Journal of Theoretical and Applied Information Technology
30th September 2016. Vol.91. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

DYNAMIC GESTURE RECOGNITION FOR NATURAL HUMAN SYSTEM INTERACTION

[12]Hajar HIYADI, [1]Fakhreddine ABABSA, [1]Christophe MONTAGNE, [2]El Houssine BOUYAKHF, [2]Fakhita REGRAGUI

[1]IBISC laboratory, University of Evry, Evry, FRANCE

[2]LIMIARF laboratory, University of Rabat, Rabat, MOROCCO

E-mail: hajar.hiyadi@ibisc.univ-evry.fr, fakhr-eddine.ababsa@ibisc.univ-evry.fr, christophe.montagne@ibisc.univ-evry.fr, bouyakhf@fsr.ac.ma, regragui@fsr.ac.ma

ABSTRACT

This paper addresses two problems: 3d dynamic gesture recognition and gesture misallocation. In order to solve these problems, we propose a new approach which combines Hidden Markov Models (HMM) and Dynamic Time Warping (DTW). The proposed approach has two main phases; first, recognizing gestures using a hidden Markov model. Second, avoiding misallocation by rejecting gestures based on a threshold computed using DTW. Our database includes many samples of five gestures obtained with a Kinect and described by depth information only. The results show that our approach yields good gesture classification without any misallocation and it is robust against environmental constraints.

Keywords: *Dynamic Gesture Recognition, Hidden Markov Model, Dynamic Time Warping, Kinect, Depth image.*

1. INTRODUCTION

Gesture recognition has motivated many researchers concerned with human system interaction through gestures. In fact, human gesture recognition is becoming a way to input information in gaming, consumer, mobile devices and robots. In order to accomplish this task, computer vision, and pattern recognition techniques, namely, feature extraction, object detection, clustering, and classification, have been successfully used [1] [2]. A gesture can be static or dynamic. Accordingly, there are two kind of gesture recognition approaches static hand posture recognition and dynamic hand gesture recognition. In this paper, we focus on dynamic gesture recognition which can be performed by the upper body members. Unlike static gesture, dynamic gesture communicates much more information whence the need of a robust recognition system. A gesture recognition system contains different steps: detection, tracking, gesture extraction and finally classification. Detection can be done using template matching [3] [4] [5]. Skin color can be an important clue for both detection and track hand. In [6], skin color is used to extract hand then track the center of the corresponding area. The extracted surface has an elliptical shape into each chrominance space. Consequently, considering this fact, the authors proposed a skin color model named elliptical contour. This idea was extended in [7] to detect and localize the head and hands. Segmentation is also a good method which can keep pertinent objects and remove noninteresting areas. In [8] the authors use segmentation method based on clustering in order to detect hand by combining two methods k-means and expectation maximization. These methods use exclusively the RGB imges. Nonetheless, approaches based on color are highly influenced by environmental conditions as lighting variations and background complexity. Accordingly, new studies aim to combine new and robust information like depth. In fact, depth information provided by a depth sensor can improve the performance of gesture recognition systems. In [9] [10] [11] [12], depth and color are combined for segmentation and tracking. In [13], the authors propose a system that can learn, from examples, how users indicate objects using speech and gesture. In order to track, filtering methods like the Unscented Kalman Filter has been used in [14] and points of interest in [15].

Dynamic gestures provide more and relevant information than static ones. However, their spatial-temporal variability makes their recognition very difficult and problematic. In fact, the same gesture can be performed differently so that his shape, duration and speed change. After their wide success in speech, handwriting and character recognition [16] [17], Hidden Markov models were successfully

Journal of Theoretical and Applied Information Technology
30th September 2016. Vol.91. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

used in gesture recognition [18] [19] [20]. Indeed, Hidden Markov Models can model spatial-temporal time series and keep the spatial-temporal identity of gesture. Other mathematical models extracted from Hidden Markov Models are also used in order to extract and recognize gestures like Input-Output Hidden Markov Model (IOHMM) [21] and Hidden Conditional Random Fields (HCRF) [22]. Gesture recognition can be also done using Dynamic Time warping by comparing the distance between gestures of test and gestures of dataset [23].

In this paper, we propose a robust gesture recognition system based on a fusion of Hidden Markov Models and Dynamic Time Warping. Our goal is to propose a system which can solve many problems in gesture recognition. The system (1) should recognize natural gestures provided their variability among individuals; (2) should recognize gestures based on relevant information not affected by lighting change , and (3) have the capacity to detect a badly performed gesture and would reject it rather than misclassifying it. Figure 1 illustrates the main framework of the proposed system. Body tracking is done using the skeleton algorithm provided by the Kinect SDK. From the extracted 3d coordinates (x, y and depth) of joints, angles between each two joints are calculated. Gestures are modeled and classified by using HMM.

