EAI Endorsed Transactions
on Context-aware Systems and Applications
Research Article EAI.EU

Hidden Markov Model for recognition of skeletal data-based hand movement gestures

Bui Cong Giao[1,*], Trinh Hoai An[1], Nguyen Thi Hong Anh[1] and Ho Nhut Minh[2]

[1] Faculty of Electronics and Telecommunications, Saigon University, Vietnam
[2] Faculty of Electronics Engineering II, Posts and Telecommunications Institute of Technology, Ho Chi Minh Campus, Vietnam

Abstract

The development of computing technology provides more and more methods for human-computer interaction applications. The gesture or motion of a human hand is considered as one of the most basic communications for interacting between people and computers. Recently, the release of 3D cameras such as Microsoft Kinect and Leap Motion has provided many advantage tools to explore computer vision and virtual reality based on RGB-Depth images. The paper focuses on improving approach for detecting, training, and recognizing the state sequences of hand motions automatically. The hand movements of three persons are recorded as the input of a recognition system. These hand movements correspond to five actions: sweeping right to left, sweeping top to bottom, circle motion, square motion, and triangle motion. The skeletal data of hand joint are collected to build an observation database. Desired features of each hand action are extracted from skeleton video frames by using the Principle Component Analysis (PCA) algorithm for training and recognition. A hidden Markov model (HMM) is applied to train the feature data and recognize various states of hand movements. The experimental results showed that the proposed method achieved the average accuracy nearly 95.66% and 91.00% for offline and online recognition, respectively.

Keywords: Skeletal data, Hand movement recognition, PCA algorithm, HMM.

Received on 30 April 2018, accepted on 25 May 2018, published on 18 June 2018

doi: 10.4108/eai.18-6-2018.154819

1. Introduction

The development of interactive technology between human and computer-based machine is creating many advanced applications. For years ago, a person communicated with computers primarily via the basic access systems involving in keyboard, mouse, window, icon, menu, and point-and-click device. But these approaches did not establish a natural interface between human and device in the space. The development of virtual reality technology has changed the human-machine interaction. In the game, developers propose new ways which release players from the constraint of using keyboard and mouse to control characters. The sensors such as Microsoft Kinect and Leap Motion are considered as state-of-the-art devices toward the implementation of fully natural interfaces in which gestures, body poses, and hand activities become the controllers. Kinect provides the great opportunities not only in the development of games applications but also in academic fields such as automatic surveillance [1], human healthcare applications [2], and human-computer interaction applications [3].

In the field of robotics, especially in self-propelled robot applications, the Kinect camera plays an important role as a visual sensor for capturing input signals quickly and accurately. The Kinect systems are developed to detect objects using RGB and depth features. An algorithm for automatic exploration was used in a mobile robot accompanied with a Kinect camera to generate a map automatically [4] or to follow a person based on RGB-D features [5]. In these designs, depth sensor and RGB camera from Kinect are used in the role of data acquisition tools to generate the map in the unknown environment or to follow the moving target.

* Corresponding author. Email: bcgiao@sgu.edu.vn

For healthcare applications, the fall recognition system for the elderly [6] was used in various projects to detect and recognize the postures, gestures, activities, movements, body parts, gait, and frailty of elderly people. Furthermore, based on the advantages of depth data and skeletal data, a human fall recognition system [7] was developed to detect the fall states and warn to supporters. These data are important to analyse and evaluate abnormalities in human health. Results of these researches showed that it could help the doctor easily offer the diagnosis and warning about health hazards to patients.

With human-computer interaction systems, Kinect is a useful tool to support communication between people and computers via voice, human postures, hand gestures, and object movements. Depth and color images are recorded to trigger the control program which can execute and begin a program or software on the computer. A non-touch input system [8] was proposed for tapping keys by using a virtual keypad which is made up of the depth data from Kinect sensor in a virtual reality space. In addition, the hand gesture recognition system was also applied to detect the speed of hand motion [9], to recognize of Arabic numbers (0-9) [10], and to simulate a mouse application which allowed a person could perform the functions of a real mouse on the computer [11].

In order to build an object recognition model successfully, it is very necessary to apply the suitable algorithms in the training and recognition to ensure the accuracy of the system. The HMM algorithm [13] is applied to build a human activities system from skeletal data. The result showed that achieved the high efficient recognition by 90% - 100%.

This research aims to establish a model to recognize the hand movements using a Kinect camera. The obtained dataset are five separate states sequences corresponding with five motions performed by the hand. Raw data are pre-processed using specialized algorithms. The pre-processing step consists of scaling and normalization. For the recognition of gestures, a Principle Component Analysis (PCA) algorithm is employed to extract the characteristic data of each motion. After that, the HMM algorithm will be applied for training and classifying each state of hand movement.

The main contributions of the paper are as follows:

(1) The use of skeletal data created by combination of RGB and depth images for the training and recognition processes supports the recognition system to offer the high accuracy. An advantage of skeletal data is that it can remove the influences of noise and unexpected factors from the recording environment.

(2) The raw data, which are recorded from Kinect sensors, would be normalized in order to facilitate for extracting features of database and to enhance the accuracy in the recognition phase.

The paper is organized as follows. Section 2 shows some related works. Section 3 describes data acquisition and materials using the Kinect system. Section 4 then presents the proposed method. After that, Section 5 reports the experimental results. Finally, Section 6 gives conclusions.

