


RGB-D camera based walking pattern recognition by support vector machines for a smart rollator

He Zhang¹ · Cang Ye¹ 

Received: 27 August 2016 / Accepted: 10 November 2016 / Published online: 4 January 2017
© Springer Singapore 2016

Abstract This paper presents a walking pattern detection method for a smart rollator. The method detects the rollator user's lower extremities from the depth data of an RGB-D camera. It then segments the 3D point data of the lower extremities into the leg and foot data points, from which a skeletal system with 6 skeletal points and 4 rods is extracted and used to represent a walking gait. A gait feature, comprising the parameters of the gait shape and gait motion, is then constructed to describe a walking state. K-means clustering is employed to cluster all gait features obtained from a number of walking videos into 6 key gait features. Using these key gait features, a walking video sequence is modeled as a Markov chain. The stationary distribution of the Markov chain represents the walking pattern. Three Support Vector Machines (SVMs) are trained for walking pattern detection. Each SVM detects one of the three walking patterns. Experimental results demonstrate that the proposed method has a better performance in detecting walking patterns than seven existing methods.

Keywords Smart rollator · Walking pattern recognition · Gait features · Markovian chain · Support vector machines

1 Introduction

Walking therapy is a particular physical therapy (or physiotherapy) that assists a motor-impaired patient to recover their walking ability. This treatment requires interaction and cooperation between a therapist and a patient. The patient is offered instructions to perform the physiotherapy exercises in a monitored manner that provides feedbacks to the therapists for evaluating the effectiveness of the exercises and adjusting the therapy parameters. However, due to the lengthy recovery process and the need of travel, one-to-one in-clinic treatment is prohibitively expensive. As a result, the patient is taught in clinic about the therapy exercises and performs the exercises at home. While it is cost-effective and save the patient time in travel, at-home physiotherapy does not provide the therapist with feedback in a timely fashion for evaluation and adjustment of the exercises. Often, a patient uses a rolling walker (aka rollator) (Joly et al. 2013; Alwan et al. 2007; Tung 2010; Dune et al. 2012) as a walking aid and to support the therapy exercises during the recovery process. Our work is therefore to develop a computer vision method for automatic detection of walking patterns and devise a smart rollator system that is able to provide persistent monitor on the user's walking patterns for at-home walking therapy. The system can be used to score a physiotherapy exercise by monitoring the change in the user's walking patterns during the course of recovery.

We define a walking pattern as a sequence of walking postures and speeds. In the course of a walking therapy, a patient undergoes changes in both walking posture and speed. If the prescribed exercises are effective, the patient's walking gait will change from abnormal to normal and the speed from slow to normal. Otherwise, there will be no noticeable change in the walking pattern. In other words, detection of walking pattern change plays a critical role to the therapist in judging the effectiveness of the at-home

✉ Cang Ye
cye@ualr.edu

He Zhang
hxzhang1@ualr.edu

¹ Department of Systems Engineering, University of Arkansas at Little Rock, 2801 S. University Ave, Little Rock, AR 72204, USA

walking therapy sessions. Walking pattern recognition by computer vision involves lower limb detection, gait feature (including gait shape and gait motion parameters) extraction, walking pattern representation (as a sequence of gait features), machine learning for pattern detection.

In the literature, force and moment sensors (Alwan et al. 2007) have been used on a smart rollator to estimate the step count, pace and stride time of the rollator user. However, these sensors cannot measure the walking posture. In (Tung 2010), a video camera is mounted on the front bar of a rollator to monitor the lower limb behavior of the user for balance control. The system measures the displacements and velocities of the feet and it requires the user to wear markers on the shoes for foot detection. Recently, RGB-D camera has been employed to measure a person's walking postures and speeds (Gritti et al. 2014; Joly et al. 2013). An RGB-D camera provides reliable depth data for lower limb detection. Gritti et al. (2014) propose a histogram based lower limb detection algorithm that extracts a person's feet and legs from an RGB-D camera's depth data and tracks the feet and legs over time. However, it does not measure the walking postures. Joly, et al. (2013) propose a model-fitting method to detect bare legs and bare feet from the depth data of a Kinect sensor. The method models a bare leg as a cylinder and a bare foot as a plane and fit the parametric models to the Kinect data to detect the leg and foot. Although a skeleton representation of the lower limb can be created from the detected leg and foot, the work in (Joly et al. 2013) mainly focuses on determining foot orientation and ankle angle. However, the cylinder model fitting approach cannot be used in our case where the human legs are covered by the deformable pant. It is not possible to use any parametric model to describe the motion-induced deformation of the pant. In this paper, we propose a new method to determine the leg skeleton by least square plane fitting. Based on the skeletons of the lower limbs, we introduce a new gait feature to describe the gait shape and the gait motion and use it for walking pattern recognition. The gait feature representation resembles the action feature, consisting of shape and motion parameters of a full skeletal system of human body, that has been used in (Xiaodong and YingLi 2012) for human action recognition.

Existing methods (Xiaodong and YingLi 2012; Chaaraoui et al. 2014) for human action recognition may be applied to walking pattern recognition. In (Xiaodong and YingLi 2012), an image-to-image difference of the action features between a test video and a class—a video representing a particular class of action—is computed and the sum of the differences for all image frames is used for action detection. The sum does not take into account the transitions between action states. The method in (Chaaraoui et al. 2014) allows comparison of one image frame against multiple image frames. However, transitions between action states are not considered. In (Zhang and Ye 2015), we propose method to

recognize walking pattern for a smart rollator by analyzing the point cloud data stream of an RGB-D camera. This paper is an extended version of (Zhang and Ye 2015). The proposed walking pattern detection method uses a Markov chain model to capture the characteristics of a gait feature sequence. A gait feature consists of the shape and motion parameters of the walking gait. The transition matrix of the Markov chain model records both the state and state transition information of a walking sequence. If a walking sequence has a fixed pattern, the transition matrix should converge and thus the stationary distribution represents the walking pattern. To automatically identify the walking pattern, we used Support Vector Machine (SVM) (Laptev et al. 2007) because it has been proved efficient in recognizing human actions with discriminative feature descriptors.

