = \frac{N_c(k)}{N}, \tag{3.2.8}$$

where $N_c(k)$ is the number of correct recognitions using the first k seconds of data, and N is the total number of recognitions. The results are shown in Figure 3.8 as a function of k.

3.3 Support vector machine based classification and results

3.3.1 SVM classification

In this section, we present a brief introduction of SVM, which is widely applied for classification and regression analysis. The form of SVM's training data can be denoted as $G = \{(\mathbf{x}_i, y_i), i = 1, ..., M\}$, where $\mathbf{x}_i = [x_{i,1}, x_{i,2}, ..., x_{i,m}]^T \in \Re^m$ is a feature vector and $y_i \in \{+1, -1\}$ denotes a class label of the vector $\mathbf{x}_i$. Kernel-based SVM can separate the set of training data G into two classes by a hyperplane. In

our paper, we use the the C-SVM model to find the optimal hyperplane by solving the following minimization problem [13]:

$$\begin{aligned} \min_{\mathbf{w},b,\xi_i,i=1,...,M} \quad & \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{M}\xi_i \\ \text{subject to} \quad & y_i(\mathbf{w}\cdot\mathbf{x}_i) + b \geq 1 - \xi_i, \\ & \xi_i \geq 0, \end{aligned} \tag{3.3.1}$$

where $\mathbf{w}$ is the direction vector of a optimal separation hyperplane, $C > 0$ denotes the penalty parameter of the error term, and $\xi_i, i = 1, ..., M$, are slack variables. Minimizing the object function yields the optimal separation hyperplane. Thus, for test data vector $\mathbf{x}$, the decision function is expressed as

$$\sum_{i\in Z} y_i\alpha_i\langle\mathbf{x}\cdot\mathbf{x}_i\rangle + b = 0, \tag{3.3.2}$$

where Z is a subset of the indices $\{1, ..., M\}$, which means that not all of training data G can be the support vectors, and α_i is the Lagrangian multiplier. Substituting $\langle\mathbf{x}\cdot\mathbf{x}_i\rangle$ by the kernel function $K(\mathbf{x}, \mathbf{x}_i)$ in the above expression results in the following decision function:

$$f(\mathbf{x}) = \text{sgn}\left(\sum_{i\in Z} y_i\alpha_i K(\mathbf{x}, \mathbf{x}_i) + b\right). \tag{3.3.3}$$

We use the MPD features of the radar signal as the training and testing data for the SVM. The MPD feature vector can be denoted as a 1×200 row vector $\mathbf{x}_i = [\tau_1\ \nu_1\ \sigma_1\ \omega_1\ \tau_2\ \nu_2\ \sigma_2\ \omega_2\ \ldots\ \tau_{50}\ \nu_{50}\ \sigma_{50}\ \omega_{50}]$.

3.3.2 SVM based results

In this thesis Gaussian radial basis function is used as the kernel for SVM classification. The obtained results are summarized in Table 3.3 with Type 1 denoting the repeated sitting and standing, Type 2 falling backward to floor along the radar

radiation direction, and Type 3 falling backward to floor with a 45° angle to the radar radiation direction.

Table 3.3: SVM classification results based on MPD features

	Type 1	Type 2	Type 3
Type 1	0.7	0	0.2
Type 2	0.1	0.8	0.1
Type 3	0.1	0.3	0.6

3.4 Discussion

In Figure 3.8, we regard the entire 10 seconds of the radar recording as an event, and use the early information of an event to forecast which motion is going to take place. As we can see from Figure 3.8, the results improve as more number of observations of an event becomes available. A 75% forecast accuracy is achieved using the information of the first 5 seconds, whereas the accuracy is increased to 85% when the first 8 seconds of the observed data are used.

In Table 3.3, the correct recognition rate of Type 3 is the worst compared with other two types. Type 3 is easily mis-classified as Type 2, which is caused by similarity of human activities, falling backward and falling backward with 45° angle, by contrast, Type 2 is not similar to Type 1 and 3, so they can be distinguished with a high recognition rate. Overall, the average correct classification rate based on SVM is lower than that based on HMM, the reason is that the feature extracted from radar signal is a sequence in time domain and the HMM based approach is more suitable to describe and analyze time-varying and variational state based sequential features.

Chapter 4. Bed exit detection based on sequential modeling of 3D head movement trajectories

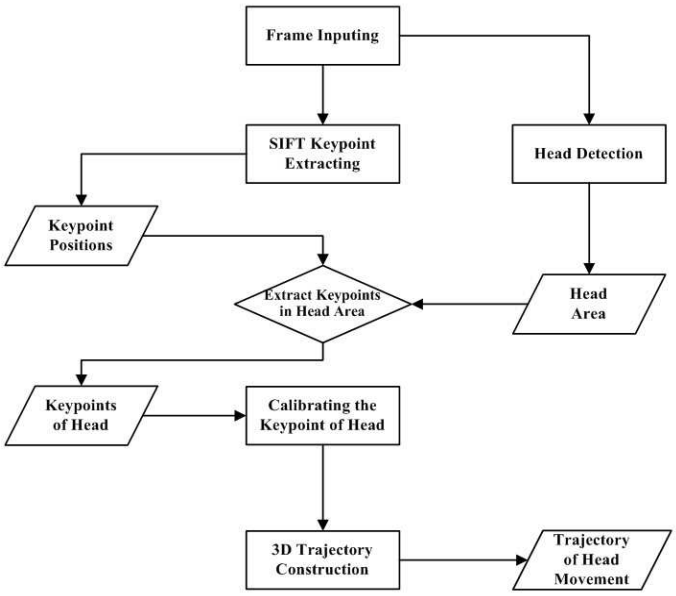


Figure 4.1: The algorithm and data flow of stereo vision motion tracking.

