A Survey of Sign Language Recognition Research: Postprint

Authors: Zhang Liangguo, Xilin Chen

Date: 2016-11-02T00:00:00+00:00

Abstract

This paper surveys recent research hotspots and domestic and international research progress in sign language recognition, primarily covering: according to different input modalities, introducing representative domestic and international works from both data glove-based and vision-based recognition approaches; subsequently summarizing mainstream recognition methods in current research; and based on analysis of the research status, identifying persistent challenges in sign language recognition along with possible countermeasures. Furthermore, this paper briefly introduces the progress of our research group in related research work.

Full Text

Preamble

Vol.7 No.3
Information Technology Letter

A Survey of Sign Language Recognition Research
Zhang Liangguo, Chen Xilin

Abstract: This paper reviews recent research hotspots and advances in sign language recognition, covering representative domestic and international work from two perspectives: data glove-based and vision-based approaches according to different input modalities. On this basis, we summarize current mainstream recognition methods. Through analysis of the research status, we identify remaining challenges in sign language recognition and possible countermeasures. Additionally, we briefly introduce the progress of our research group in related work.

Keywords: Sign language recognition; Gesture recognition; Human-computer interaction

1 Introduction

With the rapid expansion of computer influence in modern society, high performance, high intelligence, and high usability are widely recognized as the main directions of current computational science development. Especially today, with the rapid development of computing, communication, and display technologies, traditional mouse-and-keyboard-based human-computer interaction increasingly reveals its limitations. These limitations become particularly evident before new display technologies such as virtual reality and wearable computers, making research on multimodal human-computer interface technology increasingly important in real life. The goal of multimodal human-computer interface research is to master the core technologies for achieving high intelligence and usability of computing devices, establishing a harmonious and natural human-computer interaction environment that allows users to conveniently and naturally use computers in ways familiar to humans. A crucial aspect of this is enabling computers to accurately perceive different human information expression modalities, including natural language, gesture language, and facial language, as well as information jointly conveyed through body gestures, head movements, mouth shapes, expressions, and other body languages beyond spoken and written natural language. While normal-hearing people primarily use spoken and written language supplemented by gestures and expressions, sign language plays an equally vital role in communication among deaf people, just as spoken language does for hearing individuals.

Sign language is the most natural method for deaf people to exchange information and communicate, and it serves as a tool for teaching and conveying ideas in schools for the deaf. Sign language is a human body language that expresses thoughts through hand shapes, arm movements, and is supplemented by facial expressions, lip movements, and other body gestures. It possesses standardized grammar, clear semantics, and a complete vocabulary system. As an important branch of sign language, Chinese Sign Language (CSL) consists of two categories: finger spelling and gesture signs. Finger spelling uses finger trajectories to describe a Chinese pinyin letter and forms language according to pinyin rules. Finger spelling has 30 basic units of finger letters, based on written language, spelling out words by sequentially forming corresponding finger letters to represent a sentence. Any word in Chinese vocabulary can be expressed in finger spelling form. Gesture signs primarily simulate object shapes and movements, supplemented by postures and expressions. Gesture signs constitute the main part of the language used by deaf people in daily life. Currently, Chinese sign language textbooks include approximately 5,500 conventional gestures, with each gesture corresponding to a Chinese word. Finger spelling and gesture signs each have advantages and disadvantages. The advantage of finger spelling is its simplicity and clarity, being completely consistent with pinyin, allowing it to accurately express content spoken or written, including function words, technical terms, and abstract vocabulary that are difficult to express in gesture signs. Its disadvantage is the lack of intuitive and visual meaning,

making it less comprehensible. The main advantages of gesture signs are their vividness, agility, simplicity, and liveliness, making meanings immediately clear. The disadvantage is that due to the large number of Chinese characters and words, gestures cannot comprehensively and accurately express their meanings. Therefore, deaf people in China currently primarily use gesture signs in communication, appropriately supplemented by finger spelling.

The goal of sign language recognition is to provide an effective and accurate mechanism through computers to translate sign language into text or speech, making communication between deaf and hearing people more convenient and efficient. It has become one of the most important research topics in the human-computer interface field, attracting increasing attention from experts and scholars. Notably, many developed countries in Europe and America have dedicated research funding for this area. With over 20 million deaf people in China, research on sign language recognition will undoubtedly directly benefit this group by providing a more natural and convenient way to communicate with hearing people, helping them better integrate into society and positively contributing to building a harmonious society with diverse care.

