Progress in Automated Computer Recognition of Sign Language*

Barbara L. Loeding[1], Sudeep Sarkar[2], Ayush Parashar[2], and Arthur I. Karshmer[3]

[1] Department of Special Education, University of South Florida, Lakeland, Florida
Barbara@lklnd.usf.edu
[2] Computer Science and Engineering, University of South Florida, Tampa, Florida
sarkar@csee.usf.edu
[3] Department of Information Technology, University of South Florida, Lakeland, Florida

Abstract. This paper reviews the extensive state of the art in automated recognition of continuous signs, from different languages, based on the data sets used, features computed, technique used, and recognition rates achieved. We find that, in the past, most work has been done in finger-spelled words and isolated sign recognition, however recently, there has been significant progress in the recognition of signs embedded in short continuous sentences. We also find that researchers are starting to address the important problem of extracting and integrating non-manual information that is present in face and head movement. We present results from our own experiments integrating non-manual features.

1 Introduction

Human computer interaction is gradually moving towards a modality where speech recognition is playing a major role because of the advances in automated speech understanding and synthesis. This shift to speech-based I/O devices is likely to present a barrier in the near future for people with disabilities. In particular, it may be next to impossible for people who rely on sign language for communication to access state of the art technology unless these devices also incorporate automated recognition of sign language into speech and vice-versa. Research in speech-to-sign translation using a computer-generated signer has progressed to the level of being potentially useful, public, or commercial applications in various countries [1–3]. However, development of end-to-end products in automated recognition of sign language have proceeded at a slower pace. We could find information about only two kiosks [4, 5] that recognize a limited number of individual signs or sentences made by people wearing special gloves.

Many natural sign languages throughout the world incorporate a manual alphabet with 26-40 hand shapes or letters that allows people to spell out a word. This process is called finger spelling. To allow a deaf person to communicate proper names and concepts for which there is no readily known manual sign, a translation system needs to incorporate recognition of continuous finger spelling also. Since 1990, researchers have focused on automated recognition of isolated hand shapes (some are actual letters in manual alphabets) and individual signs. Representative examples of this work

* This material is based upon work supported by the National Science Foundation under Grant No. IIS 0312993.

K. Miesenberger et al. (Eds.): ICCHP 2004, LNCS 3118, pp. 1079–1087, 2004.

include [6–14]. Fortunately, in the last ten years, researchers around the world are gradually moving beyond isolated recognition to work on automated recognition of continuous letters, i.e., finger spelling [15–17]) and sentences of continuous signs, e.g., signs without intentional pauses between them [18–25]. Due to space limitations, we review only work in continuous signs in this paper. Table 1 compares data sets, features, technique(s) used by different research groups working on continuous sign language recognition and indicates the recognition rates reported in the literature. From this review, we make the following observations about the state of the art.

2 Input Data

Researchers are using a variety of input devices to collect sign data ranging from data gloves such as Cybergloves, to magnetic markers and 3SPACE position trackers, combinations of the above or vision-based systems that have no special input devices at all. Wearable devices bypass the segmentation problem by directly conveying location features but are unnatural and cumbersome. Number of cameras used also varies from one to three yielding 2D and 3D images. In some studies, the background has been uniform, signers wear dark clothes and/or colored gloves to make it easier for the computer algorithms to segment the hand and face regions.

3 Recognition Approach

Most groups up through the year 2000 used a Hidden Markov Models (HMMs) approach with one HMM per sign to automate the process of sign recognition [26, 19, 18, 27–32], with Taiwan also using a word segmentation approach and Japan using word segmentation and minCost approach. These HMM-based models have problems in scaling with the vocabulary. Many groups seem to have abandoned this model and adopted a parallel HMM model or a model where each HMM models a phoneme or subunit of a sign although different groups have different techniques for defining a subunit (movement-hold [21, 33], fenones [34]) An excellent example is the Chinese system [24] that uses a large hierarchical decision tree and state tying HMMs at three levels for 2400 basic phonemes of Chinese signs. They also have a model for distinguishing transitional movements that signers make between signs.

4 Databases

In terms of databases, typically, researchers have created their own video recordings of gestures, signs and sentences. The test data sets vary with one group focusing on continuous sentence recognition based on a sign vocabulary of 5 [35] to China's 5113-sign lexicon [24]. Three articles [19, 25, 35] report sentences composed from a sign vocabulary of 10 or less signs. Ten articles reporting a vocabulary of 22-49 [18, 20, 26, 36, 28, 37, 38, 33, 39, 22] and six report vocabulary of 52-250 [21, 27, 30–32, 40]. Most groups have attempted to recognize a set of 10 to 196 sign sentences of 2-12 signs in each sentence with recognition rates varying from 58% - 95.8%. A few groups have

tested between 400 and 1000 sentences [26, 36–38, 33, 41, 24] and report recognition rates of 52.8% - 90+%.

