Using Multiple Sensors for Mobile Sign Language Recognition

Helene Brashear & Thad Starner
College of Computing, GVU Center
Georgia Institute of Technology
Atlanta, Georgia 30332-0280 USA
{brashear, thad}@cc.gatech.edu

Paul Lukowicz & Holger Junker
ETH - Swiss Federal Institute of Technology
Wearable Computing Laboratory 8092
Zurich, Switzerland
{lukowicz, junker }@ife.ee.ethz.ch

Abstract

We build upon a constrained, lab-based Sign Language recognition system with the goal of making it a mobile assistive technology. We examine using multiple sensors for disambiguation of noisy data to improve recognition accuracy. Our experiment compares the results of training a small gesture vocabulary using noisy vision data, accelerometer data and both data sets combined.

1. Introduction

Twenty–eight million Deaf and hard–of–hearing individuals form the largest disabled group in the United States. Everyday communication with the hearing population poses a major challenge to those with hearing loss. Most hearing people do not know sign language and know very little about Deafness in general. For example, most hearing people do not know how to communicate in spoken language with a Deaf or hard–of–hearing person who can speak and read lips (e.g. that they should turn their head or not to cover their mouth). Although many Deaf people lead successful and productive lives, overall, this communication barrier can have detrimental effects on many aspects of their lives. Not only can person–to–person communication barriers impede everyday life (e.g. at the bank, post office, or grocery store), but also essential information about health, employment, and legal matters is often inaccessible to them.

Common current options for alternative communication modes include cochlear implants, writing, and interpreters. Cochlear implants are not a viable option for all Deaf people. In fact, only 5.3% of the deaf population in America has a cochlear implant, and of those, 10.1% of these individuals no longer user their implant (complaints cited are similar to those of hearing aides) [3]. The ambiguity of handwriting and slowness of writing makes it a very frustrating mode of communication. Conversational rates (both spoken and signed) range from between 175 to 225 WPM, while handwriting rates range from 15 to 25 WPM [7]. In addition, English is often the Deaf person's second language, American Sign Language (ASL) being their first. Although many Deaf people achieve a high level of proficiency in English, not all Deaf people can communicate well through written language. Since the average Deaf adult reads at approximately a fourth grade level [8, 2], communication through written English can be too slow and often not preferred.

Interpreters are commonly used within the Deaf community, but can have high hourly costs and be awkward in situations where privacy is of high concern, such as at a doctor or lawyer's office. Interpreters for Deaf people with specialized vocabularies, such as a PhD in Mechanical Engineering, can be difficult to find and very expensive. It can also be difficult to find an interpreter in unforeseen emergencies where timely communication is extremely important, such as car accidents.

Our goal is to offer a sign recognition system as another choice of augmenting communication between Deaf and hard of hearing people and the hearing community. We seek to implement a self contained system that a Deaf user could use as a limited interpretor. This wearable system would capture and recognize the Deaf user's signing. The user could then cue the system to generate text or speech.

2 Related Work

2.1 Language Models

Contact Sign is a modified form of American Sign Language (ASL) that is often used by Deaf signers when they encounter non–native signers [12]. It is a simplified version of ASL, with less complex combinations of movement and a grammar that is more analogous to English. We chose to constrain the scope of the language problem to a variant of Contact Sign. By using Contact Sign, we reduce the complexity of the language set we are seeking to recognize,

1530-0811/03 $ 17.00 © 2003 IEEE

while maintaining a language set that is useful to the Deaf community.

We choose to further constrain the problem by leveraging the idea of "formulaic" language. Formulaic is language that is ritualized or prefabricated. It includes routines, idioms, set phrases, rhymes, prayers and proverbs[22]. The DARPA one–way translation systems used by peace–keeping troupes, maritime law enforcement and doctors uses this idea to employ questions designed for specific responses. The system provides translations of predetermined phrases used to provide information or elicit feedback. Informative phrases include sentences like "I am here to help you" and "The doctor will be here soon". Requests and questions include "Please raise your hand if you understand me", "Is anybody hurt?" and "Are you carrying a weapon?"[16].