This chapter consists of six sections. Section 4.1 introduces the procedure of calibrating the stereo camera used in this project. Section 4.2 introduces keypoint matching based on SIFT features in stereo images. In Section 4.3, a procedure for training a feature model for human head areas based on HOG features is explained.

In Section 4.4 the 3D head movement trajectories are constructed and evaluated. In section 4.5 HMM based bed exit detection is introduced using 2D and 3D trajectories of head movements. In section 4.6, the discussion of the results is presented.

4.1 Camera calibration

Camera calibration is to confirm the intrinsic parameters which describe camera's geometric and optical characteristics, and the extrinsic parameters representing the relationship between the camera frame's 3D position, orientation and a world coordinate system. The camera calibration is, based on known image coordinates and a 3D world coordinate, to solve the model parameters of the camera including intrinsic and extrinsic parameters [25, 26].

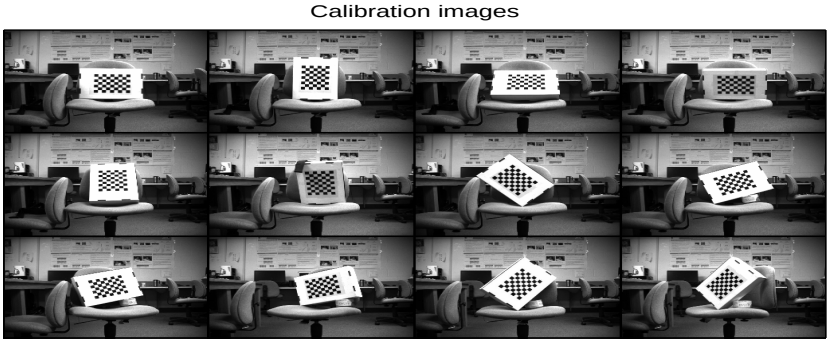


Figure 4.2: The twelve input images with the chessboard for camera calibrating.

In our project, we utilize the Matlab camera calibration toolbox developed by Bouguet. We complete the calibration based on 12 images of an $A3$ size chessboard. Figure 4.3 shows the spatial relationship between left camera and right camera, and

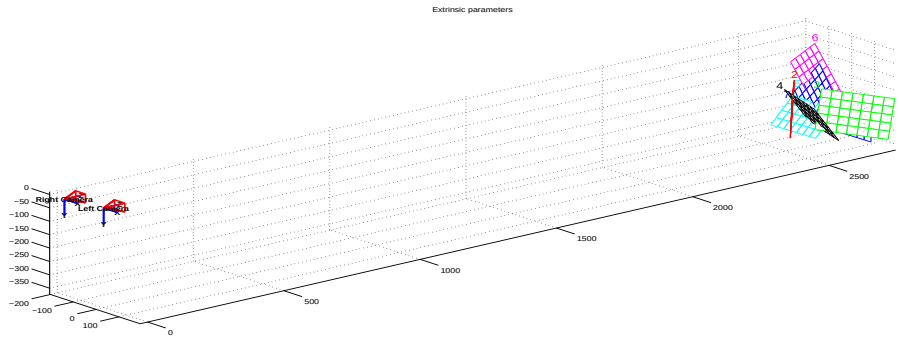


Figure 4.3: The spatial relationship between the stereo camera and the chessboard for calibration.

also the position and orientation of chessboard in camera coordinates. The resluts of calibration will be saved for latter stereo triangulation.

4.2 Keypoints matching based on scale invariable feature transform

The SIFT features own the following characteristics. (1) SIFT features are the local features of an image, which keep invariant to rotation, translation, and image scaling, and are highly robust to viewpoint and illumination changes. (2) The quantity of the features are very large even in a image with a few objects. (3) The SIFT features have a high expansibility that they can be combined with other type feature vectors. Those characteristics make the SIFT features to be the best choice to identify common feature points between left and right images from a stereo camera. Feature matching based on SIFT features can be described as following [15]: (1) Constructing scale space, (2) space-scale extrema detection, and recording the localizations and scales of keypoints, (3) assigning orientation

parameters of keypoints, (4) extracting descriptor on keypoints, and (5) matching keypoints through the descriptors.

In Figure 4.4, 246 keypoints are extracted in the image from the left camera. In Figure 4.5, 137 keypoints are extracted in the image from the right camera.