2. Related Works

There have been three typical research works dealing with recognition of human activities by using depth cameras. They are briefly described as follows.

- Hai and Kha [14] proposed a human action recognition system (HARS) using a Kinect to collect video frames of human activities. The operation of the HARS consists of the following phases. At first, the HARS uses the star skeleton algorithm [15] to extract the desired features of each human skeleton from the video frames. The HARS then maps the skeletal features to observation symbols, which constitutes an observation sequence, for the training process. Subsequently, for the learning process, the HARS uses the Baum-Welch algorithm [16] to transform in succession the observation sequences into seven HMMs corresponding to seven prespecified human actions. Finally, for the recognition process, the HARS can recognize the current human activity by labelling the observation sequence with the most suitable model in the seven trained HMMs. The HMMs are then used to classify human actions in both indoor and outdoor environments. The experimental results demonstrated that the HARS might recognize human actions with the high accuracy over 85% and the fast processing time.

- Dubois and Charpillet [16] proposed a regcognition system using a RGB-D camera to facilitate real-time analysis of the movemement of a person. Furthermore, based on HMMs, the recognition system can detect falls of a person. The operation of the regcognition system is described as follows. When the person is walking, the RGB-D camera records the depth images of the activity. Using the depth images, the regcognition system would analyse the trajectory of his centre of mass to measure gait parameters. After that, the paramters are input of the HMMs in order to result in a frailty evaluation for him. The experimental results showed that the regcognition system could offer daily information of the person's activities. Most important, the regcognition system can detect his falls to quickly alert emergency services. Thereby, the regcognition system is very useful to secure the elderly.

- Uddin et al. [12] proposed a method for human activity recognition using the joint angles from a 3D model of a human body. The method estimates directly from time-series activity images obtained with a single stereo camera by co-registering a 3D body model to the stereo information. Next, the estimated joint-angle features are mapped into codewords to generate discrete symbols for a HMM of each activity. Using these symbols, the method trains each activity through the HMM. All the trained HMMs are then used for activity recognition. The system achieved high accuracy from 92.8% to 98.1% with input data of 3D body-joint-angle features.

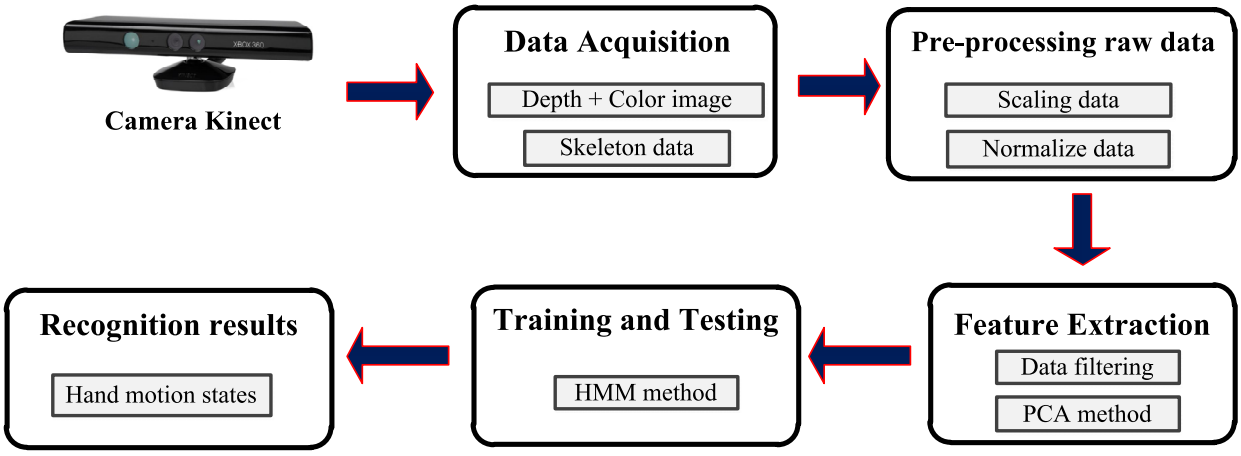


Figure 1. Proposed hand motion recognition system

3. Data Acquisition and Materials

In this research, a recognition system of hand movements has proposed with five main processes: data acquisition using the Kinect system, normalization of raw data, data features extraction, and training and recognizing hand movement gestures. Figure 1 shows the above-mentioned processes.

In particular, joints data of skeleton are produced using the Kinect camera which has an RGB camera and a laser sensor inside. The obtained parameters in each joint are the values of three axes *x, y, z*, in which *z* coordinate is the depth information. Therefore, the skeletal data with the 3-dimensional system is valuable information which could make easier to detect and recognize hand motion gestures. After acquired, the raw data are analyzed to scale and normalize all received parameters. The PCA algorithm is applied to extract features for training. For the recognition of hand movement gestures, an HMM method will be used in order to train samples and recognize separate gestures. The results will demonstrate the accuracy of the proposed method.

The major requirement of the system is to detect and recognize the motion states of a hand. Every state of the hand movements has a large number of datasets which is analyzed, calculated before trained by using optimal algorithms. Therefore, the process acquisition of raw data is performed carefully to achieve a high performance.