This paper is organized as follows. Section 2 briefly describes our RGB-D camera based smart rollator system. Section 3 introduces the data processing pipeline of the walking pattern recognition method. Section 4 presents the method for leg and foot extraction and the construction of gait feature. Section 5 first briefly describes three popular human activity detection methods and then introduces the Markov chain modeling of a gait feature sequence and the SVMs for walking pattern recognition. Section 6 presents the experimental results of the proposed method and the comparisons with seven existing methods. The paper is concluded in Sect. 7.

2 Smart rollator setup

As depicted in Fig. 1a, an RGB-D camera (ASUS Xtion PRO LIVE) is installed on a rollator, facing towards the lower-extremity of the user with a tilt-down angle $\theta = -20^\circ$. This view angle ensures that the feet and the lower parts of the legs are inside the camera's field of view when the user is walking. The RGB-D camera provides a color video and a depth video with 640×480 pixels at 30 fps. Given a depth image, the 3D point cloud of the user's lower body can be obtained in real time. Figure 1b shows the point cloud of the user's legs and feet from a depth image frame. The camera coordinate system $X_c Y_c Z_c$ and the coordinate system $X_w Y_w Z_w$ that is used to analyze the lower extremity motion are depicted in Fig. 1a. $X_w Y_w Z_w$ is obtained by rotating $X_c Y_c Z_c$ around X_c for 20° . Each point q_i of the point cloud in $X_c Y_c Z_c$ is transformed into a point p_i in $X_w Y_w Z_w$ by

$$p_i = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -\cos(\theta) & -\sin(\theta) \\ 0 & \sin(\theta) & -\cos(\theta) \end{bmatrix} q_i \quad (1)$$

For each data frame, the floor plane (shown as the purple rectangle in Fig. 1b) is extracted from the point cloud p_i

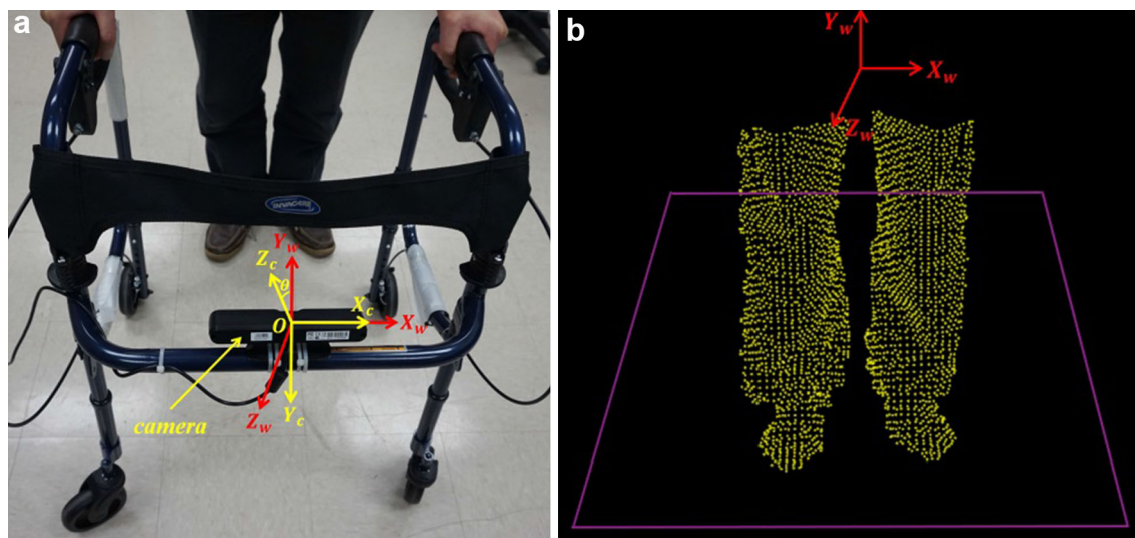


Fig. 1 Smart rollator and its coordinate systems: **a** the rollator with the Xtion camera; **b** floor plane and point cloud of the lower-extremity

using a RANSAC plane segmentation method (Qian and Ye 2014) and the points belonging to the plane are removed. The rest of the points are then used for foot and leg extraction as described in Sect. 4.2.

In this paper, a simulated walking therapy case is used for the development and validation of the proposed method. The rollator user imitates the walking patterns of a patient with knee injury during the course of recovery. Both normal walk and abnormal walk will be performed and video data (stream of RGB and depth data) will be captured by the RGB-D camera for training and testing the walking pattern recognition method. A Normal Walk (NW) is one with a sequence of normal walking gait at a regular speed. An abnormal walk is one with a sequence of normal/abnormal walking gaits at a much slower speed, sometimes near zero (i.e., halt). The abnormal walking gait is a lame walking gait. Abnormal walks include Slow Walk with Halt (SWH), Slow Lame Walk (SLW) in this paper.

3 Data processing pipeline of the smart rollator system

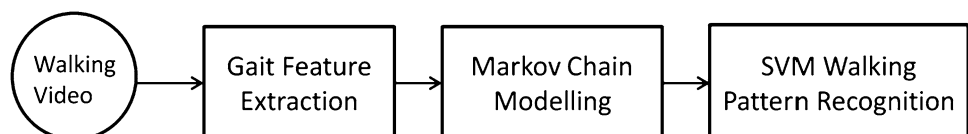
As depicted in Fig. 2, the data processing pipeline of the smart rollator system consists of three main modules: Gait Feature Extraction (GFE), Markov Chain Modeling (MCM), and SVM-based Walking Pattern Recognition (SWPR). The GFE extracts the point cloud of the user's

lower-extremity from a frame of a walking video. After locating the data for the foot and leg, the GFE models the lower-extremity as a skeletal system consisting of skeletons and skeletal points. It then computes the position and motion parameters of the left and right skeletal systems' skeletons and skeletal points. Using these parameters, the GFE constructs a gait feature for the frame. The MCM first clusters the gait features extracted from a number of walking videos into six classes, each of which is a key gait feature. It then describes each video frame by one of the six key gait features. This turns the walking video into a Markov chain whose stationary distribution represents a certain walking pattern. The SWPR maps the stationary distribution to a walking pattern by a trained SVM. The technical details of the three modules will be given in the following sections.

