Handshapes and movements: Multiple-channel ASL recognition

Christian Vogler[1] and Dimitris Metaxas[2]
[1]Gallaudet University, Christian.Vogler@gallaudet.edu
[2]Rutgers University, dnm@cs.rutgers.edu

A revised version of this paper was published in the Springer Lecture Notes in Artificial Intelligence 2915, pp. 247–258, © Springer Verlag, 2004.

Abstract

In this paper we present a framework for recognizing American Sign Language (ASL). The main challenges in developing scalable recognition systems are to devise the basic building blocks from which to build up the signs, and to handle simultaneous events, such as signs where both the hand moves and the handshape changes. The latter challenge is particularly thorny, because a naive approach to handling them can quickly result in a combinatorial explosion.

We loosely follow the Movement-Hold model to devise a breakdown of the signs into their constituent phonemes, which provide the fundamental building blocks. We also show how to integrate the handshape into this breakdown, and discuss what handshape representation works best. To handle simultaneous events, we split up the signs into a number of channels that are independent from one another. We validate our framework in experiments with a 22-sign vocabulary and up to three channels.

1 Introduction

Sign languages are the primary mode of communication for many deaf people, just as speech is for hearing people. The rate of progress in the speech recognition field has been impressive in the past decade, with some systems now becoming borderline mainstream. At this moment, the predominant way to interact with a computer is still the keyboard and mouse, but as speech recognition becomes more established, the situation may change.

Such a change would pose major challenges to sign language users, unless we can also advance the field of sign language recognition. Sadly, the current state of the art in sign language recognition research is still lagging far behind speech recognition. One major reason for this disparity is that sign language recognition is much harder than speech recognition. The reason why it is harder can be expressed succinctly in just two words: *simultaneous events.* In speech, we can — on an abstract level — represent every word as a sequence of sounds. In contrast, a major element of sign languages is that several things happen at the same time. For instance, many signs use both hands, which move at the same time. The question is how to capture all the simultaneous events without having to consider every imaginable combination of them. If we had to consider every combination, the computational complexity of the task would be prohibitive, as we would quickly run into a combinatorial explosion.

To complicate matters further, sign languages are highly inflected; that is, the appearance of a sign can change according to the subject and the object of the sentence. In addition, there are other processes beside inflection that contribute to a large number of different appearances, such as changing the handshape depending on what type of object is under consideration [15]. The bottom line is that there are too many appearances to model them all explicitly, so it is necessary to cast for more basic building blocks, from which we can build all these appearances. Just like in spoken languages, phonemes take on the role of these basic building blocks. However, it is far from clear how exactly to break down a sign into its constituent phonemes, for two reasons: First, linguistic research into the phonology of signed languages is still in its infancy, overall. Second, it has not been clearly established yet what the computational requirements of recognition systems are, so a breakdown of signs into phonemes must somehow balance linguistic and engineering requirements.

In this paper we present a hidden Markov model (HMM)-based framework for American Sign Language (ASL) recognition that addresses these two challenges. We loosely follow the Movement-Hold phonological model [8] to devise a representation of signs suitable for a recognition system. In contrast to earlier work based on this model [17, 18] we now integrate the handshape, as well. We then extend this representation to multiple channels, which capture the simultaneous events in ASL, one in each channel. The key idea to making this approach work is that we assume that the channels are independent from one another. As a consequence, we can recognize the channels independently from one another, and put together combinations of simultaneous events on the fly, which greatly reduces the computational complexity of the recognition task. Although this assumption is unlikely to hold from a theoretical standpoint, the practical experimental results justify it as a reasonable engineering tradeoff.

The rest of this paper is organized as follows: We briefly discuss related work, and then continue with the mainstay of this paper: how to model ASL in a manner that makes it computationally tractable for recognition systems, including how to represent the handshape. Afterward, we briefly discuss HMMs and how they fit into the recognition framework, and conclude with experiments that validate our assumptions.