3.1. Data Acquisition

For sampling data, three people were invited to record patterns with different hand gestures. These subjects were conducted to exercise separate actions matching with hand movement states. Each person performed alternately five hand gestures in which 20 video samples were collected per gesture. This system collected five sequence states of hand motion for training and recognizing, including sweeping right to left, sweeping top to bottom, circle motion, square motion, and triangle motion. Figure 2 demonstrates clearly the process to obtain sample data.

In order to minimize the factors that affect quality and reliability of samples, the installation location of the Kinect system in space has to meet the specifications of the manufacturer. Figure 3 shows the suitable position to place the Kinect, corresponding to the 1.2 – 1.4m height and the 2 – 3m depth to secure the collected images and 3D skeletal data with high quality.

The program for obtaining data is mainly based on Kinect for Xbox 360, Motion Sensor along with Window SDK version 1.8, which is supported with C/C++ and MATLAB language. Furthermore, the program allows to

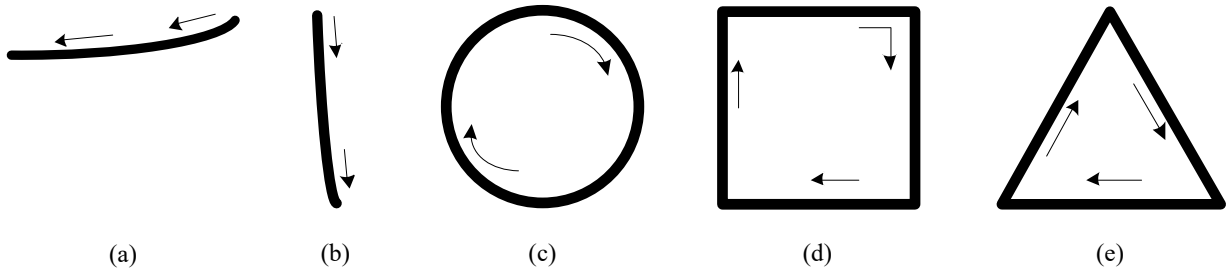


Figure 2. The direction of hand movement for each gesture: (a) sweeping right to left, (b) sweeping top to bottom, (c) circle motion, (d) square motion, (e) triangle motion

detect 2D-color, depth image, and skeleton tracking. The advantage of the SDK toolkit is described as follows.

- Noise is eliminated but the skeletal image is maintained with characteristics coordinates of the main body parts.
- The Kinect can extract parameters of each bone joint corresponding to depth image.
- The 3D data of a joint are illustrated by three coordinates (*x, y, z*) which referred to (*horizontal, vertical, depth*) of a joint position.

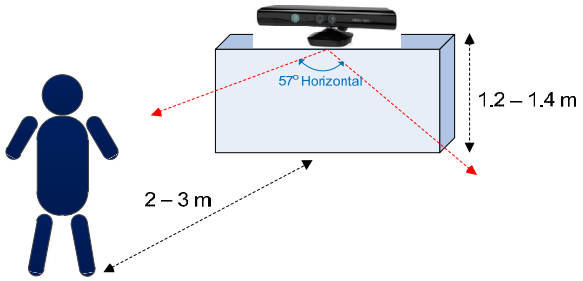


Figure 3. Standard parameters to install Kinect in outer space

For the model of the human body, there are 20 important joints produced from the Kinect data shown in Figure 4.

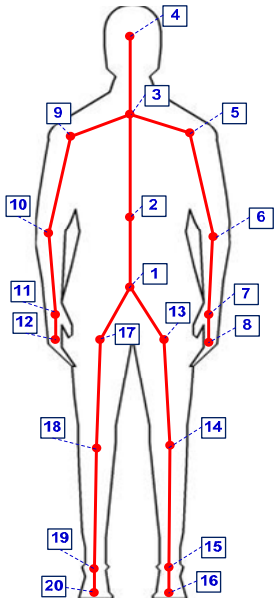


Figure 4. Diagram of the human joints with 3D coordinates (*x, y, z*)

The position of these joints is fixed at the basic places in the human model as the following points: point 1 – "hip centre" joint, point 2 – "spine" joint, point 3 – "shoulder centre" joint, point 4 – "head" joint, point 5 – "shoulder left" joint, point 6 – "elbow left" joint, point 7 – "wrist left" joint, point 8 – "hand left" joint, point 9 – "shoulder right" joint, point 10 – "elbow right" joint, point 11 – "wrist right" joint, point 12 – "hand right" joint, point 13 – "hip left" joint, point 14 – "knee left" joint, point 15 – "ankle left" joint, point 16 – "foot left" joint, point 17 – "hip right" joint, point 18 – "knee right" joint, point 19 – "ankle right" joint, and point 20 – "foot right" joint.

3.2. Skeleton Tracking

The parts of a human body in front of the Kinect camera is detected by a depth stream of Natural User Interface (NUI). The twenty positions of main joints are recorded based on the original coordinate to form a complete human skeletal. Each position is defined by the 3-dimensional coordinate (*x, y, z*), which expresses the place of a skeleton in the real space, compared to the origin position (0, 0, 0) around the sensor. From the point view of the user, the axes described as in Figure 5 in which the *x, y, z*-axis includes the horizontal line toward to the right, the vertical line upward, and the perpendicular line with (*x*, *y*) surface from the sensor, respectively.