4 Gait feature extraction

In this paper, a gait feature contains parameters describing gait shape and gait motion. The gait shape parameters encode the current information of the user's lower extremity posture while the gait motion parameters describe how one gait shape evolves into another. The gait shape parameters are the positions of skeletal points of each lower extremity and the gait motion parameters are the velocities of these skeletal points. Collectively, these

Fig. 2 Data processing pipeline



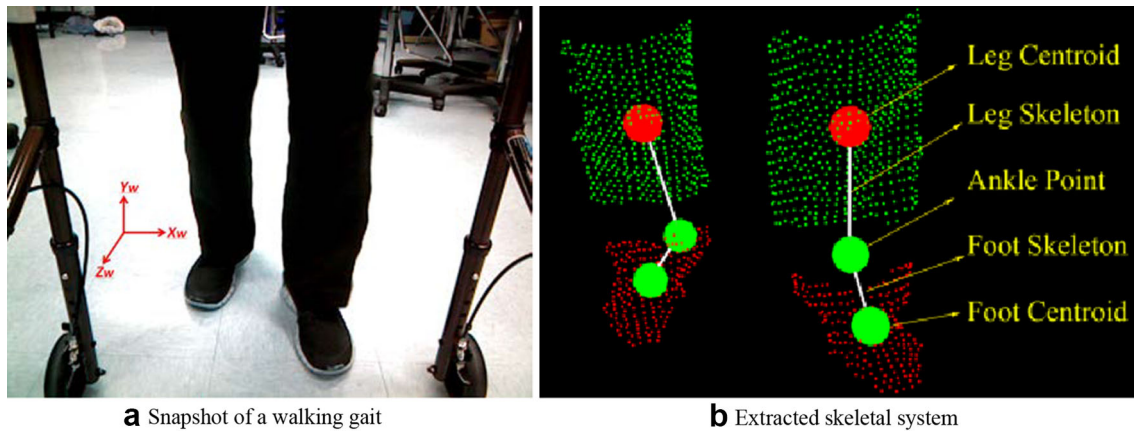


Fig. 3 Skeletal system extracted from the point cloud data of a normal walking gait

parameters describe a walking state of the rollator user. The process of gait feature extraction is divided into four steps: lower limb detection, leg and foot segmentation, leg and foot skeletons extraction, and gait feature construction.

4.1 Lower limb detection

Considering the case of walking on a flat ground with the rollator, we use the RANSAC plane segmentation method (Qian and Ye 2014) to extract the floor plane from the first frame of the camera's point cloud data. The extracted floor plane is then used to initialize the rollator's coordinate systems as mentioned in Sect. 2. Data points within the view volume clipped by the chassis of the rollator and the depth limit of the ASUS Xtion PRO LIVE (0.8–3.5 m) are identified, out of which we select the clusters within the first 70 cm above the floor plane as lower limb cluster \mathbf{P} containing feet and legs.

4.2 Leg and foot segmentation

A 2-stage processing is employed to find the foot and leg segments from each lower limb cluster \mathbf{P} . In the first stage, the minimum y coordinate y_{\min} of the lower limb cluster's data points is obtained and the coarse foot and leg segments, \mathbf{P}'_f and \mathbf{P}'_l , are located based on the data points' y -coordinates. Assuming the y -span between the toe and the ankle of a human's feet is smaller than 0.2 m, we locate points within $[y_{\min}, y_{\min} + 0.2 \text{ m}]$ as the foot segment \mathbf{P}'_f and the rest the leg segment \mathbf{P}'_l . In the second stage, the normal vector of each point in \mathbf{P}'_f is first computed. Then, the normal vector based region growing segmentation algorithm (http://pointclouds.org/documentation/tutorials/region_growing_segmentation.php) is implemented to extract the accurate foot segment \mathbf{P}_f from \mathbf{P}'_f . The leg segment is determined by $\mathbf{P}_l = (\mathbf{P} - \mathbf{P}_f) \cap \mathbf{P}'_l$.

Figure 3b depicts the segmentation result on a frame of depth data of the camera. The points of the leg segment are shown in green while the points of the foot segment are shown in red.

4.3 Leg and foot skeleton extraction

Three skeletal points are computed and used to form the leg and foot skeletons. The first two skeletal points are the centroids of the leg and foot segments. And the third point—the ankle point—is determined as the point where the leg intersects the foot-plane. Similar to (Joly, C., Dune, C.: Feet and Legs Tracking Using a Smart Rollator Equipped with a Kinect. IEEE/RSJ International Conference on Intelligent Robots and Systems. Tokyo, Japan 2013), a Least-Square Plane (LSP) to the data points of the foot segment \mathbf{P}_f is first computed and its normal vector \bar{n} is used to describe the foot orientation. The LSP is called a foot-plane.