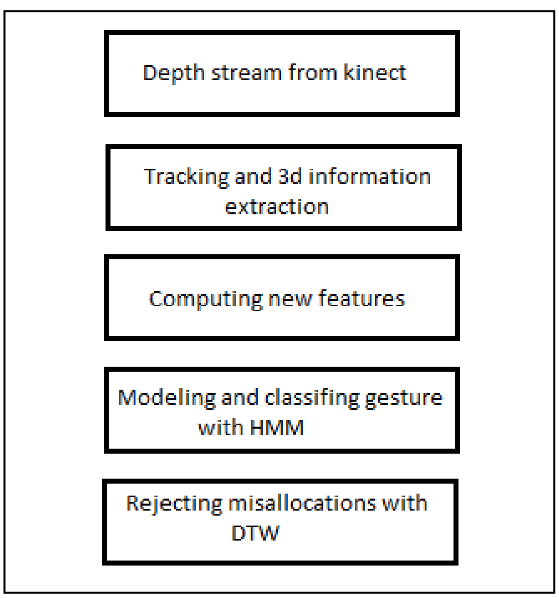

Figure 1: Flowchart of the proposed gesture recognition system

In the end, DTW is fed by the output of HMM to calculate the distance between the gesture studied and a reference sequence of the gesture class found by HMM. The result given by DTW is computed to a fixed threshold which was calculated beforehand. If the DTW distance is above the threshold then the gesture is rejected.

2. PROPOSED METHOD

In the domain of human-system interaction, the purpose of our work is to recognize 3d dynamic gestures. To make this interaction natural and intuitive for users, we defined gestures similar to human-human interaction ones. Figure 2 shows the five gestures that we propose to recognize {come, recede, point to the right, point to the left, stop}. Our gesture recognition system includes two major phases. The first phase is tracking and data extraction. The second one is classification.

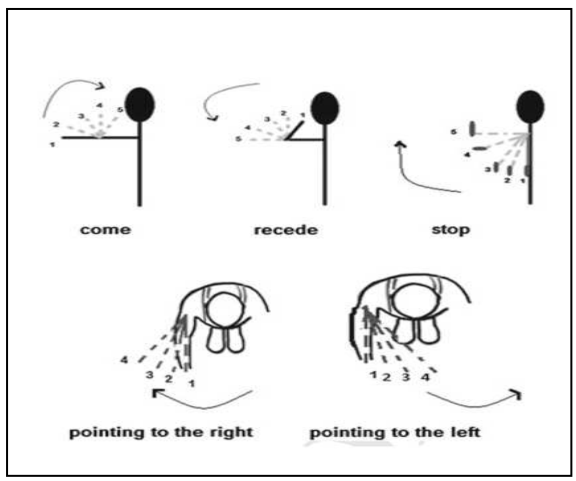

Figure 2: The five gesture kind to recognize

2.1 Tracking and data extraction:

A robust gesture recognition system requires a robust tracking. Most tracking methods use color information. However, the color is mostly influenced by lighting variations and occlusions which make it an instable clue. Consequently, color based tracking methods do not succeed at several times. In our work, we use Kinect to extract more accurate data. Tracking is done by the skeletal tracking method provided by the Kinect SDK [24]. This method tracks 20 joints of the body of each detected person. For each joint, it creates an ID and

Journal of Theoretical and Applied Information Technology
30th September 2016. Vol.91. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

gives the x, y positions and the corresponding depth value. While performing the five gestures defined beforehand, we noticed two facts. The first one is that we need only the upper body joints information. The second is that, for each gesture, one angle changes more than others. Taking account the second remark, we chose to create a descriptor for each gesture based on the active angles while executing it. Finally, we found three major angles which the variations are different from gesture to another. Figure 3 shows the angles joints α, β, and γ that refer respectively to elbow, shoulder and armpit angles.

Figure 3: α, β et γ angles

Each angle is computed from its joints coordinates, so:

- The angle α is calculated from the 3d coordinates of the elbow, wrist and shoulder joints.
- The angle β is calculated from the 3D coordinates of the shoulder, elbow and shoulder center joints.
- The angle γ is calculated from the 3D coordinates of the shoulder, elbow and hip joints.

While executing a gesture, the three angles are calculated in real time and their variations are stored as vectors:

$$V_\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_T] \quad (1)$$
$$V_\beta = [\beta_1, \beta_2, \ldots, \beta_T] \quad (2)$$
$$V_\gamma = [\gamma_1, \gamma_2, \ldots, \gamma_T] \quad (3)$$

Each gesture will be described with the combination of these three vectors as follows, where T refers to the length of the gesture sequence:

$$V = [\alpha_1, \alpha_2, \ldots, \alpha_T, \beta_1, \beta_2, \ldots, \beta_T, \gamma_1, \gamma_2, \ldots, \gamma_T] \quad (4)$$

In addition, among these three angles, one of them is very active during the gesture comparing to the two others which change lightly and take at most two or three values. This main angle is α in both come and recede gestures, β in both point to the right and point to the left gestures, and γ in stop gesture. Table 1 gives more information of angles variations for each gesture.