Cox describes the TESSA system, a system that combines formulaic language with speech recognition and semantic phrase analysis to create a system for generating phrases in British Sign Language for Deaf customers at the post office [5]. A set of formulaic language phrases were compiled from observed interactions at the post office. These phrases were then translated into sign and recorded on video. The postal employee speaks to a system that performs speech recognition and uses semantic mapping to choose the most likely phrase. The clerk may say "Interest in the UK is tax free", and the system would cue the phrase "All interest is free of UK income tax" which would then reference the video of a signed translation for the Deaf customer to see. The language processor achieved a 2.8% error rate on phrase matching for the post office domain.

The use of formulaic language allows for a reduction vocabulary size and allows for better error handling. Cox showed a progressive decrease in error rates for the language processor, by allowing a user to select from larger N best lists: 1–best was 9.7%, 3–best was 3.8% and 5–best was 2.8% [5]. The application of the phrase selection options also resulted in a significant increase in user satisfaction with the system. The TESSA system was scheduled to go on trial in five British post offices in May 2002.

2.2 Hidden Markov Models for Gesture Recognition

HMMs are stochastic models that represent unknown processes as a series of observations. Gesture recognition researchers have found HMMs to be a useful tool for modeling actions over time [23, 17, 4]. In particular, gesture recognition researchers have had some success with using HMMs for sign language recognition [11, 21, 18]. For an in–depth introduction to HMMs, the interested reader is referred to the tutorial by Rabiner [15].

2.3 First Person Camera View

Much of the research on sign language recognition has been done using cameras pointed at a person, in a "third person" view[21, 6]. The person then signs to the camera in the same manner they would to another person. These systems require that the camera be at a predetermined position in relation to the signer. While these systems have worked well in the lab, their design inherently limits their mobility.

A camera mounted in the brim of a hat can be used to observe the user and is sufficient to capture most of the normal signing space. Though sign language is designed to be viewed looking directly at another person, the head mounted camera captures the signs very effectively. The view of the camera is similar to a person looking at their own signing, which means that signing captured by the camera can be clearly understood. These motivations led to the head mounted camera design seen in Figure 1. A benefit of this design is that the user can monitor the camera's view via the head–mounted display. This allows the user some idea of what the camera see and provides feedback about system input.

2.4 Accelerometers for Gesture Recognition

Data gloves have been used by researchers for sign language recognition research [20, 14, 10, 13, 9, 11]. These data gloves are usually neoprene gloves with a network of accelerometers that send detailed information about rotation and movement of the hand and fingers. While these gloves provide large amounts of information about hand shape and movement, they also have problems associated with daily wear. Neoprene can be uncomfortable for long term use and interferes with a user's tactile feedback. Current data glove technology is not intended for daily use; the gloves deteriorate quickly with extended use and output becomes increasingly noisy as they break down.

If data gloves are not appropriate for the task, then we look to find another way of using the accelerometer technology. The information about rotation and movement could be extremely helpful in providing information about hand movement. Accelerometers could be made easy to wear if they were wireless and mounted on the wrist in a bracelet or watch. This application would not provide the detailed information about hand shape that the data gloves provide, but would provide information that is complementary to a vision system or other sensors.

2.5 Previous System

In the past, we have demonstrated a HMM based, sign language recognition system limited to a forty word vocabulary and a controlled lighting environment. [19]. The user

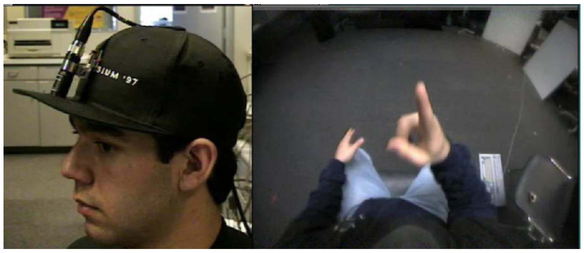

Figure 1. Hat–mounted camera and its view of signing

wore a hat–mounted camera (see Figure 1) to capture their signing. Data sets were taken in a controlled lab environment with standard lighting and background. The images were then processed on a desktop system and recognized in real–time. It was trained on a 40 word vocabulary consisting of samples of verbs, nouns adjectives and pronouns. The system trained to an accuracy of 97.8% on an independant test set using a rule–based grammar. The full statistics can be seen in Table 1.

