TURKISH SIGN LANGUAGE RECOGNITION USING HIDDEN MARKOV MODEL

Kakajan Kakayev[1] and Ph.D. Songül Albayrak[2]

[1,2]Department of Computer Engineering,
Yildiz Technical University, Istanbul, Turkey
kkakajan@gmail.com
songul@ce.yildiz.edu.tr

ABSTRACT

In past years, there were a lot of researches made in order to provide more accurate and comfortable interaction between human and machine. Developing a system which recognizes human gestures, is an important study to improve interaction between human and machine.

Sign language is a way of communication for hearing-impaired people which enables them to communicate among themselves and with other people around them. Sign language consists of hand gestures and facial expressions. During the past 20 years, researches were made to facilitate communication of hearing-impaired people with others.

Sign language recognition systems are designed in various countries. This paper presents a sign language recognition system, which uses Kinect camera to obtain skeletal model. Our aim was to recognize expressions, which are used widely in Turkish Sign Language (TSL). For that purpose we have selected 15 words/expressions randomly (repeated 4 times each by 3 different signers) which belong to Turkish Sign Language. We have used 180 records in total. Videos are recorded using Microsoft Kinect Camera and Nui Capture. Joint angles and joint positions have been used as features of gesture and achieved close to 100% recognition rates.

KEYWORDS

Hidden Markov Model, Turkish Sign Language Recognition, Gesture Recognition, Microsoft Kinect, skeleton model

1. INTRODUCTION

In past years, new ways in human-computer interaction have been enabled. The researchers developed various applications on gesture and sign language recognition. Hearing-impaired people use sign language. When hearing-impaired people communicate, they need an interaction point and gestures are used as interaction point. Especially they use upper torso gestures and facial expressions. Handicap of sign language recognition is that signs change in 3D space. Sign language recognition developed using various input devices, such as wearing data gloves, stereo cameras, etc. Madabhushi and Aggarwal [1] developed sign recognition by tracking particular body parts. There are also studies based on depth information acquired from sensors such as Kinect [2, 3]. Ra'eesah Mangera used skeleton model, which is captured with Kinect camera [4]. Each country generally has its own native sign language and some have more than one. It is not clear how many sign languages there are. The 2013 edition of Ethnologue lists 137 sign

Natarajan Meghanathan et al. (Eds) : ACSIT, SIPM, CMIT, CoNeCo - 2016
pp. 11–18, 2016. © CS & IT-CSCP 2016 DOI : 10.5121/csit.2016.60802

languages [5]. There are researches made on Turkish [6, 7], Polish [8], American [9] Sign Language Recognition systems. In this study, we used 3D skeleton information of human skeleton model generated from Microsoft's Kinect sensor using Nui Capture.

In this paper, it is intended to recognize sign language by analysing skeleton model captured with Kinect camera. With the use of Kinect and Nui Capture application, the upper skeleton information of the human participants are recorded and used for training and testing the system. In order to recognize the signs we used K-Means with Hidden Markov Model (HMM). A k-means classifier is used to cluster the data. Every sign or gesture is shown with a series of frames. Features extracted from frames (joint angles, joint distances) and converted into observation sequence by means of k-means method and trained with Hidden Markov Model. Baum-Welch algorithm is used for HMM training [10].

For this work, 18 signs recorded by 3 persons. Each sign repeated 4 times by each signer. The content of recorded signs is shown in Table 1.

Table 1. Dataset used in training and testing phase.

No	Words
1	Let's meet again
2	See you
3	Good bye
4	Good night
5	Who?
6	Hello
7	Where?
8	Sometimes
9	Thanks
10	Yesterday
11	Cook
12	Doctor
13	Pharmacist
14	Baker
15	Driver
16	How?
17	I'm fine
18	Enjoy your meal!

The rest of this paper is ordered as follows: Section 2 describes calculations used for feature extraction from skeleton model and describes algorithm for training and testing procedure used in k-means and HMM. Section 3 provides experimental results obtained from records and Section 4 conclusion.

2. FEATURE EXTRACTION

Kinect Sensor generates depth maps, skeleton model, and RGB images. To interface with the device, NuiCapture and Kinect SDKs are used. NuiCapture is software used to record and analyse Kinect for Windows sensor data easily [11]. NuiCapture can export depth, color, and skeleton data to Matlab, Maya, 3DS Max, and MotionBuilder. The skeleton model extracted with nuiCapture is shown in Figure 1. Kinect camera tracks the 3-D coordinates of these joints. Sign languages consider hand gestures, upper body joints used in feature extraction.