Table 1: The main angle's variations in each gesture

	α	β	γ
Come	180°→30°	-	-
Recede	30°→180°	-	-
Point to the right	-	90°→150°	-
Point to the left	-	90°→40°	-
Stop	-	-	30°→80°

This description of gestures depending on angles gives to each gesture a unique description which makes its classification relevant. We consider the vector V as the descriptor of the gesture; as a result, this descriptor will be fed as an input to our gesture classification system.

2.2 Gesture classification system:

2.2.1 Overview:

Our classification system combines two of the best methods in classification domain which are Hidden Markov Models (HMM) and Dynamic Time Warping (DTW). In one hand, Hidden Markov models were successfully used in gesture recognition. They can model spatial-temporal time series and keep the spatial-temporal identity of gesture. Dynamic gestures provide more and relevant information than static ones. However, their spatial-temporal variability makes their recognition very difficult and problematic. In fact, the same gesture can be performed differently, so that his shape, duration and speed change. But HMMs still are able to preserve the identity of spatiotemporal gesture even though its speed or duration varies. In the other hand, while we work with time series data, we use DTW to calculate the similarity between two sequences with different duration and speed. In fact, DTW wraps the sequences and provides a distance between them. First of all, gestures are classified with HMM. They are recognized by finding the best probability

Journal of Theoretical and Applied Information Technology
30th September 2016. Vol.91. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

among the five probabilities of belonging to the five classes. Thus, a class label is giving as an output of HMM-based recognizing process. Second, we measure using DTW, the similarity between the variations of the angle that characterizes the gesture kind recognized by HMM and, a variation sequence of this same angle and the same kind of gesture. This sequence was fixed beforehand as a reference. As a result of this computing, a distance between the two sequences is given and compared to a pre-calculated threshold. If this distance is less than the threshold, we consider that the result given by HMM is confirmed and we keep it. Else, if the distance is above the threshold, we consider that the gesture was misallocated by HMM and we reject it. Figure 4 illustrates the steps of our recognition system. First, HMM method will classify a given gesture test (Gtest) into one of the five classes. As a result, HMM method will give the type of the gesture (for example come). As mentioned before, the angle characterizes the gesture come is α which means elbow angle. Thus, we take the first part of the gesture sequence (Gtest) which corresponds to α angle variation, and we take a reference sequence of α angle variation of come gesture from the database. Then, we calculate the distance between these two sequences using DTW algorithm. The resulting distance is compared to a threshold that was fixed for the gesture come.

2.2.2 HMM based classification method:

An HMM can be written as $\lambda = (A, B, \pi)$ where:

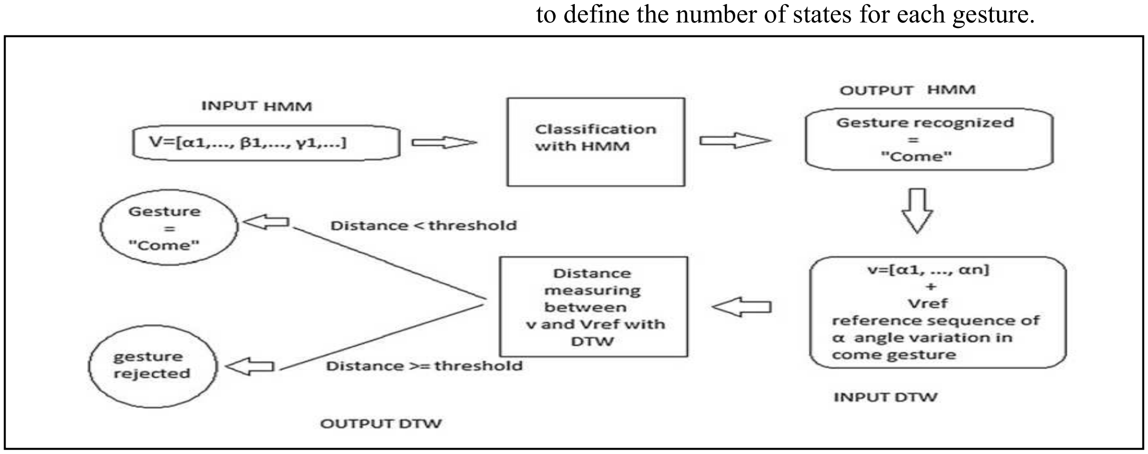

Figure 4: Our gesture recognition system combining HMM and DTW

- $A = \{a_{ij}\}$ is the transition matrix with a_{ij} is the transition probability from the state i to the state j.
- $B = \{b_{im}\}$ is the observation matrix with b_{im} is the probability of generating the symbol m in the state i.
- $\pi = \{\pi_i\}$ is an initial probability distribution for each state with π_i=Prob(state= i at t=1).