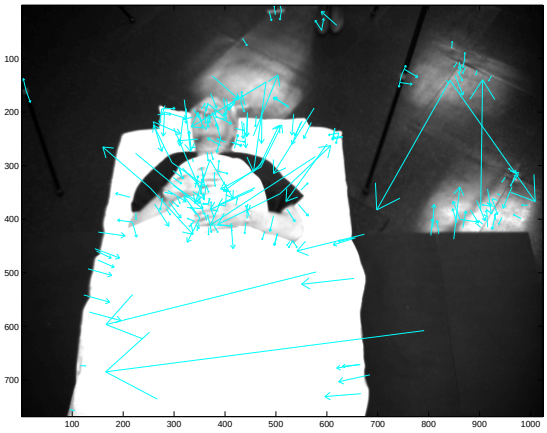


Figure 4.4: SIFT keypoints extracted in the image of left camera.

After extracting the keyoints in the image, the parameters of each keypoint's position, magnitude, orientation were recorded, and the next step is to extract keypoint descriptors. In order to obtain orientation invariance, SIFT rotates the coordinates relative to the keypoint orientation, and the descriptor, or SIFT feature vector, is with 128 elements obtained from 4×4 histogram in the 16×16 pixel region. Each the histogram contains 8 directions is the projections from the 4×4 pixel region.

After obtaining all the keypoint descriptors in the left and right camera images separately, we can use the Euclidean distance of the keypoint descriptors as the measure of keypoint similarity. We assume that point A is a keypoint in the left camera frame, and we need to find two keypoints in the right camera frame with the smallest Euclidean distances d_1 and d_2 to the point A. If $\frac{d_1}{d_2}$ is less than a threshold

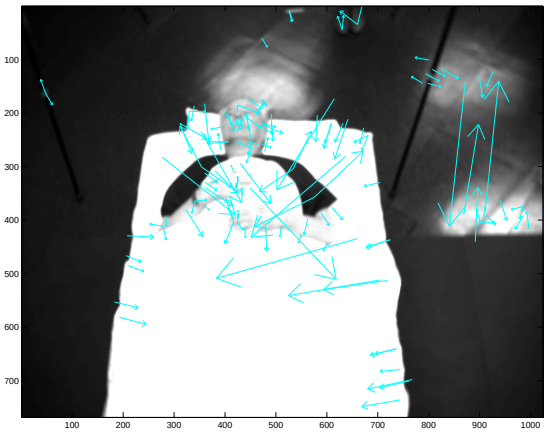


Figure 4.5: SIFT keypoints extracted in the image of right camera.

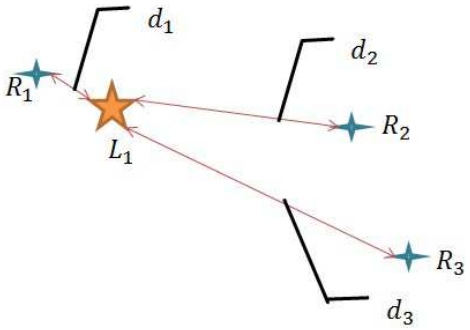


Figure 4.6: Match of a keypoint pair.

value, the keypoint with the smallest Euclidean distance d_1 can be considered as the matched keypoint for A.

In Figure 4.6 we assume L_1 is the keypoint in the image of left camera, and R_1, R_2 and R_3 are the keypoints in the image of the right camera. Firstly, we find two keypoints from R_1, R_2 and R_3 with the smallest distance to L_1, and assuming they are R_1, R_2, and the distances are d_1 and d_2. Then we calculate the ratio of d_1 and d_2, if the ration is less than a threshold value, we can confirm this point R_1 is matched with L_1.

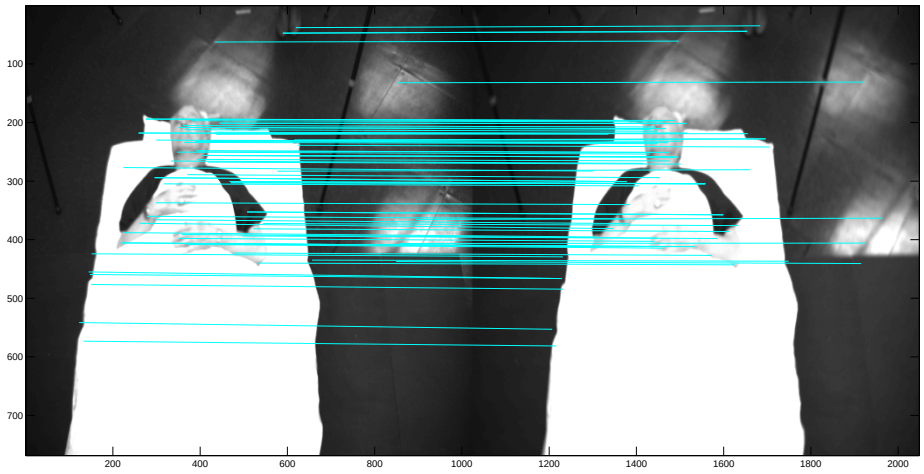


Figure 4.7: Feature points matching of left and right camera images based on SIFT.

Figure 4.7 shows the result of keypoints matching. Totally 75 keypoint pairs are matched which are connected with cyan lines.

4.3 Training and testing human head detector with HOG features

4.3.1 Train a head detector with HOG features

The method of extracting HOG [4] features is similar to SIFT, however, the HOG descriptors are extracted from a dense grid of cells with same pixel size without smoothing in the Gaussian scale space.