Sign language recognition research involves multiple disciplines including robotics, spatial geometry, psychology, physiology, pattern recognition, probability statistics, and computational linguistics. It is also a typical case study in pattern recognition, artificial intelligence, and natural language understanding. From a theoretical and technical perspective, sign language recognition research not only serves as a platform for practical application of cutting-edge technologies from various disciplines but also, in turn, drives development in these research fields. Reproducing and recognizing sign language through computers represents both a comprehensive test of computational power and knowledge representation capabilities, and an exploration of the working mechanisms of human physiological cognitive functions and pattern recognition abilities—making it a highly challenging scientific research endeavor. Therefore, sign language recognition research holds significant importance for socio-economic development and scientific advancement [**?**][**?**]:

- Sign language recognition can help deaf people, especially those with lower literacy levels, communicate with normal-hearing people using sign language, which is particularly important in public service institutions such as hotels, stations, and hospitals.
- From a cognitive science perspective, sign language research involves the mechanisms of human visual language understanding, helping to improve computers' ability to comprehend human language.
- Sign language research enables us to use gestures to control intelligent agents in virtual reality, remotely operate multi-joint artificial arms, and control household appliances such as televisions [**?**][**?**][**?**].
- Sign language research provides demonstration learning functions for robots [**?**], while recognition can also provide correction functions for sign language learners.

- As an important component of multimodal human-computer interfaces, sign language recognition can provide a completely new input method.

In summary, sign language recognition research possesses not only profound theoretical value but also broad practical application prospects.

Research on simple gesture recognition began in the 1980s. Historically, G.J. Grimes of AT&T obtained the "Digital Data Entry Glove Interface Device" patent in 1983 [**?**]. The glove could recognize 72 one-handed letters, making him the pioneer in gesture recognition research. Sign language recognition research started later, beginning in the 1990s. In recent years, with the development of multifunctional perception, intelligent human-computer interfaces, and virtual reality research, sign language recognition has gained increasing international attention. Universities, major companies, and research institutions worldwide have joined the research effort, studying sign languages including at least Chinese Sign Language, American Sign Language, Japanese Sign Language, Korean Sign Language, Irish Sign Language, British Sign Language, German Sign Language, Arabic Sign Language, Turkish Sign Language, and Australian Sign Language.

Currently, implemented sign language recognition systems are mainly divided into sensor-based systems (such as data gloves + position trackers) and vision-based systems according to different input modalities. Sensor-based recognition systems use data gloves and position trackers to measure hand joint angle information, as well as hand trajectory and temporal information in space for recognition. The advantage of this method is its ability to conveniently obtain accurate hand shape, position, orientation, and motion trajectory information, making it suitable for large-vocabulary sign language recognition with high recognition rates. However, the disadvantage is that the system requires users to wear complex data gloves and position trackers, causing inconvenience and affecting the naturalness of human-computer interaction. Additionally, the input devices are relatively expensive, making large-scale promotion somewhat difficult. Vision-based recognition methods use digital cameras to capture gesture video/image information for subsequent recognition processing. This approach provides better naturalness for human-computer interaction. The difficulty lies in the fact that current computer vision technology still struggles to reliably obtain sufficiently detailed gesture feature information from natural images. To improve system robustness, some systems use gloves with color markers, fingertip markers, or gloves with highlighted joint markers as gesture input. Overall, vision-based methods have relatively lower recognition rates and poorer real-time performance, making them currently unsuitable for large-vocabulary sign language recognition. Recently, some researchers have adopted hybrid recognition methods combining the two approaches [**?**][**?**], which serves as a feasible temporary compromise.

Below, we first review representative domestic and international research work on data glove-based sign language recognition and vision-based sign language recognition, then summarize the main recognition methods, and subsequently

identify the challenging problems that remain in sign language recognition research along with possible countermeasures.

2.1.1 Data Glove-Based Sign Language Recognition

Data gloves can accurately capture hand shape, position, orientation, and motion trajectory information for sign language, making them suitable for large-vocabulary sign language recognition with high recognition rates. Currently, data gloves used in research mainly include: the VPL Data Glove developed by Thomas Zimmerman in 1985 [**?**], the low-cost PowerGlove developed by toy manufacturer Mattel in 1989 [**?**], and the CyberGlove developed by James Kramer [**?**].

The VPL Data Glove collects data through fiber optic sensors located on the back of fingers (two per finger). Finger bending causes the fibers to bend, reducing the amount of transmitted light. Analog signals are sent to a processor that determines finger bending degrees. The glove is equipped with hand shape recognition software that maps joint configurations to commands. Hand position and orientation are measured by a Polhemus 3D position tracker. The PowerGlove was developed under the influence of VPL. It measures each finger' s bending degree through damping inkjet sensors, resulting in measurement inaccuracy. The PowerGlove provides 6-dimensional information about hand orientation and position through ultrasonic sensors. After developing the CyberGlove, Kramer created Virtual Technologies, specializing in the commercial sale of CyberGlove data gloves. The CyberGlove has two types: 18-sensor and 22-sensor versions. The 18-sensor CyberGlove model includes two bending sensors per finger, one abduction sensor, one sensor connecting the thumb and little finger, and one sensor at the wrist, measuring hand and wrist angles during finger bending/extension, palm curvature, and wrist rotation. The sensor output range is 8 bits (0-255). This data glove can output very accurate, high-quality sensor data.