We use Baum-Welch algorithm for training and Forward algorithm for the evaluation. Among the different topologies used in HMM, we choose Left-Right topology where each state can go back to itself or jump to the following state only. We chose this model because it is the best for modeling-order-constrained time-series whose proprieties change over time sequentially. This topology is shown in Figure 5.

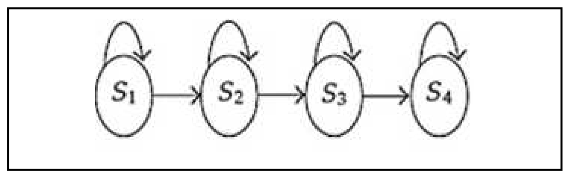

Figure 5: Left-Right topology

We created an HMM for each gesture kind. First of all, every parameter should be initialized. We begin with the number of states. Given that gestures are different in complexity and duration, the number of states is different from a gesture to another. In the evaluation section, we show an experiment in order to define the number of states for each gesture.

As a result, we found that our gestures need from 8 to 12 states. The second parameter to initialize is the initial probability distribution which is also, different from a gesture to another. For a gesture

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

with 8 states, the initial probability distribution vector is:

$$\Pi=(1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ ) \quad (5)$$

To be sure that the HMM starts from the first state of the gesture, we gave to the first element of the vector the value 1 and 0 to others. It means, the probability of starting the gesture with the first state is 1, and the probability of starting the gesture with one of the remaining states is 0.

The third parameter is the transition matrix. As we chose the Left Right topology we have two transitions: from state i to i (itself), and from state i to state i+1. Thus, the transition matrix $A=\{a_{ij}\}$ will be written as (for the case 8 states):

- $a_{ij}=a_{ii}$ / if j=i
- $a_{ij}=1-a_{ii}$ / if j=i+1
- $a_{ij}=0$ / otherwise

With i and j are two states and a_{ii} is initialized by a random value.

The last parameter is the observation matrix that can be written as:

- $B=\{b_{im}\}$

Where i is a state, m is a symbol and b_{im} is initialized by a random value.

2.2.3 DTW distance computing between two sequences:

We are interested to compare sequences with DTW to get the distances between them. The sequences which we use contain data with 1 dimension because every datum represents a measure of angle. The classification by using DTW does not require a base of training. Indeed, it is enough to have as a sequence of good reference which represents perfectly the gesture. The classification is made in the following way: Every given sequence of test is compared with all the reference sequences. In every comparison we obtain a distance. So, the sequence of test belongs to the class represented by the reference sequence which gives the minimal distance. The following algorithm represents the instructions realized to classify a sequence of test T among i classes, each class is represented by a reference sequence Ri, with dist_min is the minimum distance and $dist\_dtw(R_i,T)$ is the DTW distance between The test sequence T and the reference sequence R_i.

Algorithm:

1. ***Start***
2. Reference sequences {R1, R2, … Ri}; test sequence T;
3. dist_min=10000000;
4. For j=1 to i do:
   If (dist_dtw(R_i,T) < dist_min) do :
   [
   dist_min=dist_dtw(R_i,T);
   class_index=i ;
   ]
   End if
5. i=i+1 ;
6. ***End***

3. EVALUATION

3.1 Experimental protocol

An experimental protocol was given to the subjects, which explains the five gestures. We did not fix the duration of the gestures. The person can choose whether to perform the gestures slowly or speedy. The kinect sensor must remain stable and the distance between it and the person should be between 80 cm and 3 m in order to detect the body properly. No obstacle should be present between the person and the kinect.

Figure 6.1: Case 1: the distance is more than 3m.

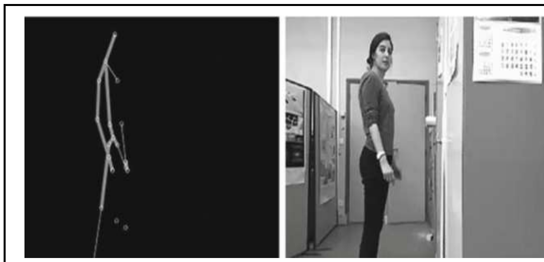

Figure 6.2: Case 2: the person is not in front of the kinect.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The environment can be more or less crowded. While performing a gesture, the person should remain in front of the kinect. Figure 6 illustrates the two cases when the kinect cannot detect totally the body.