The 3D head position is crucial information in bed exit detection, and the head recognition is the foundation for 3D trajectory of head movements. Matlab offers some detectors that could be used for head detection, for example, detectors for Frontal Face, Upper Body, Eye pair, Single Eye, Profile Face, Mouth and Nose,

however those detectors can not be applied in our bed exit project because they are trained with general purposes and are not consistent with the view angle in our application. In our bed exit detection application, the view angle of the cameras to subjects' face, eyes, mouth and nose changes unpredictably when subject is rolling in the bed or getting out of the bed, so it is necessary to train a customized detector to identify the head. We utilize the Computer Vision System Toolbox in Matlab R2013a to train a custom classifier. Three options, Haar, Local Binary Patterns(LBP), and HOG are provide as the features for the detector. HOG features are better for acquiring the edge or shape of an object [9]. The Computer Vision System Toolbox offers a concise and intuitionistic method to train custom detector as the following steps, and Figure 4.9 demonstrates the GUI interface.

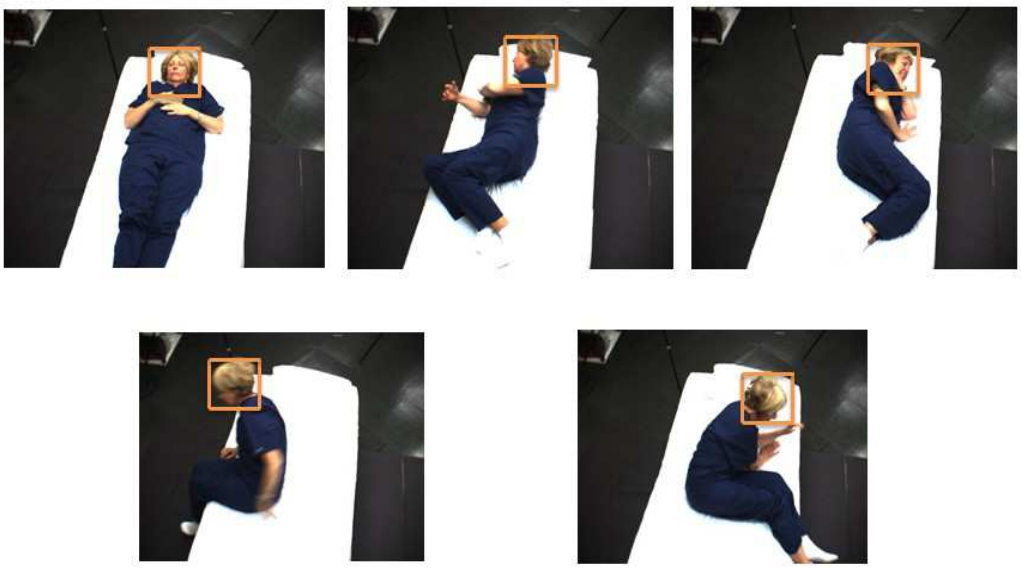


Figure 4.8: Five types of positive images.

(1) Select positive images: The positive images used for training are selected from the videos recording various postures of the subject, including lying, lying to the left side, lying to right side on the bed, getting out of the bed from left-side and

getting out of the bed from right-side. The target object for detection is the area including head and shoulder,

(2) Select Region of Interest (ROI) : The ROI is selected as the area of head and shoulder,

(3)Select negative images : The negative images are generated using positive images with ROI deleted from the images.

(4)Train the detector : The detector is trained with the following parameters: Per-stage True Positive Rate 0.995, Number of Cascade stage 8, Feature Type HOG.

(5)After training the detector, an xml file is generated which can be used for recognizing head area.

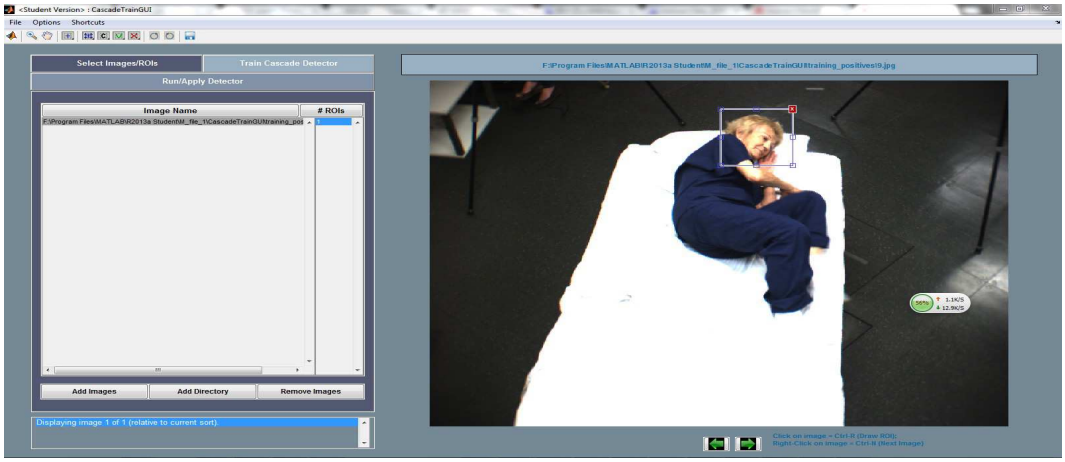


Figure 4.9: The GUI interface for training.