The skeleton of the leg can be extracted from the data points p_j , for $j = 1, \dots, N$, of the leg segment \mathbf{P}_l , where N is the total number of data points of \mathbf{P}_l . In an ideal case where the pant leg is pleat-free and the data points are noise-free, the orientation of the leg skeleton, denoted μ , is orthogonal with the surface normal of p_j , denoted $\mathbf{n}_j = (n_{jx}, n_{jy}, n_{jz})$. If we treat \mathbf{n}_j as a data point, μ is the normal of the LSP to point set \mathbf{n}_j for $j = 1, \dots, N$. The least-square problem is equivalent to find the normal of the LSP to a point set $q_j = (k_j n_{jx}, k_j n_{jy}, k_j n_{jz})$ for $j = 1, \dots, N$, where k_j is a randomly generated non-zero value for \mathbf{n}_j . By applying k_j , we spread the data points in a larger area without changing each point's vector direction. This treatment avoids the case where all data points locate in a narrow area, making the LSP sensitive to noise. For a real-world scenario, using all data points to compute μ minimizes the effects of the noise and pant-pleats. The LSP problem is solved by the singular value decomposition method. The centroid of \mathbf{P}_l and μ are then used to describe the leg skeleton.

The ankle point is determined as the intersection of the leg skeleton and the foot-plane. A lower limb skeletal system, consisting of 2 skeletons and 3 skeletal points is then formed as shown in Fig. 3b.

4.4 Gait feature representation

The extraction of the two skeletal systems results in 6 skeletal points $sp_i (i = 1, \dots, 6)$. The skeletal points' positions determine the gait shape. Using the 6 skeletal points' centroid $sp_c = \{x_c, y_c, z_c\}$ as the reference point, the skeletal points' coordinates are re-computed by $sp'_i = sp_i - sp_c$, from which a bounding box $[x_{\min}, x_{\max}, y_{\min}, y_{\max}, z_{\min}, z_{\max}]$ is created. Finally, a 18-dimensional vector f^s representing the gait shape is computed from $sp'_i = \{x'_i, y'_i, z'_i\}$ by:

$$f^s = \begin{cases} (x'_i - x_{\min}) / (x_{\max} - x_{\min}) & \text{for } j = 1 \dots 6, i = 1 \dots 6 \\ (y'_i - y_{\min}) / (y_{\max} - y_{\min}) & \text{for } j = 7 \dots 12, i = 1 \dots 6 \\ (z'_i - z_{\min}) / (z_{\max} - z_{\min}) & \text{for } j = 13 \dots 18, i = 1 \dots 6 \end{cases} \quad (2)$$

The velocity of each skeletal point is computed from its positional change between two consecutive frames. We denote the velocities of a leg point, ankle point and foot point by v_l, v_a and v_f , respectively and use superscripts 1 and 2 to represent left and right, respectively. Because the ankle point is indirectly computed from the point cloud (as the intersection between the leg skeleton and foot-plane), its position may incur a larger error than the other two skeletal points. This means that its velocity computed from two consecutive frames is not reliable. Figure 4 shows the motion parameters computed from the image frames of a 12-s walking video clip. Taking the velocity of the right ankle point v_a^2 (Fig. 4a) for instance, we can observe that v_{ax}^2 goes beyond ± 1 m/s at some frames. This should not occur because $|v_{lx}^2| < 0.26$ m/s (Fig. 4c) and $|v_{fx}^2| < 0.2$ m/s (Fig. 4b). The measurement error in $|v_{ax}^2|$ was caused by the error in extracting the ankle point. However, we found that the measurement of angle between the foot skeleton and the leg skeleton is more accurate, indicating the angular velocity ω_p (Fig. 4d) may be used as a more reliable motion parameter. Therefore, we use the velocities of the foot and the leg centroids, denoted by $v_f = (v_{fx}, v_{fy}, v_{fz})$ and $v_l = (v_{lx}, v_{ly}, v_{lz})$, and the angular velocity ω_p to form a 14-dimensional vector f^m :

$$f^m = [v_f^1, v_l^1, \omega_p^1, v_f^2, v_l^2, \omega_p^2] \quad (3)$$

Figure 4b–d depict the gait's motion parameters that are used to form vector f^m . They were computed from a sequence of image frames of a 12-second walking video clip.

By concatenating (2) and (3), a 32-dimensional feature is constructed as $[f^s, f^m]$. In order to rule out correlation between the elements of the feature vector, the principal component analysis method (Pearson 1901) is employed to reduce the feature's dimensionality from 32 to 18. The 18 eigenvalues weight over 95%.

5 Walking pattern detection

A classification algorithm is needed to detect the user's walking pattern from the extracted gait feature. In (Sagha, H., Digumarti, S.T., Millán, J.D.R., Chavarriaga, R.: Benchmarking classification techniques using the Opportunity human activity dataset. In IEEE International Conference on Systems, Man, and Cybernetics (SMC) 2011), a comparative study of four well-known classification techniques, namely Nearest Centroid Classifier (NCC), linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), and K-nearest Neighbors (KNN), are conducted by using a benchmark dataset—UCI OPPORTUNITY dataset (Ricardo et al. 2013) for human action recognition. Walking pattern detection methods based on these methods and three state-of-the-art methods, including Naive-Bayes-Nearest-Neighbor (NBNN) Classifier (Xiaodong and YingLi 2012), key pose based Dynamic Time Warping (DTW) (Chaaroui et al. 2014), and Bag-of-Video-Words (BoVW), will be implemented and compared with the proposed method in this paper. In this section, we first give a brief introduction on NBNN, DTW, and BoVW and then describe in details a new Markov chain based classification method for walking pattern detection. In Sect. 7, the proposed method's performance will be compared with the other above-mentioned techniques.

5.1 Naive-Bayes-Nearest-neighbor (NBNN)

The idea of using NBNN for human action detection (Xiaodong and YingLi 2012) is to use a number of feature collections, $C = \{C_j\}$, to describe different types of human actions. Each collection of features, C_j , represents an action of j th type. These features are extracted from video clips that have been labeled to be the j th type actions. Given a M -frame test video, features, p_i for $i = 1, \dots, M$, are first extracted and the classification result, denoted C^* , of NBNN is given by:

$$C^* = \arg \min_{C_j} \sum_i^M \|p_i - \text{NN}_{C_j}(p_i)\|^2, \quad (4)$$

where $\text{NN}_{C_j}(p_i)$ is the nearest neighbor of p_i in C_j .