3.2 HMM training and evaluation

In order to build our database, we appealed to 20 persons. Everyone is asked to execute the five gestures that we have defined before. Each gesture is executed 5 times per person. So, finally we have generated 500 sequences, 250 are used for training and 250 for testing. In the training phase, the Baum-Welch algorithm [25] is used to train the initialized HMMs parameters $\lambda = (A, B, \pi)$. Our system is trained on 50 sequences of a discrete vector for each kind of gesture by using the LR topology with the number of states ranging from 3 to 14. According to the forward algorithm, the other 50 video sequences for each type of gesture are tested, in which the algorithm is built using the discrete vector, matrix A, matrix B, and vector π as inputs. We use the forward algorithm to compute the probability of the discrete vector sequences for all the five HMM models with different states. Thereby, the gesture path is recognized corresponding to the maximal likelihood of 5 gestures HMM models.

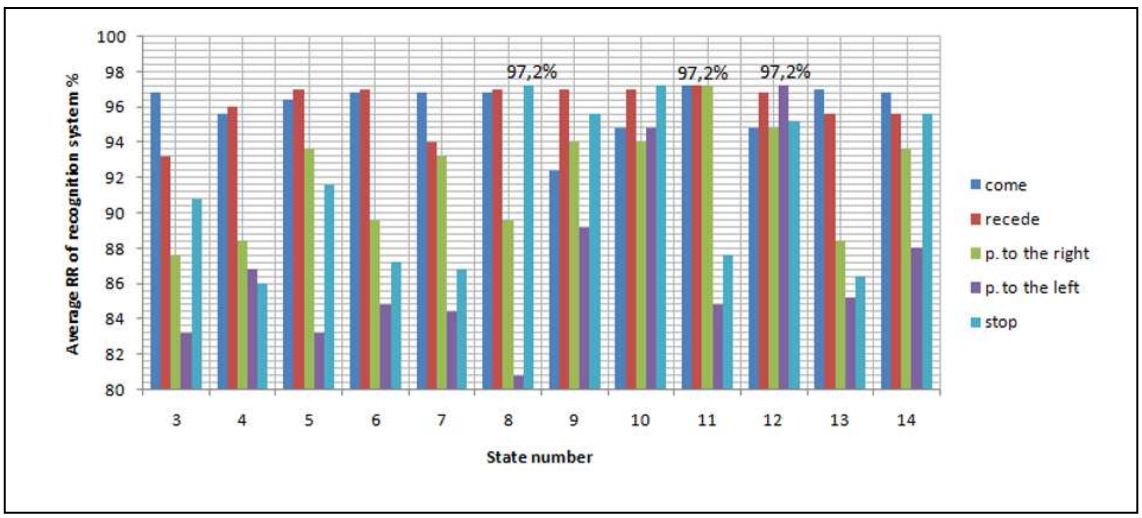

Figure 7: Average recognition rate of the recognition system with varying state number of the five HMMs from 3 to 14.

We choose the state number of the HMM for each gesture according to the experiment results and find that the recognition rate is maximal when the state numbers of gestures "come", "recede" and "point to the right" are 11, and that of gesture "point to the left" is 12, and that of the gesture "stop" is 8. Therefore, we use this setting in the following experiments. Figure 7 represents the average recognition rate of the recognition system according to the HMMs state number which varies from 3 to 14. In fact, each time, we change the state number of one HMM and compute the average recognition rate of the whole of the recognition system.

3.3 Angles variations

As mentioned beforehand, each gesture is characterized by one angle which varies more than the others. This note was done after plotting the three angles variations for each gesture. The results are illustrated in the following images (Figure 8, 9, 10, 11, 12). As result, gestures can be distinctive and yields our system to well recognize them.

Journal of Theoretical and Applied Information Technology
30th September 2016. Vol.91. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