Large, common, datasets are rare. One such database is SignStream, an annotated video database tool for analysis of ASL linguistic data [42]. Linguists continue to study how the grammar of ASL differs from that of English. A recent boon for ASL researchers is availability of an extensive video database of ASL motions, handshapes, words and sentences created by Purdue University [43].

5 Recognition Performance

Most groups measure recognition rates for isolated signs and continuous signs, with some groups choosing to look at other variables such as context independent and context dependent; short vs. long sentences; with rule-based grammar and colored gloves vs. with no grammar and colored gloves and with grammar and skin tone compared to no grammar and skin tone. Reported recognition rates for continuous sign recognition vary from 58% [44] to 99% [36]. Although less useful, rates for recognizing isolated signs range much higher (91% -99%).

We advocate a need to build consensus regarding meaningful measures of performance from a communication point of view rather than reporting signal-based accuracies. Some researchers report total number of words (actually they mean signs) in all sentences that were correctly recognized and/or percentage of words (signs) within all signed sentences correctly recognized by their system. We propose that a more meaningful performance measure is the percentage of sentences that were correctly translated or recognized. Just word level recognition rates are not good indicators of communication performance. Some signs in a sentence are more important to recognize than others. So, instead of categorizing errors as insertion, deletion or substitution, perhaps judging to what degree an error affected the meaning of the sentence or assigning weights to errors would be preferable.

6 Non-manual Features

In addition to conveying meaning through manual signs, signers also convey information non-manually through their facial expressions and body movements. Facial expressions can include movements of the eyebrows, lips and head. Different groups have now started to tackle the issue of feature extraction related to non manual features such as head motion (Xu *et al.* [45], Erdem & Scarloff [46]), facial expression and lip movement (Canzler & Dziurzyk [47]). None of these groups have yet published combination strategies for manual and non-manual information and the impact of non-manual features on sign language recognition. At the University of South Florida, we have explored one possible strategy for extracting, representing, and integrating non-manual aspects such as facial expression and head movement into the recognition process [39].

To date, it has been difficult to directly use the facial information because of the following reasons. Manual information is not necessarily synchronized with non-manual information. For example, the same facial expressions are not present at the same temporal position with respect to the manual in two instances of the same sentence. Another

problem has to do with not being able to extract facial expression information in every frame of a sequence. If there is presence of a strong non-manual indicating 'Assertion' or 'Negation' in the sentence, the facial expressions, as registered in the images, are totally swamped by the face movements which are indicated by 'head shakes' or 'head nods'. For these two reasons, we employ a sequential integration approach where the non-manual information is used to prune the list of word hypotheses generated by manual information. We show that the additional use of non-manual information increases the accuracy of recognition of continuous words from 88% to 92%, which is a significant difference. We were also able to detect 'Negation' in sentences based on simple motion trajectory based features 27 out of 30 times.

7 Multiple Signers and Background

Other issues that are important has to do with being to able to handle multiple signers and multiple backgrounds. From what we can discern from published reports, a number of groups have utilized more than one signer but only a few groups are working on signer independent systems [24, 25]. There does not seem to be any group working on pure video based systems that can handle complex backgrounds.

8 Conclusion

To summarize, we have compared and contrasted different approaches to producing continuous sign language recognition in terms of the reliance on gloves vs bare hand, recognition method, database used for testing, performance achieved, use of non-manual information, and the ability to handle multiple signers. Recognition of continuous signs has progressed to a point where we are starting to see the emergence of the design for some future commercial systems [5, 49]. Among the research challenges that remain as excellent future research directions are: (i) signer independent recognition, (ii) development of bi-directional systems for each sign language, (iii) use of non-manual information to enhance recognition, (iv) development of communication based performance measures to quantify progress in sentence level recognition, and (v) development of non-glove based systems that scale up to a large vocabulary and handles different backgrounds. One of the long term inter-modality integrative efforts include the use of wide-area, high bandwidth networks and wireless communication technologies to provide "modality translation and assistance services" so that a variety of remote services such as instant text transcription or sign language interpretation on demand would be available anywhere anytime [50]. Second, the W3C now has published the first working draft of EMMA (Extended MultiModal Annotation markup language) which takes user-supplied information from speech recognizers, handwriting recognizers, keyboard and mouse device drivers and converts it into an EMMA representation [51]. It is advisable for researchers working on automated sign recognition and translation to communicate with W3C so that EMMA will also be able to utilize information supplied from "sign translators and sign recognizers".