The initial system was composed of a head–mounted NTSC camera which fed video to either to a computer or a recorder. The video was analyzed on a Silicon Graphics 200MHZ Indy workstation at 10 frames per second. Recognition was done in real time.

experiment	*training set*	*independant test set*
grammar	99.3%	97.8%
no grammar	93.1%	91.2%

Table 1. Word Accuracy of Original System

The HMMs for the system were trained and tested using the Hidden Markov Model Toolkit from Cambridge University. The Hidden Markov Model Toolkit (HTK) was developed at the Speech, Vision and Robotics group at the Cambridge University Engineering Department as a toolkit for building and using Hidden Markov models for speech recognition research [1, 24]. HTK consists of a library of tools primarily used for speech recognition, though it has gained popularity for non–speech applications. HTK was particularly useful because it allows us to augment HMM recognition with many speech tools that take advantage of the linguistic structure of sign language.

3 Current system

While the the previous research in sign language recognition shows some success in the lab, our goal is to create a system that is a viable assistive technology. The long term goal of the project is a system that a Deaf user could wear for the recognition of their signing. The user would sign and the system would offer the recognized words for the users acceptance or modification. The system could then be used to output text or speech, depending on the application. This recognition system could ultimately be extended to act as an input device for phones (for TTY or SMS), computing devices, or as a component in a sign to English translation system.

3.1 Motivation

Mobile systems are challenged by changing and unpredictable conditions. Our previous system was a lab–based computer vision system for sign language recognition. There are many challenges involved in taking systems that are developed for constrained lab scenarios and making them mobile. Mobile computer vision can often be extremely noisy, making it difficult to process. Lighting and background environment are constantly changing as the user moves through the world. Because of the camera's view, our background is the floor or ground surrounding the user and could have considerable clutter. Many computer vision techniques are developed and tested in the lab and never exposed to the stresses of a mobile environment. Lab systems use assumptions of lighting, color constancy, and uncluttered background to help ease the vision task. Even tasks such as tracking visual markers (such as our colored wristbands) become more difficult as lighting intensity, color and angle changes.

We propose using multiple sensor types for disambiguation of noise in gesture recognition. In this case, we chose to add accelerometers with three degrees of freedom, mounted on the wrists and torso to increase our sensing information. The accelerometers will capture information that the vision system will have difficulty with such as rotation (when hand shape looks similar) and vertical movement towards or away from the camera. The camera will provide information not gathered by the accelerometers such as hand shape and position. Both sensors collect information about the movement of the hands through space. It is our goal that by adding multiple sensor types, the accuracy of the system will be improved in noisy or problematic conditions.

It is important to add that sensor selection is based on the amount of information the sensor collects and its "wearability". We could cover the user in sensors for maximum information, but if it's not practical for daily wear, our system becomes less usable. We have been working with the Deaf community to ascertain what hardware is acceptable for wearable use. The current system could be partially concealed by embedding the camera in a normal hat, such as a baseball cap, and combining visual markers and accelerometers into a watch or bracelet. Since the system is trained specifically by the user, these choices can be somewhat customized.

The current research system is still somewhat cumbersome, but reaction from our community consultants has been positive. Early on we discovered that if the heads–up display was on the eye with the dominant hand, it was in the way of head based signs and was often knocked around. Switching the display to the other eye seemed to almost eliminate the problem. The first person view of the camera seems very useful for the users to observe how the system captures their signing. Overall, people have been excited about the technology and enjoyed playing with the wearable system, but more development is needed for daily use.

3.2 Design

Our current research system consists of a wearable computer, heads–up display, hat–mounted camera, and accelerometers. The system captures video of the user signing along with accelerometer data from the wrists and body. A sample view from the cameras can be seen in Figure 2. The left hand is marked by a cyan band on the wrist and the right hand is marked by a yellow band.