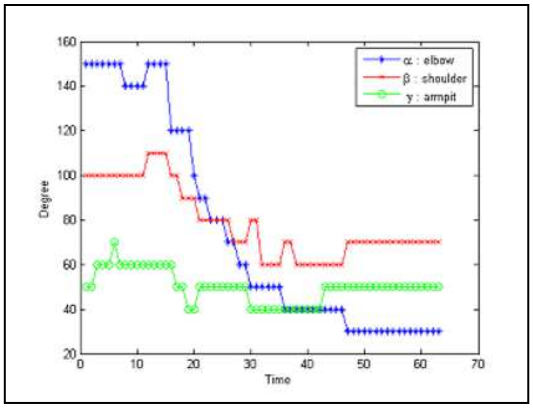

Figure 8: Angles Variations For Come Gesture

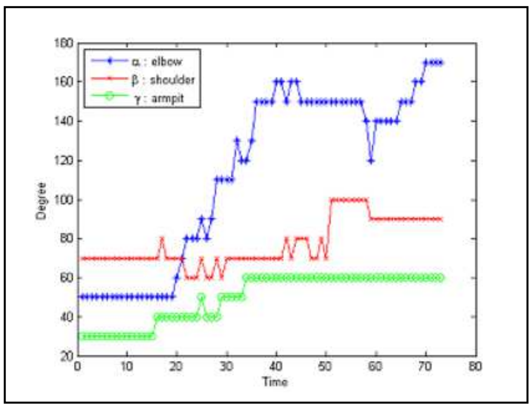

Figure 9: Angles Variations For Recede Gesture

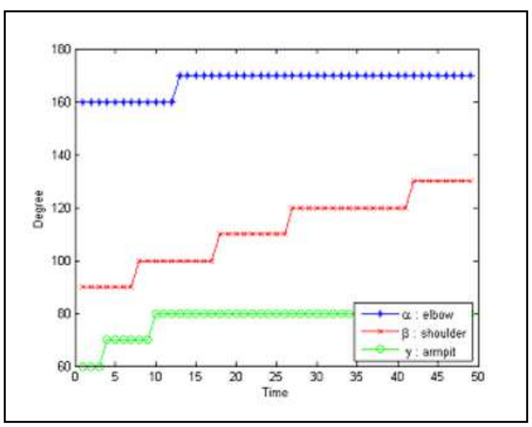

Figure 10: Angles Variations For Point To The Right Gesture

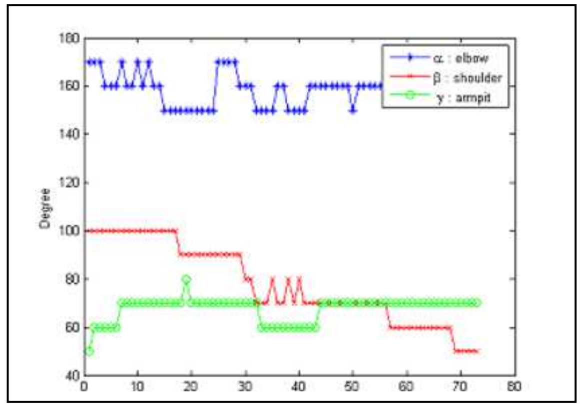

Figure 11: Angles Variations For Point To The Left Gesture

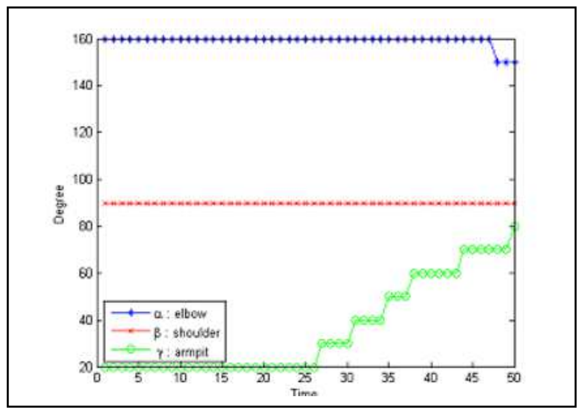

Figure 11: Angles Variations For Stop Gesture

3.4 DTW threshold calculating

The rejection method is based on thresholds. For each class of gesture, we have calculated its threshold by an empirical way. First of all, we consider a reference sequence for each gesture. For "come" class, the reference sequence contains the variations of the angle α throughout a "come" gesture. For "recede" class, the reference sequence contains the variations of the angle α throughout a "recede" gesture. For "point to the right" class, the reference sequence contains the variations of the angle β throughout a "point to the right" gesture. For "point to the left" class, the reference sequence contains the variations of the angle β throughout a "point to the left" gesture. And finally for "stop" class, the reference sequence contains the variations of the angle γ throughout a "stop" gesture. The threshold of a gesture class corresponds to the maximum distance between its reference sequence and the 50 sequences of test. The distance is

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

calculated using DTW algorithm. Using the fact that the distance which is used is the Euclidean distance, Table 2 gives the threshold computed for each kind of gesture.

Table 2: The Threshold Of Rejecting For Each Gesture

Gesture type	The threshold calculated
Come	52 800
Recede	29 400
Point to the right	85 800
Point to the left	39 101
Stop	29 100

3.5 Gesture recognition results

Table 3 shows the recognition results using HMM classification only. It can be seen that the proposed method can greatly improve the recognition process, especially for opposite gestures such as “come” and “recede”, “point to the right” and “point to the left”. Moreover, the average time of recognition for a given gesture sequence is 0.1508s. We did not find any confusion between “come” and “recede” due to the fact that the angle α decreases in “come” gesture and increases in “recede” one. Similarly, the same behavior is observed for point to the right and point to the left. However, 7 gestures are misclassified. By combining the DTW process the recognition system is improved and no gesture is misclassified but it is rejected instead as shown in Table 4. In both tables (3 and 4), lines represent real gesture and columns represent predicted gesture with 1, 2, 3, 4, 5 means respectively come, recede, point to the right, point to the left and stop gestures.