2.1. Joint Distances

For each frame, the 3-D distance between each of the 6 arm joints and the head joint was calculated:

$$d(ls, he) = \sqrt{(x_{ls} - x_{he})^2 + (y_{ls} - y_{he})^2 + (z_{ls} - z_{he})^2} \quad (1)$$

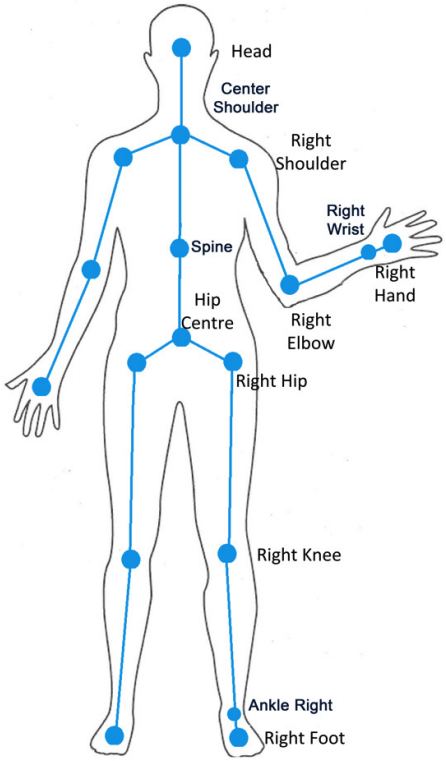

Figure 1. Kinect Skeleton Model [4]

Additionally, left hand – spine – right hand distance calculated. 7 dimensional feature vector formed from joint distances:

$$F_{JD} = [d_{ls}, d_{rs}, d_{le}, d_{re}, d_{lh}, d_{rh}, lh - rh]$$

User heights may be various. To reduce for the variation in user height, each distance was divided by the distance between spine and center shoulder [4].

2.2. Joint angles

The distance between joints is affected by the height of the user. Therefore, joint distances are not a scale invariant feature. Joint angles are not dependent on the user height or the distance from the camera. Joint angle also rotation invariant. Seven joint angles were calculated for each frame. Joint angles are shown in Figure 2. To calculate the joint angle, the vector between joints must be calculated. The calculation of shoulder-elbow-hand angle is illustrated in Figure 3. The shoulder – elbow – hand angle equation is given below:

$$\theta = \arccos\left(\frac{\overline{s-e}\cdot\overline{e-h}}{|\overline{s-e}||\overline{e-h}|}\right) \quad (2)$$

$\overline{(s-e)}$ and $\overline{(e-h)}$ is the shoulder – elbow and elbow – hand vector respectively. Numerator of equation 2 is the scalar product of the vectors and the denominator is the product of the magnitudes of the vectors. 7 dimensional feature vector is created from joint angles:

$$F_{JA} = [\gamma_L, \gamma_R, \beta_L, \beta_R, \alpha_L, \alpha_R, \partial]$$