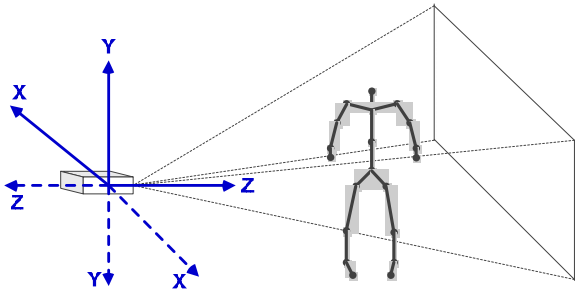


Figure 5. The 3D coordinate system of Kinect camera

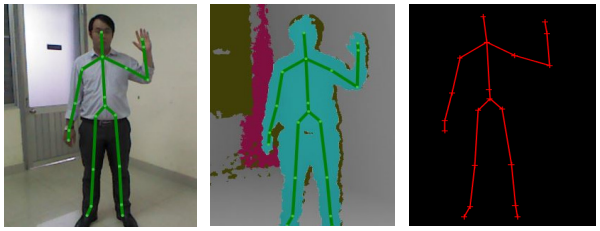
The Kinect system allows tracking color images, depth image and skeletal data automatically at the same time. Thus, the trajectory of each hand gesture is easily detected, and the location data of joints on the skeleton are accurately obtained. The skeleton of the human body is created from the image of body parts which has been produced in a manner related to the specific regions of a depth image by the different density of the feature data shown. Figure 6 depicts these specific regions.

The purpose of this study is to analyze and recognize the features of hand movement. Therefore, the "hand right" joint position is obtained to build 3D data among 20 joints of the human skeleton. The position of each image frame is defined as the transpose matrix with three values (*x, y, z*) in the 3D coordinate as follows.

$$F_1 = [x_1 \quad y_1 \quad z_1]^T. \tag{1}$$

In the case of one sample video with n frames, n = 1, 2, 3,... the matrix of this sample is

$$P = [F_1 \quad F_2 \quad \ldots F_n]. \tag{2}$$



(a) (b) (c)

Figure 6. The joint positions of
(a). Depth image
(b). Body parts
(c). 3D joint proposals

The matrix P could be expressed in detail as follows.

$$P = \begin{bmatrix} x_1 & x_2 & \dots & x_n \\ y_1 & y_2 & \dots & y_n \\ z_1 & z_2 & \dots & z_n \end{bmatrix}. \quad (3)$$

In one frame of image I, the depth dimension of a pixel q is defined by the function:

$$f_\theta(I, q) = d_I\left(q + \frac{t_1}{d_I(d)}\right) - d_I\left(q + \frac{t_2}{d_I(d)}\right) \quad (4)$$

where d_I is the depth of pixel q in the image I and $\theta = (t_1, t_2)$ is the offset between t_1 and t_2.

In this research, the data acquisition is an important process for training and recognition phases to achieve high accuracy results. The obtained data would be processed in the normalization stage of raw data prior to the extraction process using the PCA algorithm. Finally, the resulting data are recognized with HMM method.

4. Proposed method

4.1. Pre-processing raw data

In the data collection step, for each gesture, three persons will perform the hand motions repeatedly at various times. These actions produce images which have different movement trajectories and sizes. However, the coordinate data of joints after the recording step are not similar for different tasks. Such a recording circumstance leads to the large variances in the characteristics of 3D coordinates (x, y, z). Furthermore, the database is not enough reliable to use in training process, so the recognition would not achieve the precision. Therefore, the proposed method would scale the size of images and normalize the obtained coordinate values into a new coordinate system. Figure 7 illustrates the process of pre-processing of raw data.

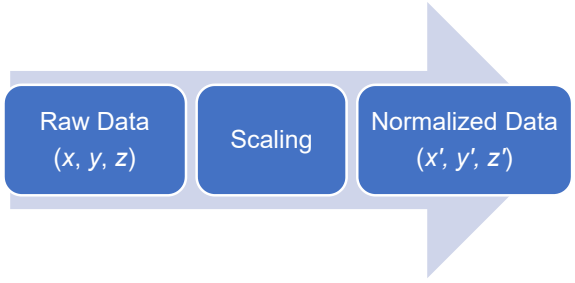


Figure 7. The pre-processing process

The scaling step converts the coordinate of (x, y, z) to a new coordinate (x', y', z'). The new coordinate is determined as follows.

$$x' = \frac{(x - x_{\min}) \times L}{(x_{\max} - x_{\min})} + x_{\min} \quad (5)$$

$$y' = \frac{(y - y_{\min}) \times L}{(y_{\max} - x_{\min})} + y_{\min} \quad (6)$$

$$z' = \frac{(z - z_{\min}) \times L}{(z_{\max} - z_{\min})} + z_{\min} \quad (7)$$

where L denotes the dimension of the square output image compressed from the initial rectangle shape image into an $L \times L$ square image.