2 Related Work

This discussion of related work focuses on previous work in sign language recognition. For coverage of gesture recognition, the survey in [9] is an excellent starting point.

C. Wang, W. Gao, and J. Ma described a large-scale HMM-based isolated recognition system for Chinese Sign Language with a very impressive vocabulary size of more than 5000 signs [19]. They used some tricks from speech recognition, such as clustering Gaussian probabilities, and fast matching, to achieve real-time recognition and recognition rates of 95%.

Most work on continuous sign language recognition is based on HMMs. T. Starner and A. Pentland used a view-based approach to ASL recognition with a single camera to extract two-dimensional features as input to HMMs with a 40-word vocabulary and a strongly constrained sentence structure [13]. G. Fang and colleagues proposed an approach to signer-independent continuous recognition of Chinese Sign Language based on an integration of simple recurrent networks (SRNs) and HMMs [5]. They used the SRNs to segment the continuous sentences into individual signs. The recognition rates were 92% over a test data set of 100 sentences with a 208-sign vocabulary.

H. Hienz and colleagues used HMMs to recognize a corpus of German Sign Language [6]. They also experimented with stochastic bigram language models to improve recognition performance. The results of using stochastic grammars largely agreed with the results in [16]. B. Bauer and K.-F. Kraiss from the same group later extended the framework to break down the signs into smaller units. These units were unlike phonemes, however, because they were determined computationally via clustering, instead of being determined linguistically. For this breakdown they achieved an accuracy of 92.5% in isolated sign language recognition experiments [2, 3].

R. H. Liang and M. Ouhyoung used HMMs for continuous recognition of Taiwanese Sign Language with a vocabulary between 71 and 250 signs [7], which they extracted from a Cyberglove in conjunction with a magnetic 3D tracker. They worked with Stokoe's system [14] to detect the handshape, position, and orientation aspects of the running signs. They integrated the handshape, position, orientation, and movement aspects through stochastic parsing and a dynamic programming algorithm at a higher level than the HMMs. The main assumption of their work is that the sign can be represented as a series of postures.

The work described in this paper is an extension of the work done in [17, 18]. In this earlier work we proposed an approach to breaking down the signs into phonemes, and devised a framework for recognizing simultaneous aspects of ASL. However, this work was restricted to the movements of the hands and ignored handshape. In this paper we validate the framework further by integrating the handshape into the framework, and extend our phoneme-based modeling approach to ASL to cover the handshape, as well.

3 Modeling ASL

A large part of the difficulty in devising an ASL recognition system is the question of how to represent the language. To keep the complexity of the recognition task small and to ensure that the system is scalable, we need to keep the number of building blocks that constitute the signs small. Likewise, we need to ensure that we can capture the simultaneous events, which happen all the time in signed languages, such as two-handed signs, and handshape changes during hand movements. Doing so without getting bogged down in a combinatorial explosion of such events is nontrivial.

In the following we address how to break down signs into their constituent phonemes, with the idea that the number of phonemes in a language is small. Hence, we can use the phonemes as the basic building blocks for a recognition system. We

(a) MOTHER (b) FATHER

Figure 1: Contrast between signs that identifies location as a phoneme in ASL. (a) and (b) differ only in location.

discuss this breakdown for both the hand movements, and the handshape. In addition, we discuss how to capture simultaneous events in a computationally tractable manner.

3.1 Phoneme-Based Modeling

A **phoneme** is defined to be the smallest contrastive unit in a language [15]; that is, the smallest unit that can distinguish morphemes (units of meaning) from another. In English, the sounds */k/, /æ/,* and */t/* (and their equivalents in regional dialects) are examples of phonemes, as can easily be seen by comparing the words “cat” - “hat,” “bat” - “bet,” and “bet” - “bed.” In ASL, the equivalents of phonemes in spoken languages are the various handshapes, locations, orientations and movements.