The current system is run on a CharmitPro with a Transmeta Crusoe 800MHz processor. The camera is an off the shelf CCD web cam connected via USB. The accelerometer system was designed at ETH (Swiss Federal Institute of Technology in Zurich) as part of an ongoing research collaboration. The current system captures and processes at 10 frames per second.

3.3 Accelerometer Network

The accelerometer network is a 3-wire bus with a dedicated master. Two wires implement the communication between the nodes using the I2C-bus and the third is used to synchronize all sensors. This hierarchical approach provides a logical separation of the sensor information increasing the amount of local processing and reducing the computational load on the central master. This allows for multiple, synchronized subnetworks.

Each of the sensor nodes is partitioned and consists of two parts each: a sub-board with two dual-axis accelerometers from Analog Devices ADXL202E which allow measurement of linear acceleration in the 3D-space, and a main board with the MSP430F149 low power 16-Bit mixed signal microprocessor (MPU) from Texas Instruments running at 6MHz maximum clock speed. The MPU reads out the analog sensor signals and handles the communication between the modules through dedicated I/O pins. Since our setup relies on the analog outputs of the accelerometers three second order Sallen-Key low pass filters are also used

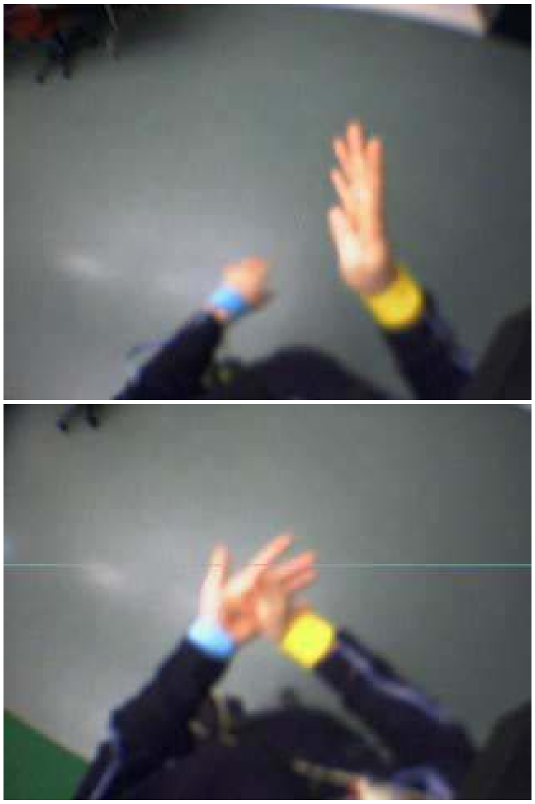

Figure 2. Sample views of signing from head–mounted camera

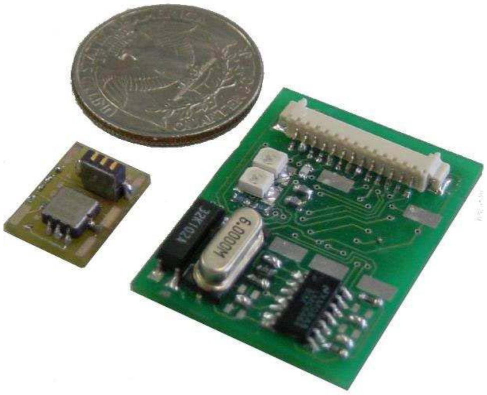

Figure 3. Image of the accelerometer components compared to a coin.

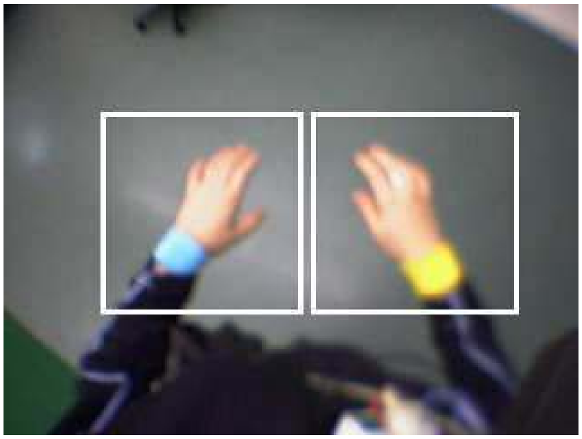

Figure 4. Example image of the user's view for the calibration phase

(fcutoff=50Hz) and located on the main-board. All modules are powered from a single central power supply consisting of a step down regulator and a small battery.