4.3.2 Data association

During detection, false alarms interfere correct decisions. We apply a simplified data association algorithm to overcome this problem. The detected results of the nth frame are recorded in a set $\mathbf{R}_n = \{\mathbf{r}_j^n\}$ including false alarms and one correct detection, the vector $\mathbf{r}_j^n = [r_{1,j}^n \; r_{2,j}^n]$ is the position of the jth detected object in the nth frame, $r_{1,j}^n$ and $r_{2,j}^n$ are the x coordinate and y coordinate, respectively. Let

$\mathbf{r}_c^n$ represent the position of the correct object in the nth frame. The relationship between $\mathbf{r}_c^{n-1}$ and $\mathbf{r}_c^n$ should satisfy as follow,

$$\begin{aligned} \hat{\mathbf{r}}_c^n &= \arg\min_{\mathbf{r}_j^n} \left|\mathbf{r}_j^n - \mathbf{r}_c^{n-1}\right|^2 \\ \mathbf{r}_j^n &\in \mathbf{R}_n \end{aligned} \tag{4.3.1}$$

where $\hat{\mathbf{r}}_c^n$ is the estimate of $\mathbf{r}_c^n$. This simplified method works well for our scenarios because of the slow movement of subjects' heads. For more complicated head movements, more advanced data association algorithms will be considered.

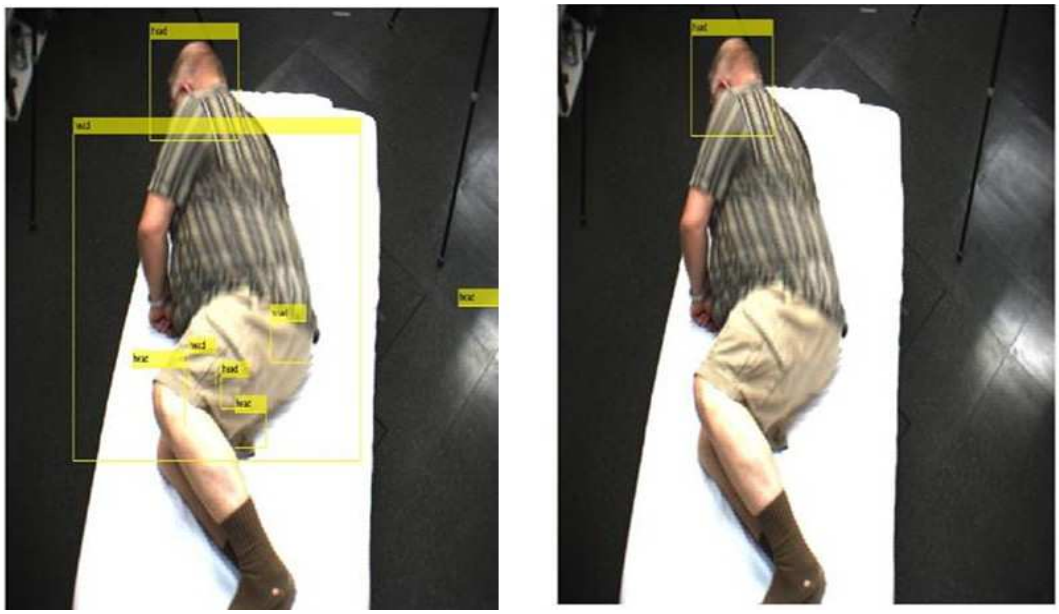


Figure 4.10: Data association for eliminating false alarms.

4.4 Stereo vision motion tracking

4.4.1 3D head position estimation

Stereo vision can determine the 3D location of an object in a scene by comparing images of two separated cameras.

In the research work of the thesis, stereo camera is used to record the human activities of rolling in bed or getting out of bed and then the acquired stereo images are used for extracting the 3D trajectory of the target objects. To achieve this goal, first we need to calibrate the stereo camera, then extract the SIFT feature keypoints in both the left and right images, and match the keypoints. Specifically, the detailed process is described as following. A sample tracking result is displayed in Figure 4.11.

1) Collect videos using stereo camera, perform camera calibration.

2) Detect the head area using the trained custom detector, perform data association and the results are displayed with a red boundary in the left image.

3) Extract feature points within the head area with SIFT algorithm, which are displayed with red points on both of the left and right images.

4) Based on the extracted feature keypoints and stereo camera parameters, calculate the 3D trajectory of the head movement using the mean of the 3D point positions. The generated trajectories are displayed in the lower figures in Figure 4.11. The color of trajectory changes with time, and the same color represents the same time.

4.4.2 Position estimation performance

To evaluate the performance of the 3D position estimation, we use a cuboid as ground truth for evaluation. In Figure 4.12, the four points, A, B, C, D of the cuboid are marked with different colors, the length AC is 29.7 cm, and the width AD is 21.3 cm, the height AB is 14.2 cm. During the evaluation we test performance when it is 1 m, 1.5 m and 2 m far away from camera.