As an example, consider the location of the hand at the chin in the sign for MOTHER. The arguments that this unit is an example of an ASL phoneme is analogous to the arguments for English: Figure 1 shows that the signs for MOTHER - FATHER differ only in location (front of the chin vs. front of the forehead).

Just like in spoken languages, the *number of phonemes in ASL is limited* and small compared to the number of signs. The exact number is still a matter of debate and depends greatly on the phonological model used. Stokoe’s system [14], for instance, identifies 55 units, whereas the Movement-Hold model identifies more than 100 [8].

In this work, we follow the basic ideas of the Movement-Hold model [8]. It is an example of a *segmental model,* in which each sign is broken down into a series of segments. The two major segments in this model are **movements** and **holds.** Movements are defined as those segments during which some aspect of the sign’s configuration changes, such as a change in handshape, a hand movement, or a change in hand orientation. Holds are defined as those segments during which all aspects of the sign’s configuration remain unchanged; that is, the hands remain stationary for a brief period of time.

Signs are made up of sequences of movements and holds. For example, the sign for MOTHER consists of three movements followed by a hold (Figure 2 on the next page). Movement segments have features that describe the type of movement (straight,

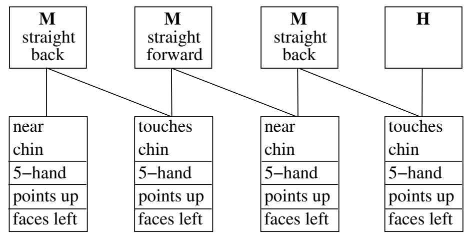

Figure 2: Schematic description of the sign for MOTHER in the Movement-Hold model. See Figure 1 to see how the sign is articulated.

Figure 3: The sign for INFORMATION demonstrates how several features in ASL change simultaneously. Both hands move, starting at different body locations. Simultaneously, the handshapes change from a flat, closed hand to a cupped open hand during the sign.

round, sharply angled), as well as the plane and intensity of movement. In addition, attached to each segment is a **bundle of articulatory features** that describe the hand configuration, orientation, and location. Further details on this model and how to map it to HMMs for sign language recognition can be found in [17, 18].

There are other segmental models, such as D. Brentari's syllable-based model [4], and further work based on the Movement-Hold model [10]. These models differ primarily in what exactly constitutes a segment, but all share a common set of assumptions about the segmental structure of signs. Whether the Movement-Hold model is truly the best choice among the segmental models for ASL recognition systems is an open question that should be resolved in future research.

3.2 Independent Channels

The representation of the sign in Figure 2 has a serious problem that makes it inadequate for a direct application to recognition systems. It occurs when we want to model and recognize the simultaneous aspects of ASL. In particular, the handshape and hand orientation can change while the hand moves. In addition, some signs are two-handed, so both hands must be modeled. Figure 3 shows an example of a two-handed sign, during which also multiple aspects of the sign change at the same time.

The problem is the sheer number of possible combinations of simultaneous features. Unlike speech, where phonemes occur only in sequence, in ASL phonemes occur both in sequence and in parallel.

To get an idea of the magnitude of the problem that recognition systems face here, it is illuminating to consider the completely naïve approach first: what about simply tossing all possible combinations of features together, without regard for linguistic constraints and interdependencies between features? Then the recognition system would have to look at all possible combinations of handshapes, hand orientations, wrist orientations, and locations and movement types of the left and right hands, respectively. This approach leads to a combinatorial explosion, as can easily be seen by multiplying all the numbers of possible respective features together. If, for example, we assume that there are 40 distinct handshapes, 8 distinct hand orientations, 8 wrist orientations, and 20 major body locations, then the number of possible feature combinations, for both the left hand and the right hand, would be $(40 \times 8 \times 8 \times 20)^2$, which is more than one billion.