Data glove-based sign language recognition systems can be further divided into finger spelling recognition, sign word recognition, and continuous sign language recognition according to the recognition target.

Finger spelling, as a static gesture word, features easy recognition and is suitable for trying new algorithms, making it a starting point for many early researchers. Finger spelling recognition algorithms mainly include machine learning methods such as ID3, NewID, C4.5, CN2, HCV, RIEVL [**?**], and neural network methods such as Radial Basis Function (RBF) [**?**] and Min-Max fuzzy neural network methods [**?**]. Representative work includes: Takahashi and Kishino from ATR Laboratory in Japan [**?**] designed a VPL data glove-based system that could recognize 34 of 46 user-dependent Japanese letters using joint angle and hand orientation encoding techniques. The system integrated principal component analysis and clustering analysis techniques. C. Lee and Y. Xu from Carnegie Mellon University (CMU) [**?**] developed a finger spelling recognition system

based on Hidden Markov Models (HMM). Using a CyberGlove data glove as the input interface, the system performed online robot demonstration learning –learning and recognition of new gestures. The system demonstrated credible recognition of 14 gestures from American Sign Language letters using only one or two examples per gesture, showing its potential as part of a robot teleoperation and example-based planning interaction interface.

Unlike static finger spelling, sign word expression is a dynamic time series that must consider its temporal characteristics in recognition and modeling. Representative work includes: S.S. Fels and G.E. Hinton [**?**][**?**] used VPL data gloves and Polhemus position trackers as input devices and neural networks as gesture classifiers. Based on features including hand motion trajectory, hand shape, hand movement direction, hand offset, and hand movement speed, they formed five functional networks to recognize 203 sign words. J.L. Hernandez et al. [**?**] used AcceleGlove to recognize 176 American Sign Language words. They divided sign language into 42 hand shapes, 6 orientations, 11 positions, and 7 motion units, then used conditional template matching to synthesize probabilities from these data, achieving a 95% recognition rate.

In continuous sign language recognition, representative work includes: R.H. Liang and M. Ouhyoung [**?**] implemented a continuous Taiwanese Sign Language recognition translator using Hidden Markov Models. The system targeted basic vocabulary and practice sentences from Taiwanese Sign Language textbooks, with a vocabulary size of 71-250, and sign language data was collected using VPL data gloves. However, to detect boundaries between words, sign language had to be performed slower than normal speed. H. Sagawa and M. Takeuchi from Hitachi Laboratory in Japan [**?**] used changes in hand shape, orientation, and position information to detect boundaries in continuous Japanese Sign Language words, achieving approximately 83% recognition rate on 200 sign samples consisting of 10 sign language sentences. Recently, the system has been extended and applied to establish Japanese Sign Language recognition information kiosks in government departments [**?**][**?**]. The system can recognize 268 Japanese Sign Language words and 52 types of sign language sentences, enabling hearing-impaired people to easily search for various information and services they need. Surveys indicate that most users highly approve of the system. Gao Wen et al. [**?**][**?**][**?**] used dynamic programming methods to obtain context-dependent models for recognizing continuous Chinese Sign Language. The system used data gloves as input and stream-tied Hidden Markov Models as the recognition method. It achieved 94.8% accuracy in recognizing 5,177 sign language words and 91.4% accuracy in recognizing 200 sentences composed of words from this vocabulary set.

2.1.2 Vision-Based Sign Language Recognition

Due to occlusion, projection, and lighting effects in 2D visual images, vision-based methods struggle to accurately track finger bending information for each finger. Consequently, vision-based sign language recognition research has fo-

cused on small to medium vocabulary sizes. According to different recognition targets, vision-based sign language recognition systems can be further divided into gesture recognition, sign word recognition, and continuous sign language recognition.

Compared with sign language recognition research, gesture recognition research started earlier. Gestures encompass a very broad range. In some literature, finger spelling and sign words are collectively referred to as gestures. Here, we use "gesture"to refer to those gestures beyond standard sign language vocabulary with clear definitions. Gesture recognition research primarily focuses on using gestures for control. References [**?**][**?**] provide a relatively comprehensive review of vision-based gesture recognition research history, which will not be repeated here.

