SIGN LANGUAGE TUTORING TOOL

Oya Aran, Cem Keskin, Lale Akarun

Department of Computer Engineering, Bogazici University, Bebek 34342 Istanbul
phone: + (90) 212 359 4523, fax: + (90) 212 287 2461,
email: aranoya@boun.edu.tr, keskinc@cmpe.boun.edu.tr, akarun@boun.edu.tr

ABSTRACT

In this paper, we present a Sign Language Tutoring Demonstrator, which is capable of teaching the basics of the sign language interactively. Instead of a passive learner, by incorporating a simple sign language recognizer to the system, the learner would be able to practice the signs and have feedbacks according to the similarity of the performed gesture to the actual gesture model.

1. INTRODUCTION

Sign language recognition is a multidisciplinary research area involving pattern recognition, computer vision, natural language processing and psychology. Sign language recognition is a comprehensive problem not only because of the complexity of the visual analysis of hand gestures and the highly structured nature of sign languages. Although sign languages are well-structured languages with a phonology, morphology, syntax and grammar, they are different from spoken languages: The structure of spoken language makes use of words linearly, i.e., one after the other, whereas sign language makes use of several body movements in parallel in a spatial and temporal space. The linguistic characteristics of sign language are different than that of spoken languages due to the existence of several components affecting the context such as the use of facial expressions and head movements in addition to the hand movements.

A very brief look into sign language grammar illustrates the challenges faced: Sign language phonology makes use of formational parameters such as the hand shape, place of articulation, and movement. The morphology uses directionality, aspect and numeral incorporation, and syntax uses spatial aspects such as localization and spatial agreement as well as facial expressions. It is clear that sign language recognition is a very complex task: a task that uses hand shape recognition, gesture recognition, face and body parts detection, facial expression recognition as basic building blocks.

Sign Language recognition requires both the hand trajectory and hand posture (position, orientation, angles of the articulations) information. In order to solve the hand trajectory recognition problem, Hidden Markov Models have been used extensively for the last decade. Lee and Kim [1] propose a method for online gesture spotting using HMMs. Starner et al. [2] used HMMs for continuous American Sign Language recognition. The vocabulary contains 40 signs and the sentence structure to be recognized was constrained to personal pronoun, verb, noun, and adjective. In 1997, Vogler and Metaxas [3] proposed a system for both isolated and continuous ASL recognition sentences with a 53-sign vocabulary. In a later study [4] the same authors attacked the scalability problem and proposed a method for the parallel modeling of the phonemes within an HMM framework.

In order to recognize even the simplest hand gesture, the hand must be detected in the image. Detection of the hand in natural environments by using only the skin color information is a challenging and complex task. To make the detection problem easier, markers on the hand and fingers are widely used in the literature. The environment can be restricted to be able to detect the hand without the need for markers.

Once the hand is detected, a complete hand gesture recognition system must be able to extract the hand shape, the hand motion and the relative position of the hand with respect to the other body parts such as head, shoulders, etc. The fusion of these different parameters in the recognition phase can be at the decision level as well as at the feature level. The recognizer must be able to differentiate between gestures (which are defined by the gesture classes in the training set) and non gestures (either unintentional movements of the hand or gestures that are not in the training set). This is especially essential for continuous recognition of hand gestures. A general framework of the system is given in Figure 1.

As pointed above, sign language recognition is a complex task. In this work, we developed a simple system which only recognizes dynamic hand signs.

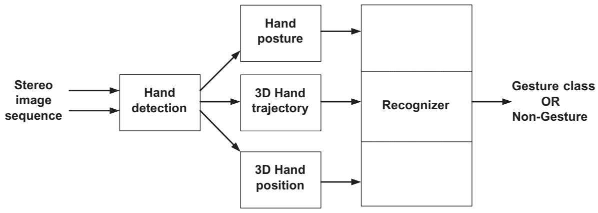


Figure 1: Hand Gesture Recognition System

2. SIGN LANGUAGE TUTORING

Sign language recognition requires the combination of different modalities: hand shape, hand motion, hand position, facial expressions. Moreover, the same gesture may have different meanings according to the context.