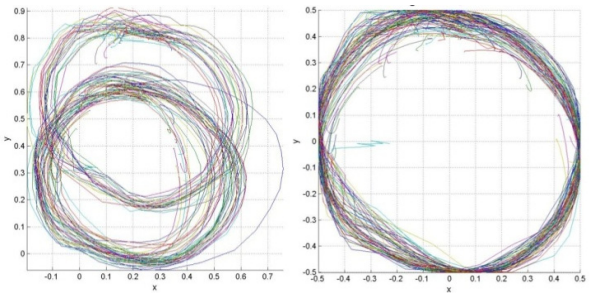
Then, the central point is normalized to the standard size as follows.

$$x' = x - \frac{(x_{\max} - x_{\min})}{2}. \quad (8)$$

$$y' = y - \frac{(y_{\max} - y_{\min})}{2}. \quad (9)$$

$$z' = z - \frac{(z_{\max} - z_{\min})}{2}. \quad (10)$$

Figure 8 shows the change from initial samples to pre-processed data.



(a) (b)

Figure 8. Images of gestures before and after the pre-processing process
(a) The initial recording data
(b) The data after the pre-processing process

4.2. Feature Extraction

The PCA algorithm is employed to extract features from the data sequence of each gesture. The joint data of right-hand motion are pre-processed in a new space where has a reduced dimension. It is capable to illustrate data better and to ensure the data variability at every dimension of new space.

In the PCA algorithm, the calculation of eigenvectors and eigenvalues is very important operation to extract features of the dataset precisely. The eigenvectors corresponding to the largest eigenvalues will describe the most crucial features characterizing the data sample S. Hence, the selection and retain the largest vector k will create a high-reliability space. The principal features of five hand movement gestures are calculated in the following five steps:

(i) The average value $\overline{P_1}$ from the row components is calculated as follows.

$$\overline{P_1} = \frac{1}{k}\sum_{i=1}^{k} P_{1,i}. \tag{11}$$

The average remaining values $\overline{P_2}, \overline{P_3}, \ldots, \overline{P_n}$ can be defined similarity.

(ii) The Standard Deviation (SD) is calculated by the following vector.

$$\overrightarrow{n_1} = \begin{bmatrix} P_{1,1} - \overline{P_1} \\ P_{2,1} - \overline{P_2} \\ \cdot \\ \cdot \\ \cdot \\ P_{n,1} - \overline{P_n} \end{bmatrix}. \tag{12}$$

The remaining vectors $\overrightarrow{n_2}, \overrightarrow{n_3}, \ldots, \overrightarrow{n_n}$ will be determined by the same as the above equation.

(iii) The set of SD vectors are built into a matrix H as follows.

$$H = [\overrightarrow{n_1}\ \overrightarrow{n_2}\ \ldots\ \overrightarrow{n_n}]. \tag{13}$$

A covariance matrix is computed to present the correlation of every vector to the vector space is

$$C = \frac{1}{n-1} H^T H. \tag{14}$$

(iv) The eigenvalue λ is in need to extract the prominent features of model. This eigenvalue is

$$\det(C - \lambda P) = 0. \tag{15}$$

(v) The eigenvector E can be calculated as follows.

$$(C - \lambda P)E = 0. \tag{16}$$

The main purpose of the process is to extract prominent features of five hand motion activities. The set of data after the PCA process is a coefficient matrix of 3300 x 3300 corresponding to feature vector k. In addition, these feature vectors will be trained using HMM to recognize five hand movement gestures separately.

4.3. Hidden Markov Model Method

The HMM method is a group of finite states linked to each other by the transition, in which each state is defined by a set of transition probability, conditional probability or emission probability, the first and second stochastic layer based on the Markov chain. In addition, the second layer of probability describes a sequence of observation but "hidden" from the observer. The key purpose of HMM training is to improve the accuracy of probability so that the state of the certain actual sequence is defined by the sequence of the observation.

There are three main algorithms used in HMM including the forward algorithm for likelihood computation, the Viterbi algorithm for decoding, and the Baum-Welch algorithm for learning. In this research, the Baum-Welch algorithm is also utilized for re-estimation and support in localizing the likelihood using HMM parameters for the set of training data. Each state model is trained for the certain movement to define probability in given test tasks and the presence or absence of a gesture under consideration is determined by a probability threshold. In addition, each state model is determined by the conditional probability of output gestures derived from a hand motion. Figure 9 illustrates the transitional probability of various states in a random process.

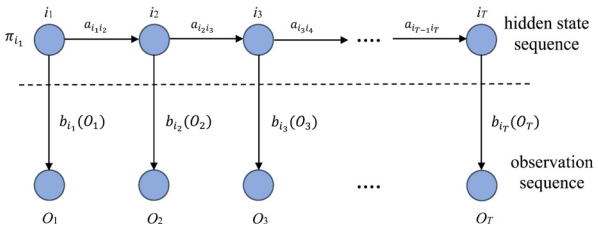


Figure 9. The process of HMM transition state [15]

The transitional probability of various states in HMM, which has N various states $S = \{s_1, s_2, \ldots, s_N\}$, is calculated as follows.

$$a(i,j) = \Pr((P_{t+1} = s_j) \mid (P_t = s_i)) \tag{17}$$

where $a(i, j)$ is the transitional probability from the state s_i to the state s_j and P_t is the state at time t.

An HMM is characterized by three parameters as follows.