The size of the cuboid is measured as following: firstly the disparities of those 4 points in stereo images with the size of pixel 0.352 mm on a reference image plane, then the 3D coordinates of those points were calculated by stereo triangulation based

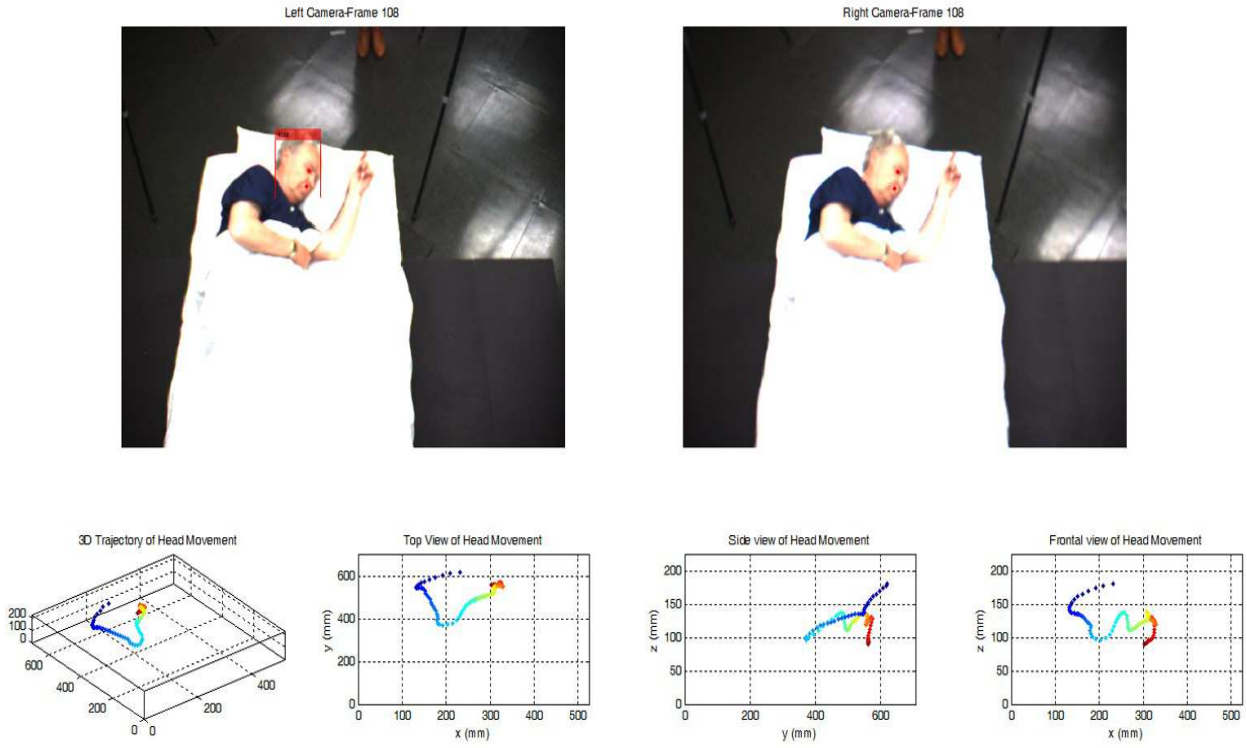


Figure 4.11: A 3D head movement trajectory extracted from a stereo video.

on the disparities and stereo camera parameters, at last the Euclidean distances is calculated between those points with their 3D coordinates. The evaluation results are shown in Table 4.1.

Table 4.1: The measured size of the box compared with the actual size

	AC	AD	AB
Actual value	29.7 cm	21.3 cm	14.2 cm
Distance 1 m	32.2	20.1	15.1
Distance 1.5 m	31.8	19.3	16.2
Distance 2 m	33.2	18.8	16.7
Average value	32.4	19.4	16

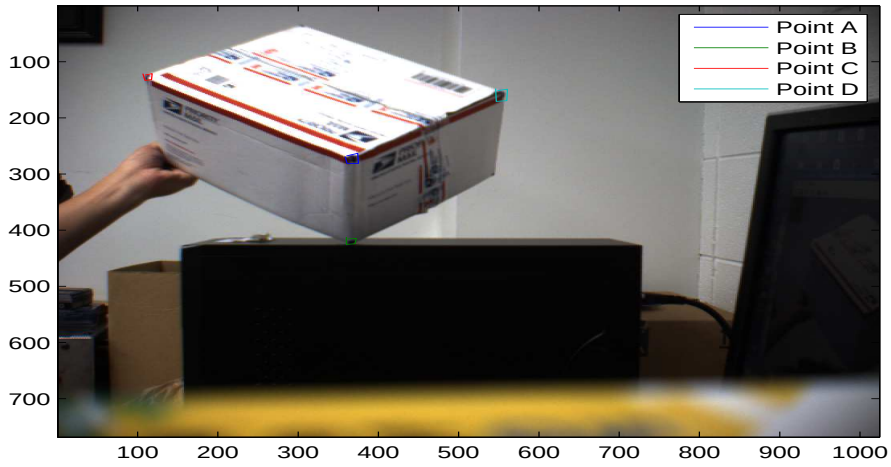


Figure 4.12: The ground truth cuboid.

Table 4.1 shows us the cuboid size and the measured values. The mean of the dimension measurments is 32.4 cm in length, 19.4 cm in width, and 16 cm is height. The deviation are 2.7 cm, 1.9 cm and 1.8 cm, respectively.