A Sign Language Tutor must be able to perform all steps in Figure 1. The hand must be detected robustly, and the hand shape, motion and position must be extracted carefully. However, a simple recognizer can be enough for the needs of the tutoring system since the aim is to compare the performed sign with a known sign. The actual type/class of the performed sign is unimportant if the performed sign is not similar to the known sign. There is also no need for the determination of non gestures.

During the learning process of sign language, one of the most important issues is to validate if the learned sign is correct or not. Instead of asking someone who knows sign language, this validation can be done by a computer program. Such interactive sign language teaching application can be used by deaf and mute people as well as by users with no disability. The availability of this application will increase the number of people who speak sign language and consequently increase the amount of communication among deaf and mute people and people with no disability.

The current version of the demonstrator has some requirements and restrictions:

- The signs currently in the application library are 7 signs form Turkish Sign Language
- These 7 signs are dynamic signs which include hand motion and two or three different hand shapes.
- We focused on recognizing the hand motion
- Hand shape recognition is done with a rule based system that checks very simple features for the hand shape
- The hand position with respect to the body parts (head, shoulders, etc.) is determined manually. The user is asked to wave hand in front of the face to aquire the position of the face in the image. Here, we assume that the body is realtively stationary and the hand is the dominant moving part in the video sequence.
- Facial expression recognition is beyond the scope of this paper.
- Only isolated signs are considered.
- The system works fast but there is no hard real time constraint since the user can wait a reasonable time for the feedback. Constraints for the waiting time can be added to the later versions of the system.
- The user must wear a uniformly colored glove. The color of the glove must be differentiable from the objects in the environment.
- The application is designed to be used at home or in office with normal lighting. Consistency problems may occur if the environment is dark.

There are two main phases of the application: Learning phase and practice phase. In the learning phase, the user first selects a sign. A pre-recorded video of that sign is shown when the user presses play button. The user can repeatedly watch this video until she/he is ready to practice. In the practice phase, the user is asked to repeat the same gesture which is selected in the training phase. The maximum duration for recording is determined with respect to the duration of the training video. At the end of the recording, the recorded avi file is processed and feedback is given to the user. The user can practice the same sign repeatedly. The training phase can be repeated for each of the signs in the application library.

3. SIGN LANGUAGE RECOGNITION SYSTEM

Figure 4 shows the recognition system used in the tutoring tool. This system is an extended version of a previous work [5], a gesture recognition system for interactive interfaces.

3.1 Marker Detection

In the developed system, we used colored gloves in order to detect the hand easily in all kinds of environments. The user can select any uniform color for the marker, as long as it is different than the colors in the environment. The initialization of the marker is done by moving the marker (i.e. waving the hand) while the marker detection utility is in progress. The hand must be the only moving object during the marker initialization. Using this assumption, the moving pixels in consequtive frames are detected. These are the pixels corresponding to the marker in the image. The corresponding average hue component is found from these pixels and this hue value is used in marker segmentation.

3.2 Marker Segmentation

Connected components algorithm with double thresholding is used to find the marker region in images acquired from the video sequence. The area of each connected region is calculated and regions smaller than a threshold are regarded as noise and are ignored. Fingertip detection is handled by extracting the simple shape descriptors such as the bounding box and the four outmost points of the hand defining the box. The elongation of the bounding box is used to determine the mode of the hand and the points are used to predict the location of the fingertip for different modes of the hand. In Figure 2 the detected marker is shown.

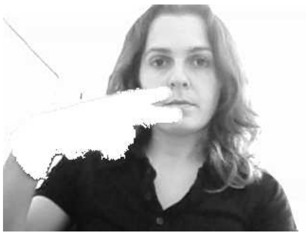


Figure 2: Detected Marker

3.3 Kalman Filtering

Kalman Filter is used to find the hand trajectory. Kalman Filter assumes that the object is linearly moving at constant velocity, subject to random perturbations in its trajectory. Since sampling is done in short time intervals, the assumption is convenient and the results of Kalman Filter on this model is satisfactory results. Figure 3 illustrates the measured and filtered trajectory of the hand motion.

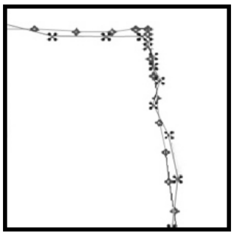


Figure 3: Effect of Kalman Filter

3.4 3D Reconstruction

For 3D tracking and recognition, two cameras are needed. Although the system supports both single camera setup and two camera setup, in sign language 3D information is essential for accurate recognition of the performed signs. The world coordinates of the fingertip is generated by 3D reconstruction from stereo using a least squares approach. To smooth the trajectory, a 3D Kalman Filter is applied on the reconstructed coordinates.