The sensor net we used in our experiments had three nodes. The master board box was affixed to the shoulder strap of the wearable bag. Two of the accelerometers were affixed to the top of the wrist. The third accelerometer was placed on the shoulder strap on the chest. The grey cabling for the accelerometers can be seen on the user's arms in Figure 2. The accelerometers themselves are covered by the colored wristbands.

3.4 Georgia Tech Gesture Toolkit

The HTK component of the system has been redesigned using the Georgia Tech Gesture Toolkit, which provides a publicly available toolkit for developing gesture–based recognition systems. The toolkit provides a framework that simplifies designing and implementing the gesture recognition component of larger systems by providing a bridge between the user and the services provided by HTK. Advantages of the redesign have been automated training and testing options and quick configuration as we test new ideas.

4 Method and Results

We designed the experiment to collect data about both sensing techniques to compare their performance. We captured a series of signed sentences to be used for testing and training. The capture method included both the vision and accelerometer data. The tests were then run using different subsets of information (accelerometer, vision, and combined), and the results were compared.

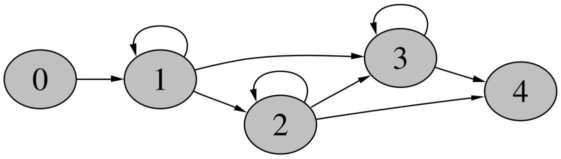

Figure 5. Topology for a 5 state left to right HMM with self transitions and 1 skip state used to model the gestures "calibrate", "my", "me", "talk", and "exit"

The test data was 72 annotated sequences of signed sentences. The vocabulary was a 5 gesture set of words {my, computer, helps, me, talk} plus a calibration gesture at the beginning and an exit gesture at the end. The calibration gesture involved the user placing their hands inside a box shown on the video stream (see Figure 4). The exit gesture was the process of dropping the hands and using Twiddler keyboard input to stop the data gathering. A rule based grammar was employed that consisted of a calibration gesture, 5 vocabulary gestures and an exit gesture. The test data contained sentences with differing permutations of the 5 vocabulary words.

Of concern is that our current language set is smaller than the training set used in our previous system experiments. However, our purpose here is to explore how we may augment computer vision so as to make a viable mobile system. Thus, our main goal is to determine if accelerometers provide features that are complementary to the vision system. We will continue to add words to our vocabulary as we collect more data sets.

The features sets used for training consisted of accelerometer data and vision data. The accelerometer feature vector consists of: (x,y,z) values for accelerometers on the left wrist, right wrist and torso. The vision feature vector consists of the following blob characteristics: x,y center coordinates, mass, eccentricity, angle, major axis x,y coordinates, and minor axis x,y coordinates. The camera captures at 10 frames a second and each frame is synchronized with 8–12 accelerometer packets. The accelerometer values are an average of the packets that accompany each frame.

The Georgia Tech Gesture Toolkit was used for the training and testing of our language models. Gestures in the vocabulary were represented by two different HMM topologies. Short words {my, me, talk, exit, calibrate} were represented with a 5 state left to right HMM with self transitions and 1 skip state, shown in Figure 5. Longer words {computer, helps} were represented with a 10 state left to right HMM with self transition and 2 skip states.