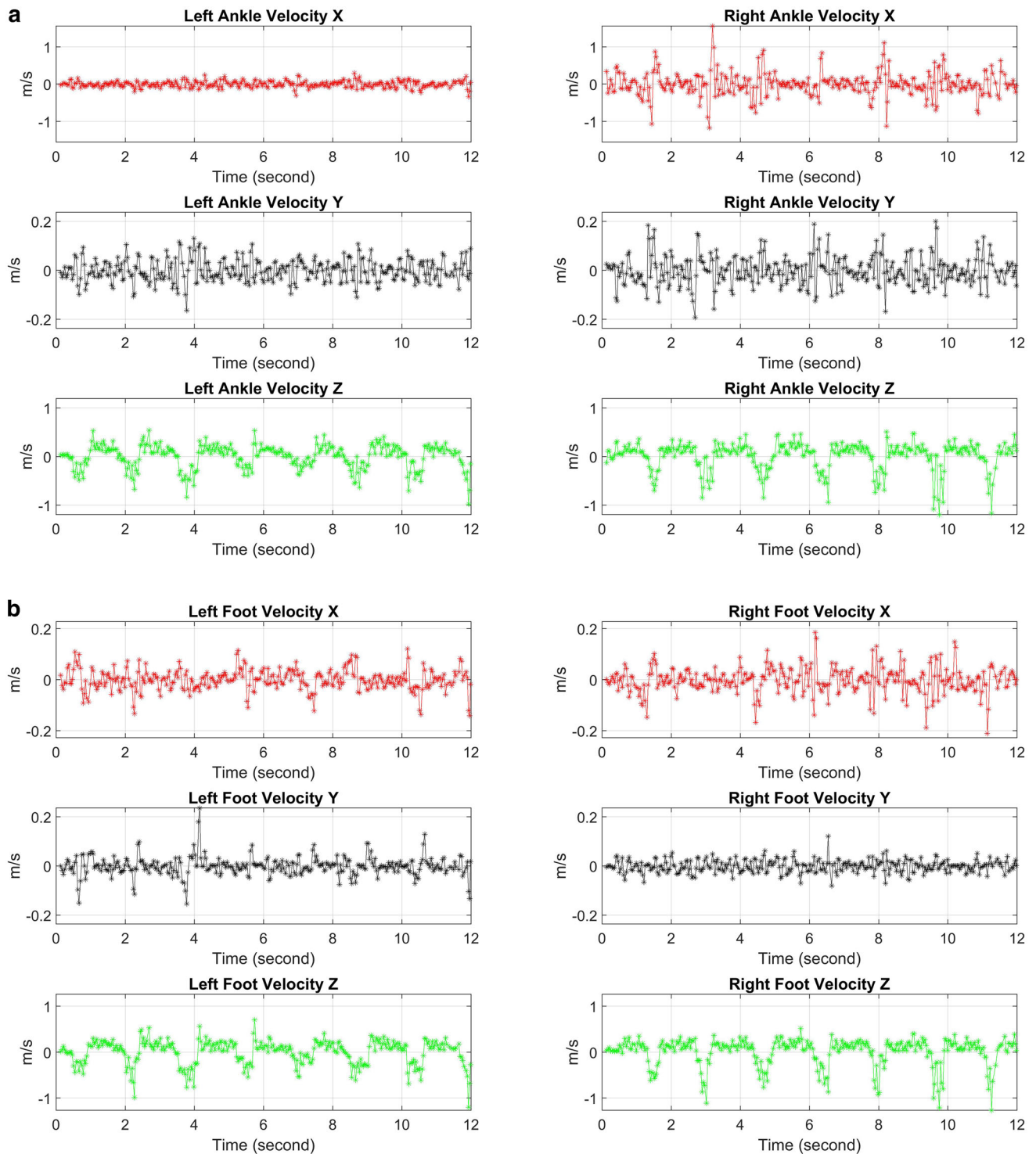


Fig. 4 Gait motion parameters computed from a 12-s walking video clip

5.2 Dynamic time warping (DTW)

In (Charaoui et al. 2014), DTW is employed for human action detection. The method models human action as a sequence of key features and identify action through sequence matching by using DTW. In training phase, the

training data (video clips) is first processed to extract gait features. Then key features are obtained from these gait features by using K-Means and each video is described by a sequence of key features. In action detection phase, a test video is processed and represent by a sequence of key features S . The distance between S and the k th key feature