The combinatorial explosion stemming from the naïve approach highlight the enormous complexity of modeling the simultaneous aspects of ASL. Modeling all these combinations *a priori* is infeasible both from a modeling and a computational point of view. From the modeling point of view, it would be impossible to collect enough data for all these combinations in a reasonable time frame. From the computational point of view, attempting to identify the correct models and to match them against the sequence of signs to be recognized would take far too long to be of practical use.

Of course, it would be ludicrous to assume that ASL really exhibits that many combinations, as is evident from the many linguistic constraints, such as handshape constraints [1], and dependencies between handshape, hand orientation, and location (e.g., signs that touch a part of the signer's body with the thumb or a specific finger have only a few orientations that are even physically possible). Unfortunately, the current state of the art in sign language linguistics and recognition makes attempts at enumerating all the valid combinations infeasible.

To make modeling the simultaneous aspects of ASL tractable, it is necessary to decouple the simultaneous events from one another. Instead of attaching bundles of articulatory features to the segments, we break up the features into **channels** that can be viewed as being **independent** from one another. For the purposes of the work presented in this paper, these channels are, one for each hand:

movement channel the channel consisting of the body locations and the movements between the body locations

handshape channel the channel consisting of the handshape

Note that future work on ASL recognition would also have to tackle the hand orientation and facial expressions, possibly in additional channels. These topics, however, are beyond the scope of this paper.

Figure 4 on the following page shows how the sign for INFORMATION is represented with the independent channel modeling approach. Note how each channel still contains movements and holds, with the same underlying idea as before: in a move-

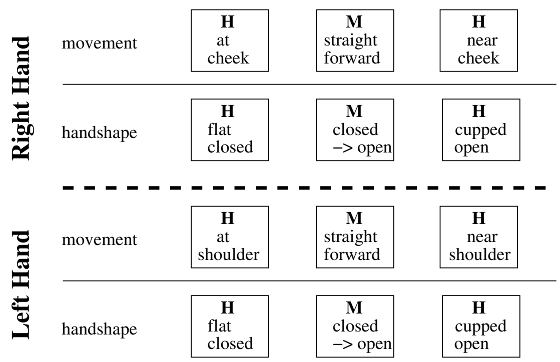

Figure 4: The sign for INFORMATION, where the different features are modeled in separate channels. Note how several channels change simultaneously. See Figure 3 on page 5 for a picture of this sign.

ment segment there is a transition from one configuration to another one in the channel, whereas in a hold segment the configuration remains static.

Splitting the feature bundles up into independent channels immediately yields a major reduction in the complexity of the modeling task. It is no longer necessary to consider all possible combinations of phonemes, and how they can interact. Instead, it is enough to model the phonemes in a single channel at a time, and just to look at the phonological and phonetic phenomena in each channel separately. In each channel, the phonemes can be represented by only a small number of different HMMs, which all belong to the same aspect of the sign's configuration, such as the handshape, or hand movement. Combinations of phonemes from different channels are easy to put together on the fly at recognition time, particularly in conjunction with the parallel hidden Markov model recognition framework that we briefly describe in Section 4.

The downside of using independent channels is that they entail making a major assumption about the structure of the simultaneous processes in ASL, which in all likelihood is not valid. In essence, these independent channels are a case of an engineering tradeoff, so as to make the recognition problem tractable, versus theoretically correct modeling of ASL. Yet, the experiments in Section 5.2 show that modeling ASL in terms of independent channels yields tangible benefits to sign language recognizers.

We now briefly discuss how to model the handshape in this framework.

3.3 Handshape Modeling

Most approaches in the past used joint and abduction angles as features for the hand, whenever these had been available, such as [7]. These are very low-level features, however. The experimental results shown in Section 5.1 indicate that these are not the best choice for recognizing the handshape.