We used the leave–one–out testing mode and collected statistics for the training and testing runs. Leave–one–out

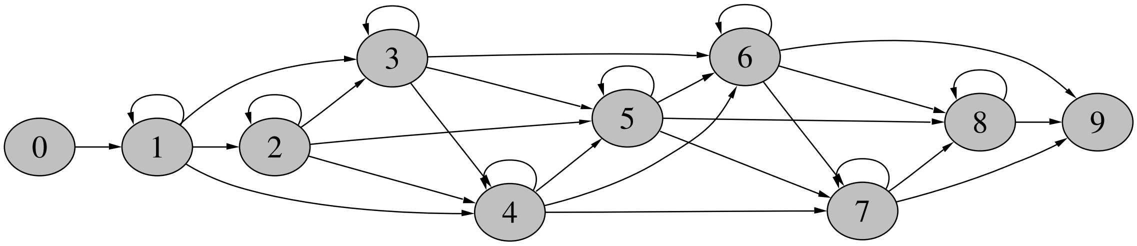

Figure 6. Topology for a 10 state left to right HMM with self transitions and 2 skip states used to model the gestures "computer" and "help"

	H	D	S	I	N
Vision	4.13	0.46	2.42	0.46	7
	292.75	30.14	174.11	30.14	497
Accelerometer	5.08	0.47	1.44	0.47	7
	405.97	30.25	60.78	30.25	497
Combined	6.40	0.7	0.53	0.7	7
	471.08	2.63	23.29	2.63	497

Table 2. Word level statistics: Rows where N=7 are testing on independant testing sets and N=497 are testing on training data. All results are averaged over the 72 runs.

testing involves running multiple training and testing cycles in which a single example is left out of the training set and used as a test set. The 72 examples allow for 72 leave–one–out training and testing runs. Statistics for each run are collected and then averaged together.

4.1 Word Level Metrics

We will use standard word level speech recognition metrics for measuring the performance of our recognition. These metrics are all evaluated based on the comparing the transcription of the recognized sentence with the actual sentence. Skipping a word results in a deletion error. Inserting extra words results in an insertion error. Substitution errors are words that were recognized incorrectly.

The following symbols are defined as:

- **H** is the number of correctly labeled gestures
- **D** is the number of deletion errors
- **S** is the number of substitution errors
- **I** is the number of insertion errors
- **N** is the total number of gestures labeled in the transcript

Correctness is calculated by: $Correct = \frac{H}{N}x100\%$

Accuracy is calculated by: $Accuracy = \frac{H-I}{N}x100\%$

Table 2 shows the word level results for the tests. Each data set (vision, accelerometer and combined) has two rows. The rows can be compared by the "N" column on the far left of the Table. The first row is "N=7", which shows the results of testing on the independant testing set (one sentence of 7 words). The second row is "N=497" which is the results of testing on the training set (71 sentences of 7 words for a total of 497 words). These word level statistics result in only whole numbers for each single run; Table 2 shows fractional numbers because they are the average over all 72 runs.

4.2 Recognition Rate Results

Table 3 shows the average sentence level accuracy over all of the runs. The "Training" column shows the results of testing on the training set and the "Testing" column shows the results of testing on a previously unseen (independant) testing set. There is marked improvement from vision (52.38% on Testing) and accelerometers (65.87% on Testing), and the combined set (90.48% on Testing). No training runs for any of the feature sets resulted in 0%. In 38 of the runs the combined vector recognized at 100% accuracy. In contrast, the accelerometer features recognized at 100% accuracy only 6 times, and the vision never did.

The "testing on training" results help show how well we can actually model the data. These runs show the results of testing the model with the data we used to create it. These recognition rates usually form an approximate upper bound for the models' performance. The difference between the accuracies on the training and testing set is less than 10% for all three sets, showing that we are approaching the upper bounds of recognition for the models we have trained. The hidden nature of HMMs makes it difficult to determine what data features are used as transitional signals, but the high correlation between the "Testing" and "Training" column statistics indicates a high probability that we are generalizing the models well and training on informative features

Data Set	Testing		Training	
	Mean	StdDev	Mean	StdDev
Vision	52.38%	8.97	52.84%	0.98
Accelerometer	65.87%	16.53	75.60%	1.30
Combined	90.48%	11.25	94.26%	0.87

Table 3. Accuracy: The "Testing" column shows testing on an independant testing set and "Training" column shows testing on training sets.

instead of noise .

4.3 Mobility and Generality

One criticism of these results is the dramatic difference in vision-only recognition rates as compared with the previous system. These questions reveal the change in direction that the project has taken. The initial system was an early proof of concept system for sign recognition, designed exclusively for in lab use. The current system is a proof of concept for mobility.