(i) The initial state probability is

$$\pi = \{\pi_i\} = \Pr(P_1 = S_i) \tag{18}$$

$$\text{where } 1 \le i \le N \text{ and } \sum_{i=1}^{N} \pi_i = 1.$$

(ii) The state-to-state transitional probability is the following matrix.

$$A = \{a(i,j)\} = \begin{bmatrix} a_{11} & \cdots & a_{1N} \\ \vdots & \ddots & \vdots \\ a_{N1} & \cdots & a_{NN} \end{bmatrix} \quad (19)$$

$$\text{where } a_{ij} \ge 0 \text{ and } \sum_{j=1}^{N} a_{ij} = 1.$$

(iii) The function of probability density of the observed output is

$$B = \{b_j, (k)\} = \Pr(v_k = O_t \mid P_t = S_j) \quad (20)$$

$$\text{where} \sum_{k=1}^{M} b_j(k) = 1.$$

In the equation (20), M is the number of observed signals of a state $V = \{v_1, v_2, \ldots, v_M\}$ for $1 \le j \le N$ and $1 \le k \le M$, and O_t is the observed signal at time t in the sequence of the observations $O = \{O_1, O_2, \ldots, O_T\}$.

HMM is a reliable method for the recognition of specific subjects such as gestures, pose, and speech by applying the well-developed algorithm in order to enhance the recognition performance. Each HMM is represented by a tuple of parameters $\lambda = (A, B, \pi)$. The process of HMM composes two phases:

- *Training:* Recording data of the observation sequence O for the states which are recognized, then optimizing HMM parameters $\lambda = (A, B, \pi)$ to the maximum of the transitional probability $\Pr(O \mid \lambda)$.
- *Recognition:* Based on the observed sequences O, and trained model λ, the phase will define the most satisfied sequence states V^* which maximize the joint likelihood $\Pr(O \mid V \mid \lambda)$ [18].

Additionally, the Baum-Welch algorithm is used to optimize HMM parameters to achieve the highest state-conditional probability $\Pr(O_t \mid p_t = S_i)$. The approach has forward and backward states to compute the forward and backward probability $\alpha_t(i)$ and $\beta_t(i)$, respectively.

At time t, the probabilities of state s_i and the transient probabilities from state s_i to s_j respectively, are as follows.

$$\gamma_t(i) = \frac{\alpha_t(i)\,\beta_t(i)}{\Pr(O \mid \lambda)}. \quad (21)$$

$$\xi_t(i) = \frac{a_t(i)\,a_{ij}\,b_j(O_{t+1})\,b_{t+1}(j)}{\Pr(O \mid \lambda)}. \quad (22)$$

Since the forward and backward probabilities are not repeatedly computed during the process, the Baum-Welch algorithm is used for these parameters re-computed in succession, the results are as follow.

$$\hat{a}_{ij} = \left(\sum_{t=1}^{T-1} \xi_t(i)\right) \Big/ \left(\sum_{t=1}^{T-1} \gamma_t(i)\right). \quad (23)$$

$$b_i(O_t) = \left(\sum_{t=1}^{T} \gamma_t(i).1(O_t = v_k)\right) \Big/ \left(\sum_{t=1}^{T} \gamma_t(i)\right). \quad (24)$$

$$\pi = \gamma_1(i). \quad (25)$$

A new $\hat{\lambda}$ model is produced from the above results of the re-estimation process based on the initial parameters of the λ model in order to generate the observation sequences O.

In this research, the HMM is set up with a set of five possible output symbols. At the training process, each observation sequence, e.g. skeletal features of human, is clustered and classified into five corresponding HMMs to optimize all parameters in these HMMs. As for the recognition action, all data of the obtained sequence will be compared to the trained HMMs to recognize and find out a model which has the highest probability corresponding to each state of the hand movement gesture.

5. Experimental Results

In the research, the sample features were extracted from offline databases using the PCA algorithm and then these features were trained using the HMM method for recognition. Experimental results were shown to describe the effectiveness and efficiency of the proposed method.

Table 1. The cluster of column data matching to the human motion gestures

No.	Hand gesture	Column data
1	Sweeping right to left	1 – 480
2	Sweeping top to bottom	481 – 960
3	Circle motion	961 – 1680
4	Square motion	1681 – 2580
5	Triangle motion	2581 – 3300

The experiments were made with three persons of 21 – 50 years old and the heights were between 166 and 172 cm to test the proposed method. Each person performed five hand gestures "Sweeping right to left", "Sweeping top to bottom", "Circle motion", "Square motion", and "Triangle motion" corresponding to five states of the HMM and each gesture is performed twenty times. Therefore, there are 300 video samples. Each sample was then extracted from 8 to 15 feature frames depending on the complexity of each gesture. The obtained data of five gestures corresponded with 3.300 image frames for training and recognition. After

that, the frames were converted into a matrix with the size of 3 × 3.300. Table 1 shows the column data of the matrix.