4.5 HMM based bed-exit detection using 2D and 3D trajectories of head movements

Following a similar procedure as in chapter 3, we obtain the HMM based classification using 2D and 3D head movement trajectories. We use the Leave-One-Subject-Out cross-validation method in the thesis training and testing the HMMs. The types and the number of the motion videos are described in Table 4.2, a dataset

Table 4.2: The types and number of motion videos

	Bed exit left side	Bed exit right side	Rolling in the bed
Number of action	18	16	29

of 63 videos of three different motions performed by five subjects is used for training and testing, the three types of human motions are, bed exit at left side, bed exit at right side, and rolling in the bed.

Table 4.3: HMM classification results based on 2D trajectories of head movements

	Bed exit left side	Bed exit right side	Rolling in the bed
Bed exit left side	94.44%	0	5.56%
Bed exit right side	12.5%	87.5%	0
Rolling in the bed	13.79%	17.24%	68.97%

Table 4.4: HMM classification results based on 3D trajectories of head movements

	Bed exit left side	Bed exit right side	Rolling in the bed
Bed exit left side	88.9%	0	11.1%
Bed exit right side	18.75%	81.25%	0
Rolling in the bed	10.34%	3.45%	86.21%

Table 4.3 shows the HMM classification results based on 2D trajectories of head movements, Table 4.4 show the HMM classification results basd on 3D trajectories of head movements.

4.6 Discussion

In Table 4.3 we notice that the recognition rates of bed exit at left side and bed exit at right side are 94.44% and 87.5% which are higher than the recognition rate of rolling in bed. This is due to the reason that the head movements are obvious during the bed exit that the subjects' heads will move from the middle to the either of the two sides. The recognition rate of rolling in the bed is less than 70%, which may cause many false alarms of bed exit during human body rolling in the bed. The reason is that head movements of rolling is similar with the head movements of bed

exit in 2D views, for example, in face up rolling to left and face up exiting left, the head will move toward left side in the image in both scenarios.

Comparing Table 4.4 with Table 4.3, the recognition rates of bed exit left and right side decrease, however the recognition rate of rolling in the bed increases from 68.97% to 86.21%, due to the reason that the head position is close to bed in 3D view during rolling, while the head moves away from bed during bed exit. In the 3D view based classification, the issue caused by the similar head movements between face up rolling to left and face up exiting left can be solved. The recognition rates of bed exit decreases, however, we are not able to explain the detection rate decrease. The decrease in detection rates may be due to inaccuracy of head trajectory or insufficient number of training data sets.

The average of recognition rate based on 2D trajectories is 83.63%, the average based on 3D trajectories is 85.45%. The performance of both 3D based and 2D based classification can be improved using a more precise and robust detection and tracking approach.

Chapter 5. Conclusions and future work

5.1 Conclusions

In this thesis, two human motion analysis approaches are developed based on sequential modeling of radar signal and stereo image features, respectively. The target applications of the developed approaches include detection of elderly falls and prediction of bed exit for fall injury prevention.

The radar-based fall detection technique is developed based on time-frequency analysis and hidden Markov models (HMMs). We first presented two techniques, the matching pursuit decomposition and the STFT, for extracting features from radar signals reflected by human body component to detect falls. Based on these features, HMM based classification approaches are utilized for classifying fall and non-fall activities. The classification results show that the proposed radar-based approach is very effective in detecting and classifying human falls. We also investigate early indication of a fall event and demonstrated promising results. The support vector machine (SVM) is also used for solving this problem, and the resulting recognition rate using matching pursuit decomposition (MPD) features shows that the SVM is also an effective classifier for the fall detection application.

A stereo image based approach for detection and prediction of bed exit activities is proposed and investigated in this thesis. This approach combines techniques from image processing, tracking, stereo vision, and machine learning aiming at bed

exit detection using 3D human head trajectories. We train a head detector using HOG features of head-shoulder areas. The head detector is used to obtain the 2D trajectory of head movement. In the meanwhile 3D trajectories of head movements are constructed from stereo images using SIFT keypoint matching, HMM are trained based on both 2D and 3D trajectory of head movement, and respectively the results of classification are investigated and compared. We conclude that this proposed approach is an effective classifier for bed exit activities and further improvement can be expected.

5.2 Future work

Although the target applications of the proposed approaches are fall detection and bed exit prediction, the technical framework that comes with the research can be extended to various human motion analysis, such as markerless human posture analysis, sport performance evaluation, injury rehabilitation evaluation, etc. Some future work we propose is summarized as the following:

- Fusion of radar signal features and stereo image features for high performance in human motion analysis
- Improvement of image based detection for better trajectory extraction and classification
- Synchronization of the stereo vision camera and the vicon motion capture system for accurate system performance evaluation

Bibliography

[1] M. Alwan, P.J. Rajendran, S. Kell, D. Mack, S. Dalal, M. Wolfe, and R. Felder. A and passive floor-vibration based fall detector for elderly. In *2nd Information and Communication Technologies, 2006. ICTTA '06*, volume 1, pages 1003–1007, 2006.

[2] PA. Bromiley, P. Courtney, and NA. Thacker. Design of a visual system for detecting natural events by the use of an independent visual estimate: a human fall detector. *Empirical Evaluation Methods in Computer Vision*, 50, 2002.