Table 3: Confusion Matrix And Recognition And Misclassification Rates Using HMM Only

	1	2	3	4	5	Accuracy
1	50	0	0	0	0	100%
2	0	50	0	0	0	100%
3	0	0	49	0	1	98%
4	0	0	0	48	2	96%
5	1	0	0	3	46	92%
Average accuracy rate 97,2%						
Average misclassification rate 2,8%						

Table 4: Confusion Matrix And Recognition And Misclassification Rates By Combining DTW With HMM

	1	2	3	4	5	Unknown gesture	Accuracy
1	50	0	0	0	0	0	100%
2	0	50	0	0	0	0	100%
3	0	0	49	0	0	1	98%
4	0	0	0	48	0	2	96%
5	0	0	0	0	46	4	92%
Average accuracy rate 97,2%							
Average misclassification rate 0%							

Table 5 gives a comparison of our approach with that of Ye et al. in [26]. The authors use raw, roll and pitch orientations of elbow and shoulder joints of the left arm. Their database contains five gestures trained by only one person and tested by two. The gesture duration is fixed beforehand. In offline mode, the accuracy of recognizing gestures executed by the person who did training was found to be 85% with their method and 97,2% with ours. And without training, the recognition accuracy reached 73% with their method and 82% with ours. Moreover, the gestures we have defined for human system interaction are natural. They are almost the same that we use daily and between people. However, most methods, in the state of the art, are based on constrained gestures which use signs. This kind of gestures is not natural. Furthermore, our proposed approach is based only on depth information that is what makes it very robust against the environment complexity and illumination variation.

Table 5: Comparison Between The Performance Of Our Approach And Ye And Ha’s Approach

Methods	Ye and Ha [26]	Our approach
Gesture nature	Dynamic	Dynamic
Used Info.	Orientations	Angles
Gestures number	5	5
Joints number	2	5
Used data	Segmented	Raw
Classification	HMM	HMM/DTW
Database	75	500
People for training	1	20
People for test	2	20
Gesture duration	Fixed	Not fixed
Accuracy	85%	97,2%

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

4. CONCLUSION AND FUTUR WORK

We described an efficient system for 3D natural and dynamic gesture recognition for human system interaction by combining two robust methods of classification (HMM and DTW).

We have identified five deictic gestures which can be recognized using only depth information. The idea is to extract the 3D coordinates of the joints of the upper part of the human body, and then compute the angles corresponding to these joints. The angle variation among the gestures is used as the input of Hidden Markov Models HMMs. HMM method affects the given gesture to one of the five classes corresponding to the maximum probability. Based on this result, the DTW measures the similarity between the variations sequence of the main angle that characterizes the gesture class which HMM method gave as output and its reference sequence. The output distance is compared to the threshold corresponding to the same class; if the distance is less than the threshold then we keep the HMM result else we reject the gesture. Experimental results presented in this paper confirm the effectiveness and the efficiency of the proposed approach.

Our approach yields potentially a better recognition rate compared to other works. Indeed, the recognition rate can reach up to 100 for certain gestures. Most of the works in state of the art are based, totally or partially, on colour information which can be influenced by the light change. We propose a system which does not need the colour information for recognizing gestures. It relies only on depth information. In addition, we give below some characteristics of the proposed recognition system: First, during the training phase, it simply saves the gesture when run. Second, the system can recognize gestures even if the distance or the location of people changes. Third, the system is able to recognize the gesture although the speed of gestures can vary from one person to another. Finally, the change in the duration of a gesture from one person to another does not affect the recognition.
The proposed work has some limitations. First, to recognize more kind of gestures, it is necessary to train them beforehand. Second, the gesture classification is done in offline. Finally, each gesture can be recognized separately. In future work, we intend to expand the data base by adding more gestures which can be simple or continuous (two or more successive gestures). Our next objective is to recognize online all gestures performed successively. In order to achieve this goal, some other techniques may be combined as sliding window and bag of words.

REFRENCES:

[1] V.I. Pavlovic, R. Sharma, T.S. Huang, "Visual interpretation of hand gestures for human computer interaction", *IEEE Trans. Pattern Anal. Mach. Intell.* 19(7), 677–695 (1997).

[2] A. Wexelblat, "An approach to natural gesture in virtual environments", *ACM Transactions on Computer-Human Interaction* 2(3), 179–200 (1995)

[3] A. Barczak and F. Dadgostar, "Real-time hand tracking using a set of cooperative classifiers based on haar-like features", *Research Letters in the Information and Mathematical Sciences*, 2005, pp 29-42.