Therefore, it makes sense to use a more high-level description of the handshape. For sign languages, in line with Sandler's phonological model of the handshape [12],

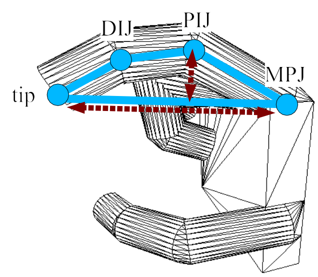

Figure 5: Measure of the openness of a finger. It depends on the width and height of the quadrilateral described by the sites on the three finger joints and the fingertip.

a representation of the degree of openness of a finger seems particularly useful. For obtaining such a representation, consider the relationship among the fingertip, the metacarpophalangeal joint (MPJ, the finger joint closest to the palm), and the degree of openness. If a finger is fully extended — that is, fully open —, the distance between the fingertip and the MPJ is maximized. Likewise, if the finger is fully bent — that is, fully closed —, the distance between these two points is minimized. For obtaining yet another measure of the openness, consider connecting the fingertip and the three finger joint angles into a quadrilateral, as shown in Figure 5, where the base is the line between the fingertip and the MPJ. The height of the quadrilateral, expressed as the distance between the proximal interphalangeal joint (PIJ, the joint closest to the MPJ) and the base line, is maximized when the finger is fully closed. Likewise, the height is minimized when the finger is fully open.

Therefore, these two measures — the width and the height of the quadrilateral described by a finger — are a direct expression of a finger's openness. Note that the MPJ angle only rotates this quadrilateral, but does not affect its dimensions. Thus, whether a finger is open or closed is largely independent of the MPJ angle. Hence, the MPJ angle should be part of the feature vector in addition to the width and height measurements. Together with the abduction angles (the angles measuring the spread between adjacent fingers), these features all together constitute a somewhat higher-level representation of the hand than the raw joint angles. The experiments in Section 5.1 justify this representation.

4 Hidden Markov Models

One of the main challenges in ASL recognition is to capture the variations in the signing of even a single human. In general, humans never perform exactly the same movement twice, even if they intend to. There are always slight variations from one movement to the next one, so a recognition framework must be able to account for them. The most common approach toward handling such variations is to use some kind of statistical model. Hidden Markov models (HMMs) are a type of statistical model well suited for capturing variations. In addition, their state-based nature enables them to describe how a signal changes over time, which is ideal for activity recognition.

HMMs have been used extensively in gesture and sign language recognition in the past, both for whole-sign modeling (one HMM per sign [13]) and phoneme-based modeling (one HMM per phoneme [17]). The phoneme HMMs are chained together to form a single HMM for a sign, and these, in turn, are chained together into a network. The recognition algorithm finds the most probable path through the network and recovers the signs that the path passes through. The parameters of the HMMs are estimated on a training data set, as described in [11].

To handle independent channels, we use parallel HMMs (PaHMMs) as an extension to conventional HMMs. Instead of using only one single network, PaHMMs use a network of HMMs per channel. The recognition algorithm searches all networks in parallel and combines the path probabilities from each path by multiplying them. This multiplication is possible, because in Section 3.2 we assume that the channels are independent from one another, and hence the HMM networks are stochastically independent from one another.

The details of mapping phonemes to HMMs are described in [17], and the recognition algorithm for PaHMMs with multiple channels is covered extensively in [18]. We now describe the recognition experiments that validate our approach.

5 Experiments

We ran two types of experiments. The first one was designed to validate our choice of handshape features, and the second one was designed to validate the modeling of the independent channels.

The data set consisted of 499 sentences, between 2 and 7 signs long, and a total of 1604 signs from a 22-sign vocabulary, the details of which can be found in [17]. We collected these data with an Ascension Technologies MotionStar™ system at 60 frames per second. In addition, we collected data from the right hand with a Virtual Technologies Cyberglove™, which records wrist yaw, pitch, and the joint and abduction angles of the fingers, also at 60 frames per second. We split the data set into 400 training sentences and 99 test sentences. No part of the testing data set was used for training the HMMs at any time, and conversely no part of the training data set was used for the recognition experiments.