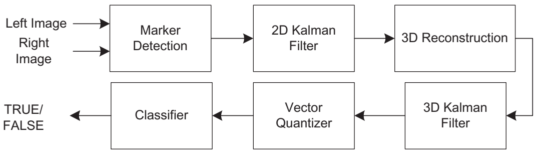


Figure 4: Sign Language Recognition System

To eliminate coordinate system dependence we transform the output of the Kalman Filter, the smoothed 3D coordinates of the marker in successive images, into sequences of quantized velocity vectors.

3.5 Classifiaction

A Hidden Markov Model is trained for each sign where the observations are the directional codewords characterizing the trajectory of the motion.

At the testing mode, the trajectory of the hand is given to each of the HMMs and the one that has the maximum likelihood is selected. If the sign associated with that selected HMM is equal to the training sign then the performed sign is TRUE, otherwise it is FALSE.

4. SIGN LANGUAGE TUTOR DEMONSTRATOR

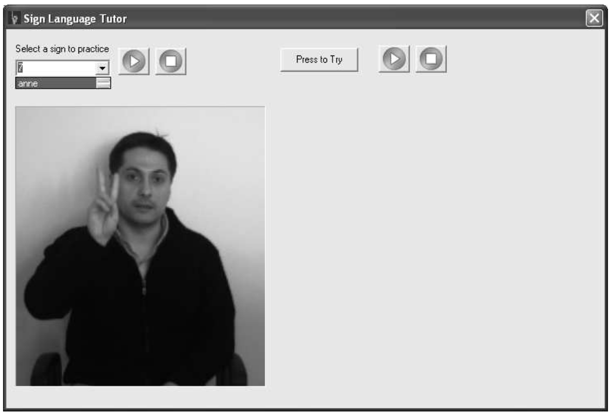
This demonstrator is designed to run on Windows XP operating system and implemented using Microsoft Visual Studio C++ 6.0. A Pentium 3 1.6 GHz PC is enough for real time operation. 2 web cameras that are capable of sending uncompressed RGB bitmap images must be installed. The bandwidths of the USB ports of two web cameras must be enough for the transmission of RGB bitmap images.

The user interface is a dialog based MFC application and designed such that the user can select among different signs. The interface form is divided into two sections which reflects the two phases of the system: learning phase and practice phase. In one of the sections, a video of the selected sign, which is performed by a natural signer, is shown to the learner (Figure 5). In the practice phase, the hand gesture performed by the signer is recorded as an Audio/Video Integrated (avi) file. After the learner performs the sign in the video, the recorded avi file is shown in the other section of the interface form (Figure 6). This avi file is processed by the application and a decision is made whether the performed gesture is correct or not. According to the decision, a feedback is given to the user as TRUE or FALSE. The feedback is given both visually (TRUE/FALSE indicator on the interface) and in audio (a pre-recorded sound that says TRUE or FALSE), to increase the usability of the system for people with different disabilities.

In the current version of the demonstrator, we used 7 different dynamic signs from Turkish Sign Language. Table 1 shows the properties of each sign.

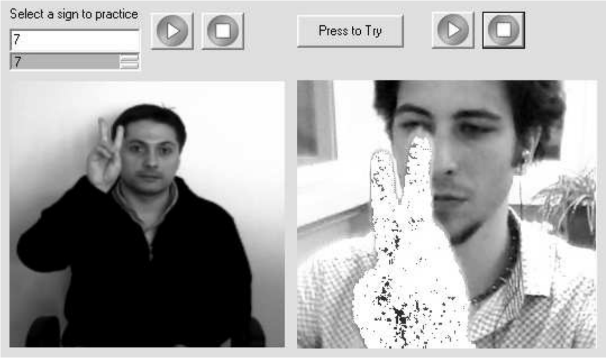
5. EXPERIMENTS

The dataset that we use in the experiments is collected from 7 people. Each person performed each of the 7 gestures at least 15 times. Data from 4 people is used as training set and the remaining data is used as the test set. Table 2 shows the total number of sequences, number of sequences in the training and test set for each sign.