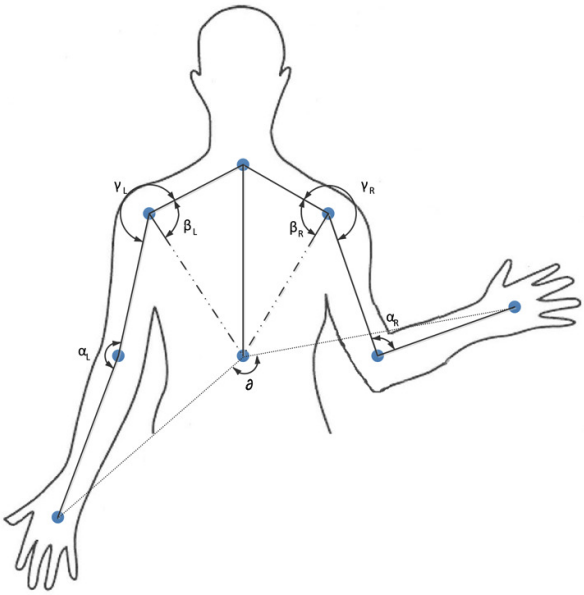

Figure 2. Joint angles calculated from skeleton model

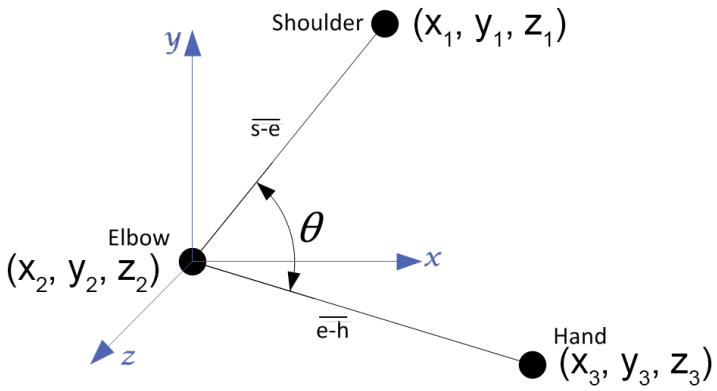

Figure 3. Calculation of shoulder – elbow – hand angle

2.3. Relative Joint Positions

Joint angles are rotation invariant, but a pose with the arms stretched on either side of the spine and arms stretched in front of the spine will have similar feature vectors. Therefore, the relative joint position between the elbow and hand joints and the head joint is calculated for each pose [4]. Figure 4 shows the position of the hand relative to the x-component of the head joint.

$\overline{he-h}$ is head-hand vector and the x-component of the head joint is:

$$he_x = x_1\hat{\imath} + 0\hat{\jmath} + 0\hat{k}$$

We can calculate the position of the hand relative to the head by equation 3

$$\varphi = arccos\left(\frac{\overline{he-h}\cdot\overline{he_x}}{|\overline{he-h}||\overline{he_x}|}\right) \quad (3)$$

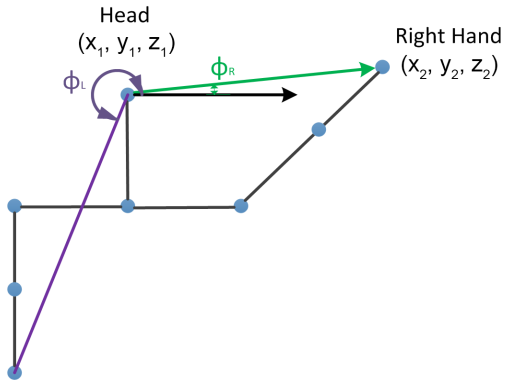

Figure 4. Demonstration of the relative position of the right and the left hands with respect to the head

2.3. Combination of feature vectors

By joining joint angles, joint distances and joint relative position we form 18 dimensional feature vectors for each frame.

$$F_C = [\gamma_L, \gamma_R, \beta_L, \beta_R, \alpha_L, \alpha_R, \sigma_L, \sigma_R, \partial, \varphi_L, \varphi_R, d_{ls}, d_{rs}, d_{le}, d_{re}, d_{lh}, d_{rh}, lh - rh]$$

Description of each feature vector provided in table 2.

Table 2. Description of feature vector elements

γ	Elbow – Shoulder – Neck angle	φ	Relative position of the elbow relative to the head
β	Spine – Shoulder – Neck angle	σ	Relative position of the hand relative to the head
α	Hand – Elbow – Shoulder angle	lh − rh	Distance between the left and right hands
∂	Left hand – Spine – Right hand angle	d	Distance between joints and head

2.4. Training and Testing

Feature extraction process is applied to all frames and signs. After feature extraction process, a k-means classifier is trained for each of the signs to obtain cluster centers. In this work, 40 used as a number of cluster (K) for each sign. This value was decided as empirically yielding the best inner-class division. Cluster centers are used for training system. HMM is trained by using Baum-Welch algorithm [10]. To calculate the recognition accuracy rate, the total number of correct recognitions is divided by total number of tests.

3. EXPERIMENTAL RESULTS

The system was tested with 2 different test types as shown in Figure 5. In Test 1, system was trained with all records of 2 signers and tested with all records of the 3rd signer. In Test 2, system was trained with 3 repetitions of each word and tested with the 4th repetition of each word and of each signer.