Unlike gestures, sign words are more structured gestures with precise meanings in sign language dictionaries. However, algorithmic research for both is very similar, as both involve recognizing time series problems. Representative work includes: C. Charayaphan and A. Marble [**?**] used image processing methods to understand 31 isolated gesture words in American Sign Language, correctly recognizing 27 of them. Y. Cui [**?**][**?**] adopted multi-class high-dimensional discriminant analysis to automatically select the most discriminative features for sign language classification, implementing a visual sign language recognition system that could recognize 28 American Sign Language words using a recursive partitioning tree for sign classification, achieving a 93.2% recognition rate. M.H. Yang et al. [**?**] used motion trajectories to extract and classify 2D motion from image sequences, then used time-delay neural networks to recognize 40 American Sign Language gestures, achieving a 98.14% recognition rate. K. Grobel and M. Assan from Germany [**?**] used Hidden Markov Model methods to recognize 262 isolated Dutch Sign Language words, achieving a 91.3% recognition rate. The system was vision-based, requiring signers to wear colored gloves, then extracting 2D features from video. Dublin City University in Ireland conducted research on Irish Sign Language recognition. A. Shamaie and A. Sutherland [**?**] proposed a hierarchical sign language recognition algorithm that recognized 100 sign words with the help of solid-color gloves, achieving an 89.6% recognition rate. J.W. Deng and H.T. Tsui from the Chinese University of Hong Kong [**?**] used a PCA/MDA (Multiple Discriminant Analysis, an extension of Linear Discriminant Analysis) combination method to recognize 100 sign words. To simplify gesture segmentation, the system used colored gloves. In reference [**?**], they further studied the system, using parallel Hidden Markov Models to recognize 192 American Sign Language words. The method consisted of two steps: first, using nine defined motion primitives to select candidate vocabulary with similar motion and position; then, unifying three channels of information (left hand motion, right hand motion, and right hand shape information) under the parallel Hidden Markov Model framework to select the word with the highest matching score from the candidate vocabulary in the first step.

In continuous sign language recognition, representative work includes: T.

Starner et al. from MIT Media Laboratory [?] conducted research on American continuous sign language recognition. They used feature vectors of hand shape, orientation, and motion trajectory information as input to Hidden Markov Models to recognize American Sign Language. To facilitate tracking, users were required to wear colored gloves (yellow for the right hand, orange for the left). Testing was conducted on short sentences randomly composed of 40 vocabulary words, with a system recognition rate of 91.3%, which increased to 98% real-time recognition after adding certain grammatical constraints. B. Bauer and H. Hienz [?] used a color camera as the input device and Hidden Markov Models as the recognition method to recognize continuous German Sign Language sentences composed of 97 vocabulary words, achieving a 91.7% recognition rate. The system was further developed based on Grobel and Assan' s system. Additionally, they used K-means clustering algorithms to automatically acquire basic units in sign language, then used these primitives for continuous sign language recognition [?]. Experiments on 12 different gesture words obtained 10 primitives through clustering, achieving an 80.8% recognition rate for continuous sign language using these primitives. In reference [?], they experimented with a 100-sign vocabulary set composed of 150 primitives, achieving a 92.5% recognition rate. The method achieved an 81% recognition rate when recognizing 50 untrained sign vocabulary words. C. Vogler and D. Metaxas [?] studied American Sign Language recognition, using a position tracker and three mutually perpendicular cameras as gesture input devices, employing computer vision methods to extract 3D parameters of signers' arm movements, and using Hidden Markov Model recognition technology to complete recognition of 53 isolated words and 486 continuous American Sign Language sentences. In their experiments, they used basic units of sign language rather than sign vocabulary [?]. Experimental results on sentences composed of 22 words showed that this method' s recognition rate was similar to traditional methods. Additionally, Vogler and Metaxas pointed out that the challenging issue in sign language recognition is developing a scalable recognition method for sign language vocabulary. They proposed using parallel Hidden Markov Models [?] to solve this problem. In parallel Hidden Markov Models, each data stream is processed independently, allowing each stream to be trained independently without considering their combination. Research on a 22-sign vocabulary set showed that this method could improve recognition robustness.

Domestic research in this area is relatively limited, mainly including: Hangzhou University studied preliminary recognition of finger spelling letters based on vision; Harbin Institute of Technology Computer Science Department used edge detection, neural networks, and other methods to implement a vision-based gesture recognition system that could recognize 13 static gestures and a simple vision-based continuous dynamic gesture recognition system [?], which did not require users to wear any additional devices; Tsinghua University also conducted extensive research on vision-based gesture recognition [?][?].