[3] E. Capezuti, L. M. Wagner, B. L. Brush, M. Boltz, S. Renz, and K. A. Talerico. Consequences of an intervention to reduce restrictive side rail use in nursing homes. *Journal of the American Geriatrics Society*, 55(3):334–341, 2007.

[4] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 886–893, 2005.

[5] J.L. Geisheimer, W.S. Marshall, and E. Greneker. A continuous-wave (CW) radar for gait analysis. In *Thirty-Fifth Asilomar Conference on Signals, Systems and Computers*, volume 1, pages 834–838, 2001.

[6] H. Ghasemzadeh, R. Jafari, and B. Prabhakaran. A body sensor network with electromyogram and inertial sensors: multimodal interpretation of muscu-

lar activities. *IEEE Transactions on Information Technology in Biomedicine*, 14(2):198–206, 2010.

[7] C. F. Juang and C. M. Chang. Human body posture classification by a neural fuzzy network and home care system application. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, 37(6):984–994, 2007.

[8] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu. An efficient k-means clustering algorithm: analysis and implementation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24:881–892, 2002.

[9] A. H Khandoker, D. T. Lai, R. K. Begg, and M. Palaniswami. Wavelet-based feature extraction for support vector machines for screening balance impairments in the elderly. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 15(4):587–597, 2007.

[10] Y. Kim and H. Ling. Human activity classification based on Micro-Doppler signatures using a support vector machine. *IEEE Transactions on Geoscience and Remote Sensing*, 47:1328–1337, 2009.

[11] Suhuai L. and Qingmao H. A dynamic motion pattern analysis approach to fall detection. In *2004 IEEE International Workshop on Biomedical Circuits and Systems*, pages pp. 2.1–5, 2004, 2004.

[12] Yun L., K.C. Ho, and M. Popescu. A microphone array system for automatic fall detection. *IEEE Transactions on Biomedical Engineering*, 59(5):1291–1301, 2012.

[13] C. J. Lin. A formal analysis of stopping criteria of decomposition methods for support vector machines. *IEEE Transactions on Neural Networks*, 13:1045–1052, 2002.

[14] L. Liu, M. Popescu, M. Skubic, M. Rantz, T. Yardibi, and P. Cuddihy. Automatic fall detection based on Doppler radar motion signature. In *Proceedings of International Conference on Pervasive Computing Technologies for Healthcare*, Dublin, Ireland, May 2011.

[15] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, November 2004.

[16] S. G. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 41:3397–3415, 1993.

[17] M. J. Mathie, A. C. Coster, N. H. Lovell, and B. G. Celler. Accelerometry: providing an integrated, practical method for long-term, ambulatory monitoring of human movement. *Physiological measurement*, 25(2):1–20, 2004.

[18] C. G. Moran, R. T. Wenn, M. Sikand, and A. M. Taylor. Early mortality after hip fracture: is delay before surgery important?. *Journal of Bone and Joint Surgery*, 87:483–489, 2005.

[19] S. L. Murhy. National vital statistics reports. Technical report, National Center for Health Statistics, 2000.

[20] L. R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77:257–286, 1989.

[21] C. Rougier, J. Meunier, A. St-Arnaud, and J. Rousseau. Fall detection from human shape and motion history using video surveillance. In *21st International Conference on Advanced Information Networking and Applications Workshops*, volume 2, pages 875–880, 2007.

[22] S. Sadigh, A. Reimers, R. Andersson, and L. Laflamme. Falls and fall related injuries among the elderly: a survey of residential-care facilities in a swedish municipality. *Journal of Community Health*, 29:129–140, 2004.

[23] T. Tamura, T. Yoshimura, M. Sekine, M. Uchida, and O. Tanaka. A wearable airbag to prevent fall injuries. *Information Technology in Biomedicine, IEEE Transactions on*, 13(6):910–914, 2009.

[24] B.U. Toreyin, E.B. Soyer, I. Onaran, and A.E. Cetin. Falling person detection using multi-sensor signal processing. In *IEEE 15th Signal Processing and Communications Applications*, pages 1–4, 2007.

[25] R.Y. Tsai. A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. *IEEE Journal of Robotics and Automation*, 3(4):323–344, 1987.

[26] M.I. Vousdoukas, P. Perakakis, S. Idrissi, and J. Vila. Svmt: A MATLAB toolbox for stereo-vision motion tracking of motor reactivity. *Computer Methods and Programs in Biomedicine*, 108(1):318 – 329, 2012.

[27] C. C. Wang, C. Y. Chiang, P. Y. Lin, Y. C. Chou, I. T. Kuo, C. N. Huang, and C. T. Chan. Development of a fall detecting system for the elderly residents. In *The 2nd International Conference on Bioinformatics and Biomedical Engineering*, pages 1359–1362, 2008.

[28] Ge Wu. Distinguishing fall activities from normal activities by velocity characteristics. *Journal of Biomechanics*, 33(11):1497–1500, 2000.

[29] J. Yamato, J. Ohya, and K. Ishii. Recognizing human action in time-sequential images using hidden Markov model. In *Proceeding of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Champaign, IL, June 1992.

[30] Y. Zigel, D. Litvak, and I. Gannot. A method for automatic fall detection of elderly people using floor vibrations and soundproof of concept on human mimicking doll falls. *IEEE Transactions on Biomedical Engineering*, 56(12):2858–2867, 2009.