[4] Q. Chen, N. Georganas and E. Petriu, "Hand gesture recognition using haar-like features and a stochastic context-free grammar", *IEEE Transactions on Instrumentation and Measurement*, 2008, pp 1562-1571.

[5] J. Xu, Y. Wu and A. Katsaggelos, "Part-based initialization for hand tracking", *The 17th IEEE International Conference on Image Processing (ICIP)*, 2010, pp 3257-3260.

[6] S. Swarup Rautaray and A. Agrawal, "A real time hand tracking system for interactive applications", *International journal of computer Applications*, 2011.

[7] Dan Xu, Yen-Lun Chen, Xinyu Wu and Yangsheng Xu, "Integrated approach of skin color detection and depth information for hand and face localization", *IEEE International Conference on Robotics and Biomimetics - ROBIO*, 2011, pp 952 - 956.

[8] S.E. Ghobadi, O.E. Leopprich, K. Hartmann and O. Loffeld, "Hand segmentation using 2d/3d images", *Proceeding of image and Vision Computing* , Hamilton, New Zealand, 2007, pp 64-69.

[9] A. Bleiweiss and M.l Werman, "Fusion time-of-flight depth and color for real time segmentation and tracking", *DAGM Symposium for Pattern Recognition, Jena*, Germany, 2009, pp 58-69.

[10] Lu Xia, Chia-Chih Chen and J.K. Aggarwal, "Human detection using depth information by kinect", *Computer Society Conference on Computer Vision and Pattern Recognition - CVPR,Colorado Springs*, 2011, 15 - 22.

[11] S. Qin, X. Zhu, Y. Yang, Y. Jiang, "Real-time Hand Gesture Recognition from Depth Images Using Convex Shape Decomposition Method",

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Journal of Signal Processing Systems, Vol. 74, 2014, pp 47-58.

[12] D. Xu, X. Wu, Y. Chen and Y. Xu, ”Online Dynamic Gesture Recognition for Human Robot Interaction”, *Journal of Intelligent Robotic Systems*, 2014.

[13] C. Matuszek, L. Bo, L. Zettlemoyer and D. Fox, ”Learning from Unscripted Deictic Gesture and Language for Human-Robot Interactions”, *I. J. Robotic*, 2014, pp 581-599.

[14] A. Boesen Lindbo Larsen, S. Hauberg, and K. Steenstrup Pedersen, ”Unscented kalman filtering for articulated human tracking*”, 17th Scandinavian Conference, SCIA*, Ystad, Sweden, 2011, pp 228-237.

[15] D. Koller, S. Thrun, C. PlagemannVarun and V. Ganapathi, ”Realtime identification and localization of body parts from depth images”, *IEEE International Conference on Robotics and Automation (ICRA)*, Anchorage, AK, 2010, pp 3108-3113.

[16] G. Saon and J. T. Chien, ”Bayesian sensing hidden markov models”, *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, 2012, pp 43-54.

[17] G. Saon and J. T. Chien, ”Viewing sea level by a one-dimensional random function with long memory”, *Mathematical Problems in Engineering*, vol. 2011, Article ID 654284, 2011.

[18] M. Elmezain, A. Al-Hamadi and B.Michaelis, ”Real-time capable system for handgesture recognition using hidden markov models in stereo color image sequences”, *Journal of WSCG*, vol. 16, 2008, pp 65-72.

[19] S. Eickeler, A. Kosmala and G. Rigoll, ”Hidden markov model based continuous online gesture recognition”, *Proceedings of 14th International Conference on Pattern Recognition*, vol. 2, 1998, pp. 1206-1208.

[20] N. D. Binh and T. Ejima, ”Real-time hand gesture recognition using pseudo 3-d Hidden Markov Model”, *Proceedings of the 5th IEEE International Conference on Cognitive Informatics (ICCI 06)*, 2002, pp 820-824.

[21] Y. Bengio and P. Frasconi, ”IEEE Transactions on Neural Networks”, *Input-output HMMs for sequence processing*, 1996, pp 1231-1249.

[22] S. Wang, A. Quattoni, L. Morency, D. Demirdjian, T. Darrell, ”Hidden conditional random fields for gesture recognition”, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2, 2006, pp 1521-1527.

[23] A. Corradini, ”Dynamic time warping for off-line recognition of a small gesture vocabulary”, *ICCV Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems, IEEE (2001)*, pp 82-89.

[24] Microsoft SDK: http://msdn.microsoft.com/enus/library/jj131025.aspx.

[25] R. R. Lawrence, “A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition”, *Proceeding of the IEEE*, Vol. 77(2), 1989, pp 257-286.

[26] G., Ye, D. Ha, O. Yongsheng, S. Weihua:” Human gesturere cognition through a kinect sensor”, *Robotics Biomimetics (ROBIO)* (2012).