Additionally, some work has been conducted on viewpoint-independent

recognition. Since images change when the viewpoint changes and viewpoint-independent features are difficult to extract, most vision-based sign language recognition research restricts the viewpoint between users and digital camera equipment. Restricting viewpoint is not conducive to user experience or the promotion of sign language recognition systems, making it necessary to research viewpoint-independent sign language recognition technology. Among existing sign language recognition technologies, three approaches can be attempted to address viewpoint variation: first, 3D feature-based methods [**?**]—3D features are independent of viewpoint and can achieve viewpoint-independent sign language recognition, but reliably extracting 3D feature information of moving hands through vision is not easy; second, carefully designed appearance-based learning methods [**?**]—this approach has some feasibility for static finger spelling but is difficult to apply to dynamic sign language; third, semantic feature-based recognition methods [**?**]—from the perspective of semantic features themselves, this approach can achieve viewpoint-independent sign language recognition, but no good solution has been found for how to extract viewpoint-independent semantic features from visual channels. There is still a long way to go to achieve viewpoint-independent sign language recognition. Notably, in the field of activity analysis, researcher C. Rao et al. cleverly solved the viewpoint variation problem in activity matching using epipolar geometry technology [**?**]. This epipolar geometry-based approach to handling viewpoint variation has good reference significance for viewpoint-independent sign language recognition.

Furthermore, sign language recognition researchers have mainly focused on small to medium-scale vocabularies (according to speech recognition research classification standards: vocabulary size less than 100 is called small vocabulary, 100 to 500 is medium vocabulary, and over 500 is large vocabulary) and have limited research to specific individuals. For non-specific person sign language recognition, P. Vamplew [**?**] used data gloves as input devices, employing four neural network modules for feature extraction of hand shape, orientation, position, and motion, then using nearest-neighbor decision classification to recognize 52 isolated sign words. Since the system used only one data glove, it could only recognize one-handed sign words. Another non-specific person work is by S. Akyol and U. Canzler [**?**], who used Hidden Markov Model methods to reliably recognize 16 German Sign Language words with the help of solid-color gloves. The system calculated hand features using the signer' s head position as the reference coordinate system, using data from 7 people for training and 3 people for testing, achieving a 94% recognition rate.

Additionally, inspired by adaptive recognition ideas in speech recognition research, some scholars have conducted work on adaptive sign language recognition, hoping to use adaptive techniques to solve the non-specific person problem in sign language recognition. Representative work includes: S.C.W. Ong and S. Ranganath from National University of Singapore [**?**] recognized 6 isolated words and 5 implied words, achieving a relative error rate reduction of 75.7% using less adaptive data; U.V. Agris et al. from RWTH Aachen University in Germany [**?**] used Maximum Likelihood Linear Regression and Maximum A Pos-

teriori (MAP) probability combination methods to adapt 153 one-handed words, achieving a 41.6% relative recognition rate improvement using 80 adaptive data sequences.

2.2 Main Recognition Methods

Currently, commonly used sign language recognition models/technologies mainly include: Hidden Markov Models, Temporal Template Matching (TTM), Artificial Neural Networks (ANN), and Artificial Neural Network/Hidden Markov Model hybrids.

Hidden Markov Model is an extension of Markov models. A general Markov model describes a random process—transitions between states; Hidden Markov Model describes two random processes: one describes the probabilistic relationship between output symbols and states, i.e., output symbols are random process functions of states; the other describes the transition relationship between states. From an external observer' s perspective, only output symbols can be seen, not state transitions—hence state transitions are hidden. Hidden Markov Models have become the mainstream method in current sign language recognition research due to their excellent ability to model temporal signals and capability for unsupervised learning. However, Hidden Markov Models themselves have some defects: 1) Traditional Hidden Markov Models assume that the density distribution of all individual observation variables follows a Gaussian mixture density or autoregressive density, which often differs from reality; 2) The Markov assumption may not hold; 3) The output independence assumption may not hold; 4) Due to limitations in training criteria and algorithms, Hidden Markov Models have relatively poor pattern recognition capabilities.

Template matching is the simplest classification method, typically including two processes: template establishment and classification. By collecting representative samples for each category, one or a set of templates is established for each gesture. Classification determines the input' s category based on similarity between the input and each template. Typically, the similarity function is defined as the Euclidean distance between the template and observed input data. During recognition, a similarity threshold is usually set—the input is recognized as the class with the closest distance within the threshold, while inputs above this threshold are rejected. The advantage of this method is that templates are easy to establish and improve, and template rules are intuitive, clear, and easy to understand. The disadvantage is that due to noise and ambiguity between different classes, classification results lack robustness when recognizing many categories. Gesture word recognition can also use template matching for similarity calculation, but this requires solving the time alignment problem: each basic gesture word in sign language contains some fundamental gestures, and the duration length of each sample of the same word collected varies randomly. How to measure similarity between gestures of different lengths is a time alignment problem. To solve this problem, two typical temporal template matching methods have been proposed in literature: Dynamic Time Warping and Finite

State Machines.