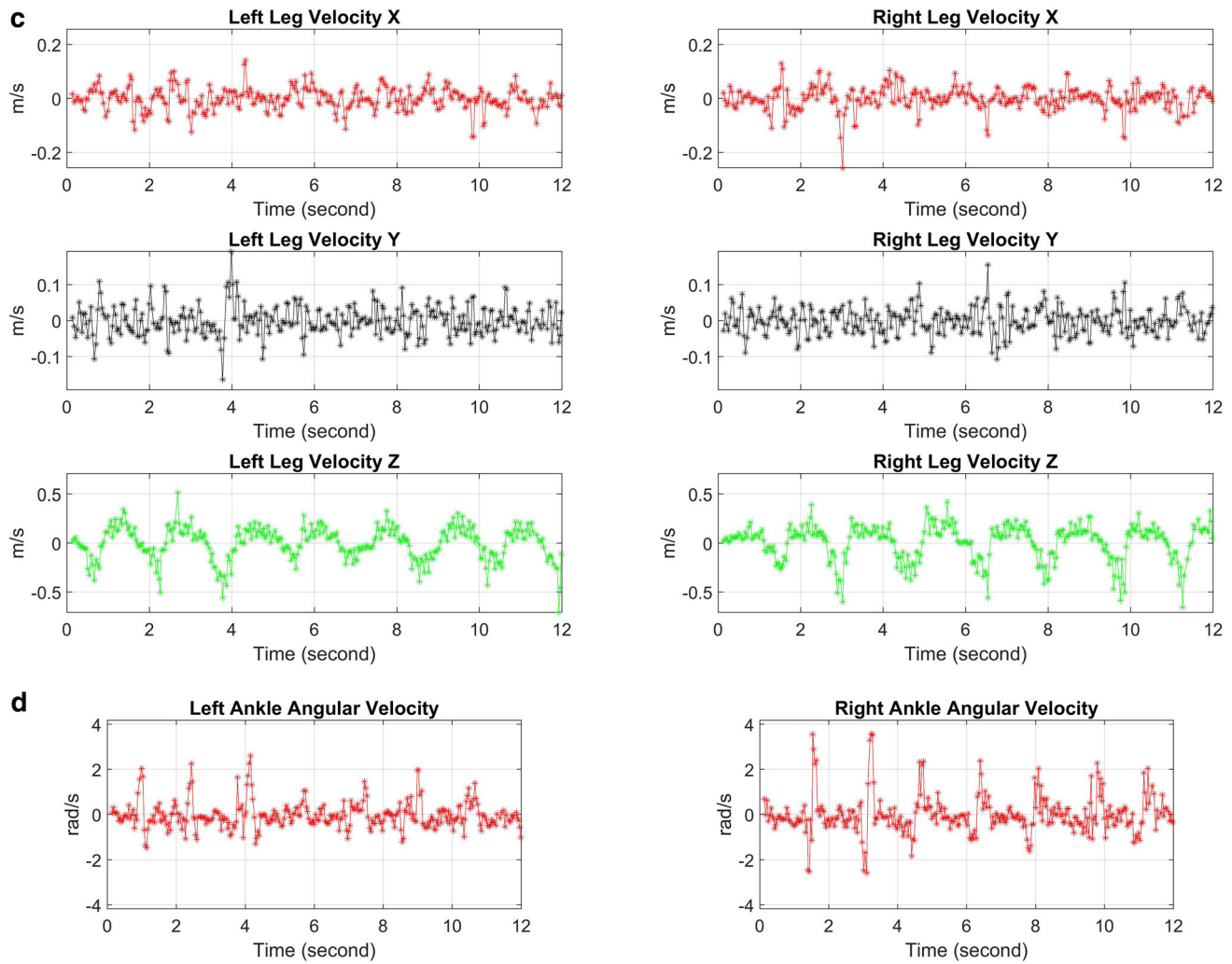


Fig. 4 continued

sequence S_k is denoted $D_k(S_k, S) = \Delta_k(M_k, N)$, where M_k and N are the sizes of S_k and S , respectively. $\Delta_k(M_k, N)$ is computed by using a dynamic programming rule as follows:

$$\Delta_k(i, j) = \min \left\{ \begin{array}{l} \Delta_k(i-1, j) \\ \Delta_k(i, j-1) \\ \Delta_k(i-1, j-1) \end{array} \right\} + d_{ij}, \quad (5)$$

where d_{ij} is the distance between key features p_i and p_j . S is classified as one belonging to the class that contain S^* with minimum $D(S^*, S)$.

5.3 Bag-of-video-words (BoVW) based SVM

In (Mona et al. 2015), BoVW technique is used for human action recognition. The method's training phase consists of five steps. First, SIFT features (Lowe 2004) are extracted

from all images of the training videos and the extended visual features, each of which includes the SIFT descriptor and the x and y coordinates of the key point, are formed. Second, the feature descriptors are clustered by using K-means algorithm and the resulted clusters' centers, call visual words, form the word vocabulary. Third, each video's feature descriptors are mapped to the vocabulary to create a word frequency histogram (the video's signature) to represent the video. Fourth, the value of each bin of the histogram is normalized over all the videos. Fifth, a multi-class SVM is trained by using the normalized histograms. In the testing phase, the training histograms are re-normalized along with the histogram of the test video. The re-normalization is performed in such a way that the resultant normalized test histogram is affected by all the histograms (training ones and test one). Afterward, the normalized test histogram is fed to the SVM for classification.

5.4 Markov chain modeling of walking

The gait feature extraction process can produce a large number of gait features from a walking video. Similar to the idea of key feature, they are classified into a few representative gait features, called Key Gait Features (KGFs), to simplify the representation. In this paper, we use K-means classification algorithm to partition the gait features (extracted from a number of walking videos) into a number of clusters. A KGF is then computed as the centroid of a cluster. After the KGFs are determined, each gait feature of a walking video is represented by one of the KGFs and the sequence of KGFs forms a Markov chain whose stationary distribution is used to detect the rollator user's walking pattern.

5.4.1 Key gait feature

We employ K-means algorithm to partition the gait features into k clusters and the centroid of each cluster is computed as a KGF (Laptev et al. 2007; Chaaraoui et al. 2014). The value of k is determined by the Bayesian Information Criterion (BIC) (Pelleg et al. 2000). The BIC is a criterion for selecting a model out of a finite set of models. In our case, we choose k with the lowest BIC. In addition, if the number of gait features belonging to a cluster is smaller than a threshold =50, this cluster is treated as an outlier and thus deleted. Using this scheme, we extract 6 KGFs from a number of walking videos and use them to represent all possible gaits for a walking video.

5.4.2 Markov chain model

Each gait extracted from a walking video is now represented by a KGF if the norm of the difference between the gait feature and the KGF is below a threshold. By treating a KGF as a state, we denote the gait sequence of a walking video by a Markov chain S . The transition matrix P of the Markov chain is of 6×6 dimensions. Each entry of the matrix p_{ij} represents the probability, with which state i evolves into state j . p_{ij} can be obtained from state sequence S by

$$p_{ij} = \begin{cases} n_{ij}/(n_i - 1) & \text{if } j \text{ is the last state in } S \\ n_{ij}/n_i & \text{otherwise} \end{cases}, \quad (6)$$

where n_{ij} is the number of transitions from state i into state j while n_i is the number of occurrences of state i in S . Therefore, P can be computed from S . It is noted that (6) guarantees $\sum_{j=1}^N p_{ij}$ for $i = 1, \dots, 6$, where N is the total number of states in S . The following is a Markov chain sample obtained from a portion of a walking video:

S : 1113232334444465652323344444665623233344

S describes how long a gait is held and what gait it transforms into. For this sequence, p_{34} can be computed by $p_{34} = n_{34}/n_3 = 3/11 = 0.273$ and $p_{46} = n_{46}/(n_4 - 1) = 2/(11 - 1) = 0.2$. The other entries can be computed in a similar way to obtain the transition matrix P .