Table 1. Automated Continuous Sign Recognition, i,e. Signs Embedded in Sentences

Citation	Type	Data Set, Accuracy Rate, Input Device, Technique Used
1995 Starner [26] and Starner & Pentland [36]	ASL	400 sentences based on 40 signs about 4 signs each 95% w colored gloves 90% w/o colored gloves w strict grammar 99%, camera-based, HMM
1996 Braffort [19]	French	7 signs, 44 sentences, 92-96%, HMMs, Dataglove
1996 Starner & Pentland [18]	ASL	40 signs, constrained sentence structure, no gloves-92%, gloves 99%, 1 camera, HMM
1997 Vogler & Metaxas [27]	ASL	53 signs 97 sentences widely varied sentence structure HMMS modeling 3 video cameras 3D movement parameters better than context dependent HMMs 92.1-95.8%
1997 Starner, Weaver & Pentland [28]	ASL	continuous, real time, sentence level, 40 sign lexicon, no gloves 92%, camera in cap 98%, (97% unrestricted grammar)
1998 Assan & Grobel [29]	Netherlands	26 signs, 14 sentences of 3-5 signs, 72.8%, signer dependent, HMMs
1998 Liang & Ouhyoung [21]	Taiwanese	71-250 signs, HMMs with explicit segmentation Movement-Hold, 1 Dataglove, signer dependent, real time, Long sentences 4-5 signs each recognition better than short sentences (2-3 signs ea), overall average rate is 80.4%, rate decreases as size of sign vocabulary increased
1998 Vogler & Metaxas [30]	ASL	53 signs, unconstrained sentence structure, HMM and 3D motion analysis
1999 Hienz, Bauer & Kraiss [31]	German	52 signs, signer dependent, sentences with 2-9 signs each, single camera, HMMs, Stochastic grammar language model, 95% bigram model
1999 Vogler & Metaxas [38, 37]	ASL	22 words, 499 sentences of diff lengths, 3 experiments, word accuracy 91.82%, phoneme level with local features 88.36%, Phoneme level with global features 91.19%, HMM recognized phonemes not whole signs
2000 Bauer, Hienz & Kraiss [32] and Bauer & Hienz [40]	German	Sentences based on 52 signs, 94%, sentences based on 97 signs, 93.2%, with bigram model, 1 HMM per sign
2000 Sagawa & Takeuchi [44, 5]	Japanese	10 sentences, 1 and 2 handed signs, 200 samples, 58%
2000 Vogler, Sun & Metaxas [33]	ASL	99 test and 400 training sentences over 22 signs, parallel HMMs each HMM based on single phoneme Movement Hold, 84.9% sentence level, 94.2% sign level

Table 1. (Continued)

Citation	Type	Data Set, Accuracy Rate, Input Device, Technique Used	Non-Manual
2001 Vogler & Metaxas [41]	ASL	400 training, 99 test sentences, 22 signs, phoneme level, Parallel HMMs	
2001 Xu [45]	Japanese	11 head motions, "good recognition", real time 2 video cameras	Head motions
2002 Erdem & Sclaroff [46]	ASL	10 ASL video sequences, 3D head tracker, no sign recognition results	Head Shakes
2002 Canzler and Dziurzyk [47]	German	Nonmanuals facial features, 24 persons, 30 images, no sign recognition results	Face
2002 Yuan *et al.* [48]	Chinese	40 signs, upto 4 sign sentences, 2 gloves and 3 trackers, 70%, HMMs (strong/weak connections)	
2003 Brashear *et al.* [35]	ASL	5 signs, 72 sentences, 90.48%, accelerometer	
2003 Chen [24]	Chinese	5113 signs, 1000 sentences, 6 signers, signer independent, 87.58%, decision tree, fuzzy VQ	
2003 Kapuscinski & Wysocki [25]	Polish	10 signs, 12 sentences, 2-4 sign each, for 1 signer recognition rate of 82.5-86.7%, 3 signers results in 81.3% - 86.4%, PaHMM + bigram = best, Camera Long sleeves, No gloves	
2003 Parashar [39]	ASL	39 signs 25 sentences, with 1-5 signs each, 2 cameras, bottom up approach using relational features, 88% sign recognition, 92% with nonmanual info	Face, Head nods
2004 Vogler & Metaxas [22]	ASL	22 signs 99 test sentences, 88.89%, movement channel, handshape channel, Parallel HMM	