The early system had very narrow working parameters. The system was designed to work in a very specific environment that was engineered to aid in the vision task. The background and user's clothing was uniform and chosen for high contrast. The user sat stationary in a specific place and did not move his head or body. Special care was taken to place lighting around the signing space so that the hands were evenly lit. Because these parameters were carefully orchestrated, the vision hardware and software could be carefully calibrated for optimimum performance. The color thresholds, gain control, and white balance were adjusted by hand.

The hardware was also chosen for maximum performance instead of cost, wearability and form factor. The camera was a $3,000 near broadcast quality camera which could be hand calibrated for the working environment. The processing was done on a SGI O2 server with a professional digitization board. Though the system sensors were wearable, and the system had been designed for migration to a wearable system, it was not truly ”wearable” at the time.

In contrast, the current system has been designed for wearability and mobility. The hardware has been chosen for cost and form factor, which has resulted in a trade-off in quality. The camera is an off-the-shelf web cam, which is significantly cheaper, more durable, and consumes much lower power (1 Watt compared to 4 Watts on the previous system), but has much lower image quality. The vision code and camera calibration are much more generalized. The camera adjusts to lighting changes with auto gain control (instead of a pre-set value), which can often change color and shading effects. The color models have been chosen for a general environment, so they are more likely to be noisy.

5 Conclusion and Future Work

Our hypothesis that the two sensing methods would collect complementary and slightly overlapping information is validated by the results. Individually, vision and accelerometer data sets performed significantly less well than the combined feature vector, even when tested on the training set. We plan to further explore what kinds of techniques we can use for dealing with noisy sensing in the environment.

We are currently collaborating with Assistive Technology researchers and members of the Deaf community for continued design work. The gesture recognition technology is only one component of a larger system that we hope to one day be an active tool for the Deaf community. Research continues on the wearable form factor, pattern recognition techniques, and user interface.

We chose to use a rule–based grammar for sentence structure in the training and testing process. Speech recognition often uses statistical grammars for increased accuracy. These grammars are built by tying together phonemes (the simplest unit of speech) and training on the transition between the phonemes. The sets are usually done with bigrams (two phonemes tied together) or trigrams (three phonemes). Training using bigrams or trigrams requires considerably more data because representations of each transition of each word are now needed. In our case, the bigrams and trigrams would be built by tying together gestures. Our current data set is too small to effectively train using bigrams or trigrams, but we intend to continue collecting data with the goal of implementing these techniques.

The current system has only been trained on a very small vocabulary. We seek to increase the size and scope of our dataset to train on a larger vocabulary with more signing examples. A larger dataset will also allow us to experiment further on performance in different environments. Such a comparison will allow us to tangibly measure the robustness of the system in changing environments and provide training examples for a wider variety of situations.

We plan “Wizard of Oz” studies to seek to determine one–way translator phrase sets for constrained situations similar to the post office environment in Cox's work [5]. The phrase sets will help us determine appropriate vocabularies for various applications.

Work on the sensing components of the system will continue throughout development of the system. Adaptive color models and improved tracking could boost performance of the vision system. Wireless accelerometers will make wearing the system much more convenient. Other sensing options will be explored.

6 Acknowledgments

Funding, in part, from the Graphics, Visualization and Usability Center Seed Grant program, by NSF career grant # 0093291 and by the Rehabilitation Engineering Research Center on Mobile Wireless Technologies for Persons with Disabilities (Wireless RERC). The Wireless RERC is funded by the National Institute on Disability and Rehabilitation Research of the U.S. Department of Education under grant # H133W010804.

We would like to thank the Center for Assistive Technology and Environmental Access at Georgia Tech. Special thanks to Danielle Ross, Harley Hamilton and our other friends in the Deaf community who have helped greatly in the evolution of this project.

The opinions contained in this publication are those of the authors and do not necessarily reflect those of the U.S. Department of Education, the National Science Foundation and other funding institutions.

References

[1] HTK Hidden Markov Model Toolkit home page. http://htk.eng.cam.ac.uk/.

[2] Stanford achievement test, 9th edition, form s. *Norms Booklet for Deaf and Hard-of-Hearing Students*, 1996. Gallaudet Research Institute.