Assuming that a walking video has a fixed pattern, the transition matrix P of the Markov chain should converge with a sufficiently large number of video data frames. In this case, the stationary distribution π of the Markov chain holds the inherent property of the walking pattern. π , a row vector whose entries are non-negative and sum to 1, is defined by:

$$\pi P = \pi \quad (7)$$

It can be seen that π is a left eigenvector of P with an eigenvalue of 1. Therefore, it can be computed from P . In this work, we use π to represent the walking pattern of a walking video.

5.5 SVM for walking pattern recognition

As indicated earlier, there are three types of walking patterns to be detected. Therefore, a multi-class classifier is required for pattern recognition. In this paper, we use the one-vs-all strategy to train three Support Vector Machines (SVMs) to detect the walking patterns. One SVM will be trained to recognize a particular type of walking patterns by using the relevant training data $(\pi_i, y_i); i = 1 \dots, N$, where N is the number of walking videos used for training the SVM while y_i is the SVM output for π_i . y_i is manually labeled. Taking the training of the 3rd SVM (for SLW detection) as an example, feature vector π_i is computed for the i th video. If the walking pattern of this video is SLW, $y_i = +1$; Otherwise $y_i = -1$. The kernel function of the SVM is Gaussian kernel whose sigma is 0.03 and regularization parameter is 0.01.

6 Experimental results

6.1 Data collection

Nine human subjects participated in data collection. They were instructed to perform the three types of walks. For each walk, the image and depth data streams were recorded from the Xtion. The video for each walk is 12–17 s long, containing 360–500 data frames. 5 video clips were recorded for the experiments performed by each human subject, resulting in 45 video clips, 9 for NW, 9 for SWH and 27 for SLW. Three SLW videos were recorded for each subject limping on his left leg, right leg, and both legs, respectively.

6.2 Performance evaluation and comparison

In our experiments, leave-one-out cross validation technique is employed for performance evaluation. The performance of the proposed Markov Stationary Distribution (MSD) based one-vs-all SVM method is compared with that of the NCC, LDA, QDA, KNN and NBNN classifiers (Xiaodong and YingLi 2012) as well as the DTW method and BoVW based one-vs-all SVM (Laptev et al. 2007) method. In spite of their simplicity, NCC, LDA, QDA and KNN have been reported in (Sagha et al. 2011) to perform well on the UCI OPPORTUNITY dataset (Ricardo et al. 2013) for human action recognition. Therefore, they are implemented and compared with the proposed method in this paper. The average detection accuracy and the F-measure (Ricardo et al. 2013) of each method are used for performance evaluation. The F-measure takes into account the precision and recall for each class and can provide a better performance evaluation in terms of accuracy. The precision for the i th class is defined as $\alpha_i = \frac{TP_i}{TP_i + FP_i}$ and recall as $\beta_i = \frac{TP_i}{TP_i + FN_i}$, where TP_i , FP_i and FN_i are the true positive, false positive and false negative numbers for the class, respectively. Considering class imbalance, the F-measure is computed by using the classes' sample proportion,

$$F_1 = \sum_i 2 \times \frac{s_i \alpha_i \times \beta_i}{S \alpha_i + \beta_i}, \quad (8)$$

where s_i is the number of samples of class i and S is the total number of samples.

The experimental results are tabulated in Tables 1, 2, 3, 4, 5, 6, 7, 8. The three types of videos to be tested are indicated in bolded letters (with the number of video clips in the parenthesis). The classification result (NW, SWH and SLW) for each type of test videos is shown in the column. Taking the first column of Table 1 as an example, out of the 9 NW test videos, 4 was detected as NW, 1 as

Table 1 Detection result using NCC

	NW (9)	SWH (9)	SLW (27)
NW	4	3	7
SWH	1	0	5
SLW	4	6	15

Table 2 Detection result using KNN

	NW (9)	SWH (9)	SLW (27)
NW	3	5	0
SWH	2	1	4
SLW	4	3	23

Table 3 Detection result using LDA

	NW (9)	SWH (9)	SLW (27)
NW	4	3	1
SWH	3	2	8
SLW	2	4	18

Table 4 Detection result using QDA

	NW (9)	SWH (9)	SLW (27)
NW	2	4	2
SWH	4	0	2
SLW	3	5	23

Table 5 Detection result using NBNN

	NW (9)	SWH (9)	SLW (27)
NW	1	0	0
SWH	1	0	0
SLW	7	9	27

Table 6 Detection result using DTW

	NW (9)	SWH (9)	SLW (27)
NW	2	4	2
SWH	4	2	1
SLW	3	3	24

Table 7 Detection result using BoW

	NW (9)	SWH (9)	SLW (27)
NW	5	1	0
SWH	4	8	2
SLW	0	0	25

Table 8 Detection result using MSD

	NW (9)	SWH (9)	SLW (27)
NW	7	2	2
SWH	2	7	0
SLW	0	0	25

SWH, and 4 as SLW. The average accuracy and the F-measure of each method (over all video clips) are computed and tabulated in Table 9. From the Table 9, it is clear that the proposed method outperforms the other methods in both average accuracy (0.87) and F-measure (0.87). In term

Table 9 Result of the walking pattern recognition

Classifier performance	NCC	KNN	LDA	QDA	NBNN	DTW	BoVW	MSD
Accuracy	0.42	0.60	0.53	0.56	0.62	0.62	0.84	0.87
F-measure	0.41	0.58	0.55	0.52	0.50	0.60	0.85	0.87

of the simple performance index—average accuracy, the performances of the other 7 methods are ranked as BoVW based one-vs-all SVM, DTW, NBNN, KNN, QDA, LDA, and NCC. However, if the more accurate performance index—F-measure—is used, they would be ranked as BoVW based one-vs-all SVM, DTW, KNN, LDA, QDA, NBNN, and NCC.