References

1. Cox, S., Lincoln, M., Tryggvason, J., Nakisa, M., Wells, M., Tutt, M., Abbott, S.: TESSA, a system to aid communication with deaf people. In: Proceedings of the fifth international ACM conference on Assistive technologies, ACM Press (2002) 205–212
2. Phelps, K.: Signing avatar characters become virtual tutors. In: Virtual Voice. (2002)
3. Toro, J., *et al.*: An improved graphical environment for transcription and display of American Sign Language. Information **4** (2001) 533–539
4. Akyol, S., Canzler, U.: An information terminal using vision based sign language recognition. In: ITEA Workshop on Virtual Home Environments. (2002) 61–68
5. Sagawa, H., Takeuchi, M.: Development of an information kiosk with a sign language recognition system. In: Conference on Universal Usability. (149–150) 2000

6. Kramer, J., Leifer, L.: The talking glove: An expressive and receptive 'verbal' communication aid for the deaf, deaf-blind and nonvocal. In: Conference on Computer Technology, Special Education, and Rehabilitation. (1987)
7. Murakami, K., Taguchi, H.: Gesture recognition using recurrent neural networks. In: SIGCHI Conference Proceedings. (237–242) 1991
8. Charayaphan, C., Marble, A.: Image processing system for interpreting motion in American Sign Language. Journal of Biomedical Engineering **14** (1992) 419–425
9. Waldron, M., Kim, S.: Isolated ASL recognition system for deaf persons. IEEE Transactions on Rehabilitation Engineering **3** (1995) 261
10. Kadous, M.W.: Machine translation of AUSLAN signs using powergloves: Towards large lexicon-recognition of sign language. In: Workshop on the integration of Gesture in Language and Speech. (1996) 165–174
11. Vamplew, P.: Recognition of Sign Language Using Neural Networks. PhD thesis, Department of Computer Science, University of Tasmania (1996)
12. Lee, C., Xu, Y.: Online, interactive learning of gestures for human robot interfaces. In: IEEE International Conference on Robotics and Automation. (1996) 2982–2987
13. Al-Jarrah, O., Halawani, A.: Recognition of gestures in Arabic sign language using neuro-fuzzy systems. Artificial Intelligence **133** (2001) 117–138
14. Fang, G., Gao, W., Zhao, D.: Large sign vocabulary sign recognition based on hierarchical decision tree. In: International Conference on Multimodal Interfaces. (2003) 125–131
15. Messing, L., Erenshteyn, R., Foulds, R., Galuska, S., Stern, G.: American Sign Language computer recognition: Its present and its promise. In: ISAAC Conference. (1994)
16. Hernandez-Rebollar, J.L., Lindeman, R.W., Kyriakopoulos, N.: A multi-class pattern recognition system for practical finger spelling translation. In: The 4th IEEE International Conference on Multimodal Interfaces. (2002) 185
17. Vassilia, P.N., Konstantinos, M.G.: Towards an assistive tool for Greek Sign Language communication. In: IEEE International Conference on Advanced Learning Technologies (ICALT'03). (125) 2003
18. Starner, T., Pentland, A.: Computer-based visual recognition of American Sign Language. In: International Conference on Theoretical Issues in Sign Language Research. (1996)
19. Braffort, A.: ARGo: An architecture for sign language recognition and interpretation. In: Progress in Gestural Interaction. (1996) 17–30
20. Grobel, K., Assan, M.: Isolated sign language recognition using Hidden Markov Models. In: International Conference System: Man and Cybernetics. (1996) 162–167
21. Liang, R., M.Ouhyoung.: A real-time continuous gesture recognition system for sign language. In: International Conference on Automatic Face and Gesture Recognition. (1998) 558–565
22. Vogler, C., Metaxas, D.: Handshapes and movements: Multiple-channel ASL recognition. In: Lecture Notes in Artificial Intelligence 2915. (2004) 247–258
23. Ma, J., Gao, W., Wang, C., Wu, J.: A continuous Chinese Sign Language recognition system. In: International Conference on Automatic Face and Gesture Recognition. (2000) 428–433
24. Chen, Y.: Chinese Sign Language recognition and synthesis. In: IEEE International Workshop on Analysis and Modeling of Faces and Gestures. (2003)
25. Kapuscinski, T., Wysocki, M.: Vision-based recognition of Polish Sign Language. In: Methods in Artificial Intelligence. (2003)
26. Starner, T.: Visual recognition of American Sign Language using Hidden Markov Models. Master's thesis, MIT, Media Lab. (1995)
27. Vogler, C., Metaxas, D.: Adapting Hidden Markov Models for ASL recognition by using three-dimensional computer vision methods. In: International Conference Systems on Man and Cybernetics. (1997) 156–161