We now discuss and give the results for the two types of experiments.

5.1 Handshape Experiments

To determine the relative merits of using joint angles and measures of finger openness, we ran continuous recognition experiments on only the handshape channel. These experiments were designed to compare the robustness of using raw joint angles versus the robustness of the quadrilateral-based representation from Section 3.3.

For each feature vector, the experiments varied the number of HMM states and parameters. Because many signs share the same handshape, it is impossible to identify a sign uniquely from the handshape alone. For this reason, the evaluation criterion was

Feature Vector	μ	σ	Median	Best	N
joint angles	83.15%	15.44%	82.66%	98.68%	1912
quadrilateral	95.21%	5.37%	96.83%	99.47%	1956

Table 1: Results of continuous handshape feature vector comparisons. μ, σ, Median, Best, and N correspond to the average handshape accuracy, standard deviation, best case, and number of experiments, respectively.

the percentage of correctly recognized handshapes, rather than the percentage of correctly recognized signs. In addition, whenever the same handshape occurred multiple times in a row, we contracted these occurrences into a single handshape. The results, given in Table 1, show clearly that the quadrilateral-based description of the handshape is far more robust than the raw joint angles.

5.2 Recognition of Independent Channels

We ran four experiments on the 22-sign set to evaluate the recognition accuracy of modeling varying numbers of channels. The first experiment was a baseline experiment with conventional HMMs for just the right hand's movement channel, from earlier work [17]. The second experiment used PaHMMs to capture the movement channels of both hands. The third experiment used PaHMMs to capture the movement and handshape channels of the right hand. The fourth experiment used PaHMMs to capture all three channels.

The results are given in Table 2 on the next page. Of particular note is that using all three channels shows no improvement over just using the movements and handshape from the right hand. We think that the reason is the relatively small size of the data set that we used. Clearly, future work needs to validate our approach with much larger data sets. Yet, overall the results are promising and validate the assumption that in practice channels can be modeled independently from one another.

6 Conclusions and Outlook

We have developed a framework for ASL recognition that is based on breaking down the signs into their constituent phonemes, and on modeling simultaneous events in stochastically independent channels. In addition, we have integrated the handshape into the framework, in an extension over previous work. We have validated our approach in experiments with up to three channels with a 22-sign vocabulary.

The next step is to validate the framework with larger data sets. It is imperative that the data set comes from native signers in the future, because their articulation of the signs shows characteristics that nonnative signers simply do not exhibit. In addition, future work needs to integrate facial expressions into the framework, and model some of the grammatical processes in sign languages, such as the use of space to denote referents.

Type of experiment	Sentence accuracy	Word accuracy	Details
baseline: movement channel right hand, HMM	80.81%	93.27%	H=294, D=3, S=15, I=3, N=312
movement channel both hands, PaHMM	84.85%	94.55%	H=297, D=3, S=12, I=2, N=312
movement channel right hand, hand-shape right hand, PaHMM	88.89%	96.15%	H=302, D=2, S=8, I=2, N=312
all three channels, PaHMM	87.88%	95.51%	H=301, D=1, S=10, I=3, N=312

Table 2: Comparison of conventional HMMs and PaHMMs. The conventional HMMs modeled the strong hand's movement channel, whereas the PaHMMs modeled a combination of multiple channels. H denotes the number of correct signs, D the number of deletion errors, S the number of substitution errors, I the number of insertion errors, and N the total number of signs in the test set.

Acknowledgments

This work was supported in part by NSF EIA-98-09209, AFOSR F49620-98-1-0434, and NSBRI-NBPF00202.

References

[1] R. Battison. *Lexical borrowing in American Sign Language*, pages 19–58. Linstok Press, Silver Spring, MD, 1978. Reprinted as "Analyzing Signs" in Lucas, C. and Valli, C. Linguistics of American Sign Language, 1995.