7 Conclusion

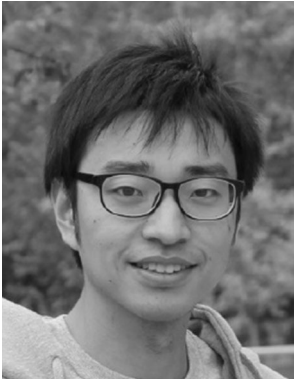
This paper presents an RGB-D camera based walking pattern detection method for a smart rollator system. The method extracts the user's lower limbs from the camera's depth data to obtain the gait information represented by a skeletal system with six skeletal points and four skeletons. By combining the parameters of the gait shape and gait motion, a gait feature is constructed to describe a walking state. K-means is employed to cluster all gait features extracted from a number of walking videos into six key gait features. Using the key gait features, a walking video sequence is modeled as a Markov chain, of which the stationary distribution represents the walking pattern. Three SVMs are trained and used to detect the three walking patterns. Experimental results validate that the proposed method outperforms seven existing methods in detecting walking patterns.

In term of future research, we will use video data collected from real patients' to test the method and compare its performance with that of the other methods. Also, we will define more walking patterns and include them in the proposed method. For real world application, the real-time video stream from the RGB-D camera will be examined by the proposed method segment by segment, each of which contains a fix number of data frames. The user's walking ability will be evaluated based on the accumulative recognition results on the video segments.

Acknowledgements This work was supported by the National Institute of Child Health and Human Development, the National Institute of Nursing Research, and the National Institute of Biomedical Imaging and Bioengineering of the National Institutes of Health under Award R01NR016151. The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies.

References

- http://pointclouds.org/documentation/tutorials/region_growing_segmentation.php
- Alwan, M., Ledoux, A., Wasson, G., et al.: Basic walker-assisted gait characteristics derived from forces and moments exerted on the Walker's Handles: results on normal subjects. *Med. Eng. Phys.* **29**, 380–389 (2007)
- Charaoui, A.A., Padilla-López, J.R., Climent-Pérez, P., et al.: Evolutionary joint selection to improve human action recognition with RGB-D devices. *Expert Syst. Appl.* **41**, 786–794 (2014)
- Dune, C., Gorce, P., Merlet, J.P.: Can smart rollators be used for gait monitoring and fall prevention? *IEEE/RSJ International Conference on Intelligent Robots and Systems* (2012)
- Gritti, A., Tarabini, O., Guzzi, J.: Kinect-based People Detection and Tracking from Small-footprint Ground Robots. *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Chicago, IL ((2014))
- Joly, C., Dune, C.: Feet and Legs Tracking Using a Smart Rollator Equipped with a Kinect. *IEEE/RSJ International Conference on Intelligent Robots and Systems*. Tokyo, Japan (2013)
- Laptev, I., Caputo, B., Schödl, C., et al.: Local velocity-adapted motion events for spatio-temporal recognition. *Comput. Vis. Image Underst.* **108**, 207–229 (2007)
- Lowe, D.: Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision* **60**, 91–110 (2004)
- Mona, M.M., Elsayed, H., Magda, B.F., et al.: An enhanced method for human action recognition. *J. Adv. Res.* **6**, 163–169 (2015)
- Pearson, K.: On lines and planes of closest fit to systems of Points in space. *Phil. Mag.* **2**, 559–572 (1901)
- Pelleg, D., Moore, A.W.: X-means: Extending K-means with Efficient Estimation of the Number of Clusters. *International Conference on Machine Learning (ICML)* (2000)
- Qian, X., Ye, C.: NCC-RANSAC: a fast plane extraction method for 3D range data segmentation. *IEEE Trans. Cybern.* **44**, 2771–2783 (2014)
- Ricardo, C., Hesam, S., Alberto, C., et al.: The opportunity challenge: a benchmark database for on-body sensor-based activity recognition. *Pattern Recogn. Lett.* **34**, 2033–2042 (2013)
- Sagha, H., Digumarti, S.T., Millán, J.D.R., Chavarriaga, R.: Benchmarking classification techniques using the Opportunity human activity dataset. In *IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (2011)
- Tung, J.: Development and Evaluation of the iWalker: An Instrumented Rolling Walker to Assess Balance and Mobility in Everyday Activities. Ph.D. dissertation, University of Toronto (2010)
- Xiaodong, Y., YingLi, T.: Eigenjoints-based action recognition using naive-bayes-nearest-neighbor. *Computer Vision and Pattern Recognition Workshops*, Providence (2012)
- Zhang, H., Ye, C.: An RGB-D camera based walking pattern detection method for smart rollators. *Lect. Notes Comput. Sci.* **9474**, 624–633 (2015)



He Zhang received BS degrees in Computer Science & Technology from China University of Mining & Technology, Beijing, China, in 2009. Since 2014 August, he is a Ph.D. student with the Department of Systems Engineering, University of Arkansas at Little Rock. His research interests include simultaneous localization and mapping, rehabilitation robotics, 2D/3D computer vision.



Cang Ye received the B.E. and M.E. degrees from the University of Science and Technology of China, Hefei, Anhui, in 1988 and 1991, respectively, and the Ph.D. degree from the University of Hong Kong, Hong Kong, in 1999. He is currently a professor with the Department of Systems Engineering, University of Arkansas at Little Rock. His research interests are in autonomous navigation of mobile robots, computer vision, assistive/rehabilitation robotics

and human–robot interaction.