28. Starner, T., Weaver, J., Pentland, A.: A wearable computer based American Sign Language recognizer. In: International Symposium on Wearable Computers. (1997) 130–137
29. Assan, M., Grobel, K.: Video-based sign language recognition using Hidden Markov Models. In: International Gesture Workshop: Gesture and Sign Language in Human-Computer Interaction. (1998) 97–109
30. Vogler, C., Metaxas, D.: ASL recognition based on a coupling between HMMs and 3D motion analysis. In: International Conference on Computer Vision. (363–369) 1998
31. Hienz, K., Bauer, B., Kraiss, K.: HMM-based continuous sign language recognition using stochastic grammars. In: Gesture Workshop. (1999)
32. Bauer, B., Hienz, H., Kraiss, K.F.: Video-based continuous sign language recognition using statistical methods. In: International Conference on Pattern Recognition. (2000) 463–466
33. Vogler, C., Sun, H., Metaxas, D.: A framework for motion recognition with application to American Sign Language and gait recognition. In: Workshop on Human Motion. (2000) 33–38
34. Bauer, B., Kraiss, K.F.: Towards an automatic sign language recognition system using subunits. In: International Gesture Workshop: Gesture and Sign Language in Human-Computer Interaction. (2002) 64–75
35. Brashear, H., Starner, T., Lukowicz, P., Junker, H.: Using multiple sensors for mobile sign language recognition. In: IEEE International Symposium on Wearable Computers. (2003)
36. Starner, T., Pentland, A.P.: Real-time American Sign Language recognition from video using Hidden Markov Models. In: Symposium on Computer Vision. (1995) 265–270
37. Vogler, C., Metaxas, D.: Toward scalability in ASL recognition: Breaking down signs into phonemes. In: Gesture-Based Communication in Human-Computer Interaction. (211-224) 1999
38. Vogler, C., Metaxas, D.: Parallel Hidden Markov Models for American Sign Language recognition. In: International Conference on Computer Vision. (116–122) 1999
39. Parashar, A.: Representation and interpretation of manual and non-manual information for automated American Sign Language Recognition. Master's thesis, University of South Florida (2003)
40. Bauer, B., Hienz, H.: Relevant features for video-based continuous sign language recognition. In: International Conference on Automatic Face and Gesture Recognition. (2000) 440–445
41. Vogler, C., Metaxas, D.: A framework of recognizing the simultaneous aspects of American Sign Language. Computer Vision and Image Understanding **81** (2001) 358–384
42. Neidle, C., MacLaughlin, D., Bahan, B., G., L.R., Kegl, J.: The SignStream project. In: American Sign Language Linguistic Research Project, Report 5 Boston University. (1997)
43. Martinez, A.M., Wilbur, R.R., Shay, R., Kak, A.: Purdue RVL-SLLL ASL database for automatic recognition of American Sign Language. In: International Conference on Multimodal Interfaces. (2002)
44. Sagawa, H., Takeuchi, M.: A method for recognizing a sequence of sign language words represented in a Japanese Sign Language sentence. In: International Conference on Automatic Face and Gesture Recognition. (434–439) 2000
45. Xu, M.: A vision-based method for recognizing non-manual information in Japanese Sign Language. In: International Conference on Advances in multimodal interfaces. (2000) 572–581
46. Erdem, U., Sclaroff, S.: Automatic detection of relevant head gestures in American Sign Language communication. In: International Conference on Pattern Recognition. (2002) I: 460–463
47. Canzler, U., Dziurzyk, T.: Extraction of non manual features for videobased sign language recognition. In: Proceedings of IAPR Workshop on Machine Vision Applications. (2002) 318–321

48. Yuan, Q., Gao, W., Yao, H., Wang, C.: Recognition of strong and weak connection models in continuous sign language. In: International Conference on Pattern Recognition. (2002) 10075
49. Clendenin, M.: Chinese lab hopes to commercialize sign-language recognition platform. In: http://www.eetimes.com/article/showArticle.jhtml?articleId=10801270. (March 2003)
50. Zimmerman, G., Vanderheiden, G.: Modality translation and assistance services: A challenge for artificial intelligence. Journal of the Australian Society of Artificial Intelligence **20** (2001)
51. Larson, J.A.: EMMA: W3C's extended multimodal annotation markup language. Speech Technology Magazine **8** (2003) http://www.w3.org/TR/emma/