[2] B. Bauer and K.-F. Kraiss. Towards an automatic sign language recognition system using subunits. In I. Wachsmuth and T. Sowa, editors, *Gesture and Sign Language in Human-Computer Interaction. International Gesture Workshop*, volume 2298 of *Lecture Notes in Artificial Intelligence*, pages 64–75. Springer, 2001.

[3] B. Bauer and K.-F. Kraiss. Video-based sign recognition using self-organizing subunits. In *International Conference on Pattern Recognition*, 2002.

[4] D. Brentari. Sign language phonology: ASL. In J. A. Goldsmith, editor, *The Handbook of Phonological Theory*, Blackwell Handbooks in Linguistics, pages 615–639. Blackwell, Oxford, 1995.

[5] G. Fang, W. Gao, X. Chen, C. Wang, and J. Ma. Signer-independent continuous sign language recognition based on SRN/HMM. In I. Wachsmuth and T. Sowa, editors, *Gesture and Sign Language in Human-Computer Interaction. International Gesture Workshop*, volume 2298 of *Lecture Notes in Artificial Intelligence*, pages 76–85. Springer, 2001.

[6] H. Hienz, K.-F. Kraiss, and B. Bauer. Continuous sign language recognition using hidden Markov models. In Y. Tang, editor, *ICMI'99*, pages IV10–IV15, Hong Kong, 1999.

[7] R.-H. Liang and M. Ouhyoung. A real-time continuous gesture recognition system for sign language. In *Proceedings of the Third International Conference on Automatic Face and Gesture Recognition*, pages 558–565, Nara, Japan, 1998.

[8] S. K. Liddell and R. E. Johnson. American Sign Language: The phonological base. *Sign Language Studies*, 64:195–277, 1989.

[9] V. Pavlovic, R. Sharma, and T. S. Huang. Visual interpretation of hand gestures for human-computer interaction: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):677–695, 1997.

[10] D. Perlmutter. Sonority and syllable structure in American Sign Language. *Linguistic Inquiry*, 23(3):407–442, 1992.

[11] L. R. Rabiner. A tutorial on Hidden Markov Models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

[12] W. Sandler. Representing Handshapes. In W. H. Edmondson and R. Wilbur, editors, *International Review of Sign Linguistics*, volume 1, chapter 5, pages 115–158. Lawrence Erlbaum Associates, Inc., Mahwah, NJ, 1996.

[13] T. Starner, J. Weaver, and A. Pentland. Real-time American Sign Language recognition using desk and wearable computer based video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(12):1371–1375, 1998.

[14] W. C. Stokoe. *Sign Language Structure: An Outline of the Visual Communication System of the American Deaf*. Studies in Linguistics: Occasional Papers 8. Linstok Press, Silver Spring, MD, 1960. Revised 1978.

[15] C. Valli and C. Lucas. *Linguistics of American Sign Language: An Introduction*. Gallaudet University Press, Washington DC, 1995.

[16] 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, 1997.

[17] C. Vogler and D. Metaxas. Toward scalability in ASL recognition: Breaking down signs into phonemes. In A. Braffort, R. Gherbi, S. Gibet, J. Richardson, and D. Teil, editors, *Gesture-Based Communication in Human-Computer Interaction*,

volume 1739 of *Lecture Notes in Artificial Intelligence*, pages 211–224. Springer, 1999.

[18] C. Vogler and D. Metaxas. A framework for recognizing the simultaneous aspects of American Sign Language. *Computer Vision and Image Understanding*, (81):358–384, 2001.

[19] C. Wang, W. Gao, and J. Ma. A real-time large vocabulary recognition system for Chinese Sign Language. In I. Wachsmuth and T. Sowa, editors, *Lecture Notes in Artificial Intelligence*, volume 2298, pages 86–95. Springer, 2002.