Artificial Neural Network is a complex information processing network formed by extensively connecting a large number of simple processing units, where processing units and their interconnection patterns are designed by borrowing from human neuron structures and connection mechanisms. This network has learning, memory, knowledge generalization, and input information feature extraction capabilities similar to the human brain. Since the 1980s, artificial neural network research has experienced a new boom, attracting great attention due to its nonlinear, adaptive, robust, and learning characteristics. Although statistical model methods dominate sign language recognition, neural networks' unique advantages and strong classification and input-output mapping capabilities are highly attractive in sign language recognition research. Currently, commonly used neural networks mainly include Multi-layer Perceptron (MLP), Self-Organizing Feature Map (SOFM), Radial Basis Function networks, Time Delayed Neural Networks (TDNN), and Recurrent Neural Networks (RNN). Since typical neural networks cannot process temporal signals, early sign language recognition applications of neural networks mostly focused on static gesture word recognition. However, Time Delayed Neural Networks and Recurrent Neural Networks can transform temporal information into spatial information, enabling direct dynamic gesture language recognition of temporal signals. Nevertheless, due to their structure, artificial neural networks lack the ability to model long-distance dependencies and thus cannot serve as a general framework for recognizing large-vocabulary, continuous sign language.

Artificial Neural Network/Hidden Markov Model hybrids utilize artificial neural network characteristics to compensate for Hidden Markov Model defects, using the unified framework of Hidden Markov Models and the pattern recognition capabilities of artificial neural networks to produce optimized results. Additionally, regarding hybrid model research for Hidden Markov Models, there are also combinations such as HMM/SVM [**?**]. However, how to combine sub-models in hybrid models requires further research.

Introducing language models into the recognition process is an important approach. Language models attempt to reflect, record, and use natural language patterns to improve the performance of various natural language applications. Common linguistic models include N-gram models, part-of-speech-based language models, decision tree language models, variable-length memory language models, maximum entropy language models, and structure-based language models. These methods have different characteristics: N-gram models are linear symbol co-occurrence-based language models that can only describe surface language information, not language structure. Part-of-speech-based language models reduce data sparsity problems in statistical models through part-of-speech information and enhance model generalization capability, but they also cannot describe language structure. Decision tree language models function similarly to N-gram models, dividing the current word' s context environment into several equivalence classes, but they differ in classification criteria: N-gram classifica-

tion is based on the previous N-1 adjacent words, while decision tree language models only select context features with representational significance, making them more suitable for describing language patterns than N-gram. Variable-length memory language models avoid the exponential parameter space growth of N-gram models with increasing N values. The context length used to predict the next word in this model is not fixed but adaptively extends according to its contribution to the model, enabling increased description of long-distance language relationships. Maximum entropy language models can unify multiple knowledge sources and integrate structural knowledge of language models within their framework, but their parameter training is computationally intensive. How to organically integrate computational linguistic knowledge such as syntax and semantics into maximum entropy language models and extract relevant features under reasonable computational complexity is key to applying maximum entropy language models to natural language processing. Structure-based language models are structured statistical language models that can describe natural language structure, but they depend on syntactic analysis results and have difficult parameter learning, so they are usually used as language feature extraction components under the maximum entropy language model framework.

Currently, commonly used search algorithms in sign language recognition mainly include the Viterbi algorithm, stack decoding algorithm, and Multi-Pass algorithm. The Viterbi algorithm is a dynamic programming-based breadth-first search algorithm that can obtain a globally optimal path using the minimum cost method among all paths. However, since it only preserves local optimal backtracking for each state, it cannot produce the top N solutions (although improved Viterbi algorithms can obtain top N solutions by preserving N local backtrackings, the storage space required is huge and rarely used). The A* stack decoding algorithm belongs to depth-first search, with the advantage that the algorithm' s path retains all prefixes (while Viterbi algorithms can typically only utilize one or two previous contexts), making it particularly suitable for decoding problems in sign language statistical models that require the entire context. However, its time and space complexity are high. Multi-pass search is a widely used method in speech and sign language recognition, organically combining the previous two algorithms to produce a better search strategy. For example, the commonly used forward-backward search algorithm is a strategy that combines A* stack decoding and Viterbi algorithms to generate N-best sequences.

2.3 Existing Problems and Possible Solutions

Obviously, an ideal sign language recognition system should be able to handle all vocabulary defined in sign language itself to maximally meet users' actual needs; it should be able to recognize in real-time, accurately, and reliably in complex environments; and the system should be able to serve non-specific users. Currently, sign language recognition research has nearly 20 years of history and has made great progress, but many challenging problems remain to be solved.