[3] Regions regional and national summary report of data from the 1999-2000 annual survery of deaf and hard of hearing children and youth. Technical report, Washington, D. C., January 2001. GRI, Gallaudet University.

[4] L. Campbell, D. Becker, A. Azarbayejani, A. Bobick, and A. Pentland. Invariant features for 3-d gesture recognition. In *Second Intl. Conf. on Face and Gesture Recogn.*, pages 157–162, 1996.

[5] S. J. Cox. Speech and language processing for a constrained speech translation system. In *Proc. Int.. Conf.. on Spoken Language Processing*, Denver, Co., September 2002.

[6] Y. Cui and J. Weng. Learning-based hand sign recognition. In *Proc. of the Intl. Workshop on Automatic Face- and Gesture-Recognition*, Zurich, 1995.

[7] J. Darragh and I. Witten. *The Reactive Keyboard*. Cambridge Series on Human-Computer Interaction. Cambridge University Press, Cambridge, 1992.

[8] J. Holt, C. B. Traxler, and T. E. Allen. Interpreting the scores: A user's guide to the 9th edition stanford achievement test for educators of deaf and hard-of-hearing students. *Gallaudet Research Institute Technical Report*, 91(1), 1997. Gallaudet Research Institute.

[9] W. Kadous. Recognition of Australian Sign Language using instrumented gloves. Master's thesis, University of New South Wales, October 1995.

[10] C. Lee and Y. Xu. Online, interactive learning of gestures for human/robot interfaces. In *IEEE Int. Conf. on Robotics and Automation*, volume 4, pages 2982–2987, Minneapolis, MN, 1996.

[11] R. Liang and M. Ouhyoung. A real-time continuous gesture interface for Taiwanese Sign Language. In *Submitted to UIST*, 1997.

[12] C. Lucas and C. Valli. *Sign language research: Theoretical issues*, chapter ASL, English, and contact signing, pages 288–307. Gallaudet University Press, Washington DC, 1990.

[13] L. Messing, R. Erenshteyn, R. Foulds, S. Galuska, and G. Stern. American Sign Language computer recognition: Its present and its promise. 1994. In *International Society for Augmentative and Alternative Communication: Conference Book and Proceedings*, pages 289–291, Maastricht, Netherlands, 1994.

[14] K. Murakami and H. Taguchi. Gesture recognition using recurrent neural networks. In *CHI '91 Conference Proceedings*, pages 237–241, 1991.

[15] L. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, Feb 1989.

[16] A. Sarich. DARPA one-way phrase translation system (PTS). http://www.sarich.com/translator/.

[17] J. Schlenzig, E. Hunter, and R. Jain. Recursive identification of gesture inputs using hidden Markov models. *Proc. Second Annual Conference on Applications of Computer Vision*, pages 187–194, December 1994.

[18] T. Starner and A. Pentland. Visual recognition of american sign language using hidden markov models. In *Proc. of the Intl. Workshop on Automatic Face- and Gesture-Recognition*, Zurich, 1995.

[19] T. Starner, J. Weaver, and A. Pentland. Real-time American Sign Language recognition using desk and wearable computer-based video. *IEEE Trans. Patt. Analy. and Mach. Intell.*, 20(12), December 1998.

[20] T. Takahashi and F. Kishino. Hand gesture coding based on experiments using a hand gesture interface device. *SIGCHI Bulletin*, 23(2):67–73, 1991.

[21] C. Vogler and D. Metaxas. Adapting Hidden Markov Models for ASL recognition by using three-dimensional computer vision methods. In *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, pages 156–161, Orlando, FL, October 1997.

[22] A. Wray and M. Perkins. The functions of formulaic language: an integrated model. *Language & Communication*, 1:1–28, 2000.

[23] J. Yamato, J. Ohya, and K. Ishii. Recognizing human action in time-sequential images using hidden Markov models. *Proc. Comp. Vis. and Pattern Rec.*, pages 379–385, 1992.

[24] S. Young. The HTK Hidden Markov Model Toolkit: Design and philosophy, 1993.