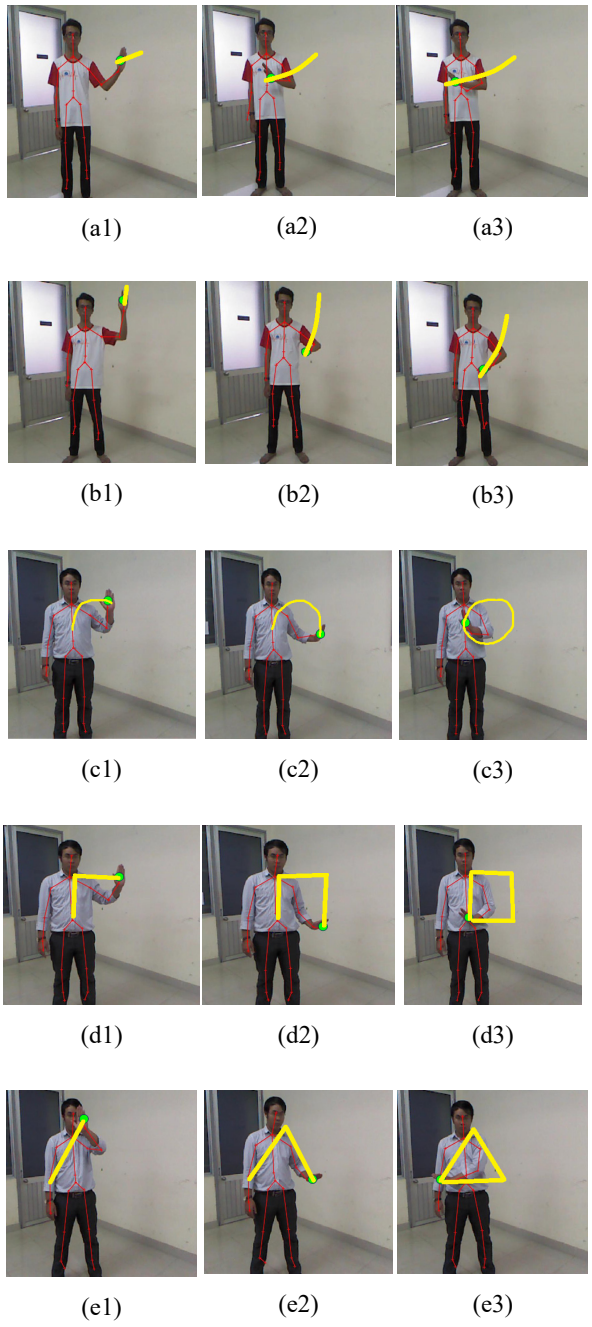


Figure 10. The process of recording samples:
(a1), (a2), (a3)_" Sweeping right to left"
(b1), (b2), (b3)_" Sweeping top to bottom"
(c1), (c2), (c3)_" Circle motion"
(d1), (d2), (d3)_"Square motion"
(e1), (e2), (e3)_" Triangle motion"

In practice, the hand movement activities were recorded as in Figure 10. Of the images of Figure 10, the "hand right" joint – point 12 (see Figure 4) describes the most basic features of hand movement. For this reason, obtained data at "hand right" joint are important to analyse different features of each hand gestures compared to the remaining joints.

5.1. Gestures recognition using the HMM method

The proposed method using the HMM was implemented on 3.300 image frames of 300 video samples for training and 60 testing videos per each hand gesture for recognition. Table 2 illustrates the recognition results of five hand movement gestures with the average accuracy of 95.66%. In detail, the recognition result of circle motion was highest at 98.3%. That means 59/60 testing videos were defined accurately. Nevertheless, the square motion showed the lowest recognition result with 93.3% corresponding to 56/60 testing videos recognized correctly. Similarly, other types of the hand movement gestures including the sweeping right to left, sweeping top to bottom, triangle motion produced the results of 96.7%, 95.0%, and 95.0%, respectively as shown in Table 2.

Table 2. Recognition results of hand movement gestures for offline database

Accuracy (%)	Sweeping right to left	Sweeping top to bottom	Circle motion	Square motion	Triangle motion	Undefined state
Sweeping right to left	**96.7**	0.0	0.0	1.7	0.0	1.6
Sweeping top to bottom	0.0	**95.0**	0.0	1.7	1.7	1.6
Circle motion	0.0	0.0	**98.3**	1.7	0.0	0
Square motion	0.0	0.0	3.3	**93.3**	1.7	1.7
Triangle motion	0.0	0.0	0.0	1.7	**95.0**	3.3

5.2. Online hand movement recognition

Another experiment of the research is online recognition. The recognition program was built with the thresholds that were obtained from the HMM mothed after the training process. Each one of gestures was tracked in the real-time setting to be recognized immediately. For one gesture, the recognition result was tested with 60 samples. Table 3 illustrates the accuracy of this experiment. The statistics in the table showed that the overall accuracy was about 91%. The highest result was 95% equal to 57/60 samples accuracy for the circle motion recognition. The least result

was 88.3% corresponding to 53/60 samples accuracy for the square movement recognition. The remaining classifications had following results: sweeping right to left with 90%, sweeping top to bottom with 91.7%, and triangle motion with 90%.

Table 3. Recognition result of hand movement gestures in online

Accuracy (%)	Sweeping right to left	Sweeping top to bottom	Circle motion	Square motion	Triangle motion	Undefined state
Sweeping right to left	**90.0**	0.0	0.0	3.3	1.7	5.0
Sweeping top to bottom	0.0	**91.7**	0.0	5.0	1.7	1.6
Circle motion	0.0	0.0	**95.0**	3.3	1.7	0
Square motion	1.7	0.0	6.7	**88.3**	0.0	3.3
Triangle motion	0.0	0.0	3.3	1.7	**90.0**	5.0

In the paper, the Kinect camera system was used to obtain hand motion data of three persons for the hand movement recognition system. The skeletal data of five gestures per person were collected into sample data to extract prominent features. Besides, the HMM method was utilized to train the feature data and then to recognize hand motion gestures with the average accuracy of offline was 95.66% and online was 91.00%. It is obvious that the proposed method can be applied effectively and efficiently to the hand gesture recognition system.