These are briefly summarized as follows:

1. **Establishing transition models between gestures in continuous sentence recognition.** When two gestures are performed sequentially, an extra gesture movement inevitably occurs between them, transitioning from the end state of the first gesture to the initial state of the second gesture. The transition model between gestures is a key challenge in continuous sign language recognition.

2. **Finding scalable sign language recognition methods for large vocabularies.** The difficulty of large-vocabulary recognition lies in the huge search space and the need to distinguish many categories. As vocabulary size increases, potential inter-word similarity increases, making discrimination more difficult. Simultaneously, as vocabulary size increases, system search computational overhead and storage overhead increase, leading to reduced recognition speed and decreased recognition rates. Taking Chinese Sign Language as an example, there are approximately 5,500 conventional gestures. Obviously, training a recognition model for each gesture is unrealistic. Researching scalable sign language recognition methods could become a breakthrough point. A possible approach is to use primitives. Defining the minimal recognition primitives for gestures and automatically finding these minimal recognition primitives is necessary work in this area. However, sign language does not have ready-made linguistically defined primitives like speech, so finding such primitives suitable for computer sign language recognition is currently a research hotspot.

3. **Finding ways to utilize non-hand features such as facial expressions and lip movements.** Sign language itself is a multi-channel language: besides hand movements, some sign words and certain tones require coordination from facial, lip, and even upper body joints. Research shows that facial expressions and lip movements play very important roles in sign language understanding. However, current research has paid little attention to this aspect. Another angle requiring attention is how to fuse these different sign language representation channels to further improve sign language recognition rates.

4. **Achieving non-specific person sign language recognition.** Due to individual differences in hand size, shape, arm length, and sign language performing styles among different signers, non-specific person sign language recognition presents greater difficulties and challenges compared to specific person recognition. One possible approach is to analyze common and personalized information in data. Since different people perform sign language in very different ways, data diversity makes extracting effective common features from non-specific person sign language recognition data extremely difficult. Another possible direction is to directly conduct adaptive research for sign language recognition, performing self-learning based on a universal sign language model using small amounts of data from specific users during use to achieve model transfer, enabling the system to

better recognize specific persons' sign language at specific times. Simultaneously, the system' s self-learning capability should not be limited to specific users to maintain the system' s universal service nature.

5. **Establishing sign language language models.** Previous research shows that introducing language models can improve sign language recognition rates. However, currently used language models almost all come from hearing people' s language corpora, which obviously cannot well reflect sign language's own language model characteristics. In-depth research on sign language language models and how to introduce these models into sign language recognition to improve recognition rates is a fundamental aspect of sign language recognition research.

6. **For vision-based sign language recognition, particularly important is achieving complete and accurate visual feature information acquisition.** Unlike simple gesture recognition (which only requires simple feature description) and data glove-based recognition methods (where sensing devices themselves can quickly obtain accurate hand information), vision-based sign language recognition for larger vocabularies requires front-end visual processing to provide as complete and useful feature information as possible for subsequent high-level recognition stages. Therefore, how to effectively and quickly extract accurate feature data from video is the bottleneck. Specifically: how to reliably and accurately obtain bare hand information; how to ensure robust detection of hand shape and position information based on skin color under complex backgrounds and different lighting conditions; how to identify and interpret occlusions between hands and between hands and face; and how to achieve fusion of different visual channel information such as two-hand shape, position, orientation, facial expressions, lip movements, and body posture.

7. **Seeking ways to achieve viewpoint-independent recognition.** Since human visual recognition is viewpoint-independent, we theoretically hope that computer-provided sign language recognition algorithms are also viewpoint-independent. However, under current technical conditions, most vision-based sign language recognition algorithms restrict camera capture viewpoints, typically limiting them to frontal views. The practical difficulty arises because image features change when capture viewpoints change. As is well known, extracting viewpoint-independent features is relatively difficult. Viewpoint-independent sign language recognition is an important aspect of vision-based sign language recognition research.

3 Our Work

This section focuses on introducing some research work of our group in sign language recognition. First, we introduce research on data glove-based sign language recognition and the integrated sign language recognition and synthe-

sis translation (dialogue communication) system for deaf and hearing people developed based on this. Second, we introduce several other related research works.