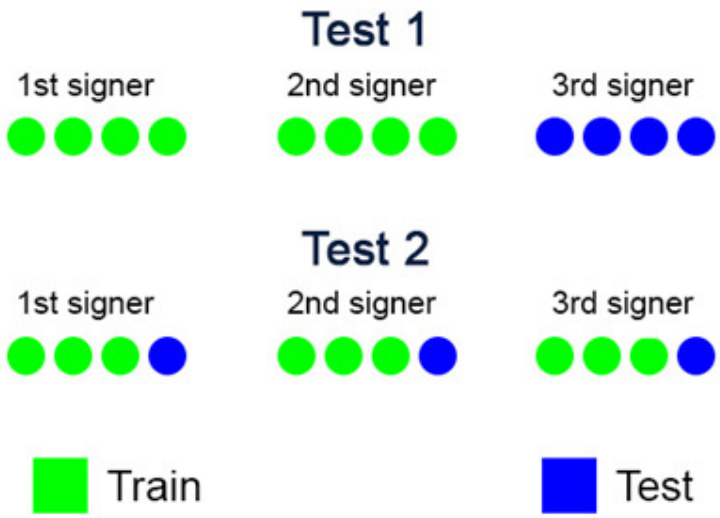

Figure 5. Test diagram

Dataset was divided into two parts, first part consists of 15 words, second part has 18 words. In the second dataset initial 15 words are the same as in the first dataset along with 3 extra words. Extra words are similar to previous words by means of movement. Test results are shown in Table 3, 4, 5 and 6.

Table 3. 15 words – Test 1

Total signs used for training	15
Total records used for training	120
Total records used for testing	60
Number of correct recognition	59
Number of wrong recognition	1
Recognition rate	59/60*100 = 98%

Table 4. 15 words – Test 2

Total signs used for training	15
Total records used for training	135
Total records used for testing	45
Number of correct recognition	43
Number of wrong recognition	2
Recognition rate	43/45*100 = 95%

Table 5. 18 words – Test 1

Total signs used for training	18
Total records used for training	144
Total records used for testing	72
Number of correct recognition	60
Number of wrong recognition	12
Recognition rate	60/72*100 = 83%

Table 6. 18 words – Test 2

Total signs used for training	18
Total records used for training	162
Total records used for testing	54
Number of correct recognition	50
Number of wrong recognition	4
Recognition rate	50/54*100 = 92%

4. CONCLUSION

In this paper we have presented sign language recognition system based on skeleton model of gestures. We have developed Turkish Sign Language recognition system using the Kinect camera and achieved close to 100% recognition rates. To increase the accuracy of the system, system can be trained by increasing the repetitions and signer number. The system works well for new words and new users.

It is observed that with Kinect's ability of recognition of human gestures adds another aspect to using computer applications. This research sets another example to sign language recognition systems. It will help hearing-impaired people as automatic translation tools.

This research has an advantage of freedom from any external devices for input except Kinect camera. In order to increase the recognition rate, the system needs as many repeated records and signs from different signers as possible. It is observed that when similar movements are used in test the successful recognition rate is reduced because of insufficient training data.

ACKNOWLEDGEMENTS

The author would like to acknowledge the colleagues, advisor for their assistance and everyone who supported at the time of developing this research.

REFERENCES

[1] A. Madabhushi and J. K. Aggarwal, "Using head movement to recognize activity", (2000) "Proceedings of 15th International Conference on Pattern Recognition", vol. 4, pp. 698 – 701.

[2] Yamato, J., Ohya, J. and ISHII, K., (1992). "Recognizing human action in time-sequential images using hidden Markov model", Computer Vision and Pattern Recognition, 379-385

[3] Biswas, K.K. and Basu, S.K., (2011). "Gesture Recognition using Microsoft Kinect", Robotics and Applications

[4] Mangera, R. (2013). "Static gesture recognition using features extracted from skeletal data"

[5] Wikipedia, Sign Language, https://en.wikipedia.org/wiki/Sign_language

[6] Haberdar, H., (2005). "Saklı Markov Model Kullanılarak Görüntüden Gerçek Zamanlı Türk İşaret Dili Tanıma Sistemi", Yıldız Technical University, İstanbul

[7] Memiş, A. and Albayrak, S., (2013). Turkish Sign Language Recognition Using Spatio-temporal Features on Kinect RGB Video Sequences and Depth Maps, Signal Processing and Communications Applications Conference, 1-4

[8] Oszust, M. and Wysocki, M., (2013). Polish Sign Language Words Recognition with Kinect, Human System Interaction, 219-226

[9] Z. Zafrulla, H. Brashear, H. Hamilton, T. Starner, and P. Presti, "American sign language recognition with the kinect," in Proceedings of the 13th international conference on multimodal interfaces, ser. ICMI'11, no. September, Sch. of Interactive Computing, Georgia Inst. Of Technology, Atlanta. New York, NY, USA: ACM, 2011, pp. 279–286.

[10] Wikipedia, Baum-Welch algorithm, http://en.wikipedia.org/wiki/Baum-Welch_algorithm

[11] Software to Record and analyse Kinect for Windows sensor data easily, http://nuicapture.com/