Select a sign to practice
7
anne

Figure 5: Sign Language Tutor Interface - Learning Phase



Select a sign to practice
7
7

Figure 6: Sign Language Tutor Interface - Practice Phase

We tried HMMs with 5 states. The results are in Tables 3. Since each sign has different characteristics, the number of states of the HMMs must be determined seperately for each sign. We applied a scheme for the automatic determination of the number of states. The idea is to start with a fixed number of states and eliminate the states where the transition probability of moving to another state is above a threshold. When we apply this scheme, the number of states is determined as either 4 or 5 for each sign.

Sign	Total # of seq.	# seq. in train set	# seq. in test set
yedi	306	184	122
anne	211	127	84
cuma	137	82	55
gormek	101	61	40
kolay	124	74	50
konusmak	96	58	38
yok	102	61	41

Table 2: Datasets

6. CONCLUSION

This sign language teaching application is useful for learning the sign language both by deaf and mute people and by peo-

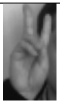
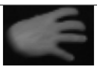
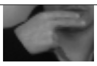
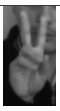
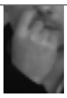
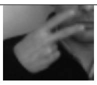
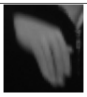
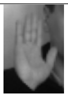
Turkish Meaning	English Meaning	Hand Shape	Hand Motion	Hand Position
7 (yedi)	7 (seven)		Vertical motion	On the side (left/right) of the head
anne	mother		Horizontal motion	On the chest
cuma	friday		Vertical motion	In front of the face
gormek	to see		Motion in the Z axis	In front of the eyes
kolay	easy		Horizontal motion	In front of the mouth
konusmak	to talk		Circular motion in the Z axis	In front of the mouth
yok	absent	 	Half circular motion in the Z axis	On the shoulders

Table 1: 7 Signs from Turkish Sign Language

Sign	Train Set	Test Set
yedi	0.8370	0.8770
anne	0.7795	0.8690
cuma	1.0000	1.0000
gormek	1.0000	1.0000
kolay	0.9865	1.0000
konusmak	0.9138	1.0000
yok	0.8197	0.9024

Table 3: Classification Accuracy for N = 5

	yedi	anne	cuma	gormek	kolay	konusmak	yok
yedi	107	0	8	0	6	1	0
anne	3	73	0	0	2	2	4
cuma	0	0	55	0	0	0	0
gormek	0	0	0	40	0	0	0
kolay	0	0	0	0	50	0	0
konusmak	0	0	0	0	0	38	0
yok	0	0	1	0	2	1	37

Table 4: Confusion Matrix for N = 5

ple with no disability. The learning process is based on two phases: showing the videos of the selected signs which are performed by native signers to the user and the user practices the sign. The output of the application is a feedback given to the user about the similarity of the performed sign to the performance of the native signer. Currently, this application is designed for a small set of dynamic signs and the user has to wear a colored glove. Even with a colored glove, if the illumination is not appropriate, it is not possible to detect the hand. The demonstrator must be improved to perform a robust detection of the hand without any gloves. The set of signs that is taught by the system can be improved by using a larger training set. The learning scheme can be improved to teach more signs in a more structured way: start with simple signs and advance as the user learns the signs.

REFERENCES

[1] Hyeon-Kyu Lee and Jin-Hyung Kim. Gesture spotting from continuous hand motion. *Pattern Recognition Letters*, 19(5-6):513–520, 1998.

[2] T. Starner and A. Pentland. Realtime american sign language recognition from video using hidden markov models. Technical report, MIT Media Laboratory, 1996.

[3] C. Vogler and D. Metaxas. Adapting hidden markov models for asl recognition by using three-dimensional computer vision methods. In *Conference on Systems, Man and Cybernetics (SMC'97), Orlando, FL*, pages 156–161, 1997.

[4] C. Vogler and D. Metaxas. Asl recognition based on a coupling between hmms and 3d motion analysis. In *International Conference on Computer Vision (ICCV'98), Mumbai, India*, 1998.

[5] C. Keskin, A. Erkan, and L. Akarun. Real time hand tracking and 3d gesture recognition for interactive interfaces using hmm. In *Proceedings of International Conference on Artificial Neural Networks, Istanbul, Turkey*, 2002.