6. Conclusions

In the research, the images of three persons with five gestures obtained from a Kinect system were converted into skeletal data. These data were extracted crucial features using the PCA method. For classification, a HMM method was employed for training the 3D data of hand right and then for recognizing hand movement gestures. The experimental results showed that the average accuracies of offline and online recognition were 95.66% and 91.00%, respectively. The results demonstrate that the proposed method offers the high-quality recognition.

For further applications, the hand gesture recognition system can be used to improve free-touch interface for controlling programs on the computer such as virtual reality E-games, windows' cursor, and document processor in the Operating Room...

Acknowledgements.

The authors would like to sincerely thanks the support of Saigon University, Vietnam, with Project No. CS2016-51. Furthermore, we would send to thank students and colleagues who have helped us complete this research.

References

[1] N. Eric and J.-W. Jang, "Kinect depth sensor for computer vision applications in autonomous vehicles," in *9th International Conference on Ubiquitous and Future Networks*, pp. 531-535, IEEE. 2017.

[2] E. E. Stone and M. Skubic, "Fall detection in homes of older adults using the Microsoft Kinect," *IEEE journal of biomedical and health informatics,* vol. 19, no. 1, pp. 290-301, 2015.

[3] W.-S. Yu and K.-Y. Huang, "Implementation of media player simulator using Kinect sensors," in *2017 International Conference on System Science and Engineering (ICSSE)*, pp. 204-209, IEEE. 2017.

[4] N. Kameyama and K. Hidaka, "A sensor-based exploration algorithm for autonomous map generation on mobile robot using Kinect," in *2017 11th Asian Control Conference (ASCC)*, pp. 459-464, IEEE. 2017.

[5] K. Shimura, Y. T. Ando Y and M. M, "Research on person following system based on RGB-D features by autonomous robot with multi-Kinect sensor," in *2014 IEEE/SICE International Symposium on System Integration (SII)*, pp. 304-309, IEEE. 2014.

[6] R. Dubey, N. Bingbing and P. Moulin, "A depth camera based fall recognition system for the elderly," in *International Conference Image Analysis and Recognition*, pp. 106-113, Springer 2012.

[7] T. H. An, T. Q. Phuc, N. T. Hai and T. T. Mai, "Support vector machine algorithm for human fall recognition Kinect-based skeletal data," in *2015 2nd National Foundation for Science and Technology Development Conference on Information and Computer Science (NICS)*, pp. 202-207, IEEE. 2015.

[8] J. Shin and C. M. Kim, "Non-touch character input system based on hand tapping gestures using Kinect sensor," *IEEE Access,* vol. 5, pp. 10496-10505, 2017.

[9] M. Elgendi, F. Picon and N. Magenant-Thalmann, "Real-time speed detection of hand gesture using Kinect," in *Workshop on Autonomous Social Robots and Virtual Humans, The 25th Annual Conference on Computer Animation and Social Agents (CASA 2012)*, 2012.

[10] F. Liu, B. Du, Q. Wang, Y. Wang and W. Zeng, "Hand gesture recognition using kinect via deterministic learning," in *2017 29th Chinese Control and Decision Conference (CCDC)*, pp. 2127-2132, IEEE. 2017.

[11] X. Xue, W. Zhong, L. Ye and Q. Zhang, "The simulated mouse method based on dynamic hand

gesture recognition," in *2015 8th International Congress on Image and Signal Processing (CISP)*, pp. 1494-1498, IEEE. 2015.

[12] A. Dubois and F. Charpillet, "Human activities recognition with RGB-depth camera using HMM," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 4666-4669. IEEE. 2013.

[13] P. T. Hai and H. H. Kha, "An efficient star skeleton extraction for human action recognition using hidden Markov models," in *2016 IEEE Sixth International Conference on Communications and Electronics (ICCE)*, pp. 351-356, IEEE. 2016.

[14] H. S. Chen, H. T. Chen, Y. W. Chen and S. Y. Lee, "Human action recognition using star skeleton," in *Proceedings of the 4th ACM international workshop on Video surveillance and sensor networks*, pp. 171-178, 2006.

[15] L. Rabiner, "First hand: the hidden Markov model," in *IEEE Global History*, 2013.

[16] A. Dubois and F. Charpillet, "Measuring frailty and detecting falls for elderly home care using depth camera," *Journal of Ambient Intelligence and Smart Environments,* vol. 9, no. 1, pp. 469-481, 2017.

[17] M. D. Uddin, N. D. Thang, J. T. Kim and T. S. Kim, "Human activity recognition using body joint-angle features and hidden Markov model," *Etri Journal,* vol. 33, no. 4, pp. 569-579, 2011.

[18] D. O. Tanguay, "Hidden Markov models for gesture recognition," Doctoral dissertation, Massachusetts Institute of Technology, 1995.