3.1 Data Glove-Based Sign Language Recognition

In data glove-based sign language recognition: In 1999, we implemented the basic architecture of a sign language recognition and synthesis system that could recognize 1,065 sign vocabulary words and 80 continuous sign language sentences with a recognition rate of 95.2% [**?**][**?**]. In 2001, the HandTalker system jointly developed by Harbin Institute of Technology and the Institute of Computing Technology, Chinese Academy of Sciences, could recognize 5,100 sign words and 200 continuous sign language sentences, and synthesize 5,500 sign words. The word recognition rate was 94.8%, and it could also recognize 200 sentences composed of words from this vocabulary set with 91.4% accuracy [**?**][**?**]. In 2004, based on this foundation, the HandTalkerII system could recognize 5,100 vocabulary words composed of data from 6 different signers with a recognition rate of 91.6%, and recognize 1,500 samples composed of 750 different sentences with a recognition rate of 91.9%. Simultaneously, it implemented a highly realistic synthesis system with coordinated lip movements and facial expressions synchronized with sign language, where the average intelligibility of finger spelling, word, and sentence synthesis were 92.95%, 88.23%, and 87.58% respectively [**?**]. Additionally, research was conducted on scalable sign language recognition methods and minimal recognition primitives for sign language recognition [**?**]. Through analysis of sign language dictionaries, over 2,400 radicals were summarized and organized, implementing a radical-based recognition system and comparing this method with gesture word-based methods. Research on large-vocabulary sign language recognition [**?**] used Self-Organizing Feature Maps to implicitly extract sign language features, transforming data into a compact, important low-dimensional representation as input for continuous Hidden Markov Models, partially solving the non-specific person recognition problem. A transition model-based method was proposed to solve large-vocabulary continuous sign language recognition, achieving good results in large-vocabulary continuous sign language recognition.

3.2 Integrated Sign Language Recognition and Synthesis Dialogue System

With the development of multimodal human-computer interface technology, communication based on heterogeneous language modes has become possible. Heterogeneous mode communication refers to communication between different languages. Unlike general spoken language communication between different languages, communication between hearing people (specifically referring to hearing individuals with spoken language ability) and deaf people should mainly be conducted through spoken language and sign language. However, most hearing people cannot understand sign language, while most deaf people cannot hear

spoken language. Using computer technology to build a bridge enabling free communication between deaf and hearing people allows direct communication between them with computer assistance. For deaf people, the system translates hearing people' s speech into sign language for display; for hearing people, the system converts deaf people' s sign language into speech.

Based on data glove sign language recognition work, we developed and improved the structural block diagram of the sign language recognition and synthesis system as shown in Figure 1 [Figure 1: see original paper]. The system provides a bidirectional translation (dialogue) system for barrier-free communication between hearing and deaf people—enabling direct communication between hearing people who speak and deaf people who use sign language. Hearing people can speak or express their thoughts in speech form via telephone. This speech first enters the speech recognition module and is recognized as text sentences. Then these texts are synthesized into sign language through the sign language synthesis module for deaf people to watch. Simultaneously, hearing people' s facial images are captured and, through facial expression and lip movement synthesis modules, synthesize human expressions and lip movements, maintaining synchronization between sign language movements and facial expressions. What appears before deaf people is thus a virtual human with rich facial expression changes, using fluent sign language to express thoughts. After deaf people understand these signs, they respond using sign language. These signs enter the sign language recognition system through input devices, allowing the machine to understand the meaning of deaf people' s sentences. These recognized sentences are then synthesized into human speech through speech synthesis modules. Through two agents—speech-to-sign conversion and sign-to-speech conversion—natural communication between deaf and hearing people can be achieved.

The key technology for implementing the sign language recognition and synthesis system lies in automatic sign language recognition. This is an extremely challenging topic and an important component of multimodal human-computer interface technology research. Our system recognized 5,113 Chinese Sign Language words collected from 6 sign language teachers and recognized 1,500 samples of continuous sign language sentences composed of 750 different sentences from this vocabulary set, achieving good recognition performance. Besides automatic sign language recognition technology, other major technologies involved include sign language synthesis, 3D face synthesis, synchronization of lip movements and facial expressions with sign language, speech recognition, and speech synthesis [**?**][**?**]. Integrating these technologies provides a new approach for barrier-free communication between deaf and hearing people.

3.3 Other Related Research

Two approaches for front-end visual processing in vision-based sign language recognition are 2D-based methods and 3D model-based methods [**?**]. However, due to high computational costs or poor 3D model recovery performance, 3D model methods are difficult for back-end high-level recognition applications,

making 2D methods the mainstream for visual pre-processing. Typically, the capture field of view in vision-based sign language recognition is the signer' s upper body, and hand detection usually utilizes color, motion, and/or edge information. Among these, color cues are used in two ways: one directly uses skin color to segment hands, but one problem to solve is how to distinguish them from faces; the other uses colored gloves to assist hand detection. Overall, hand detection/tracking (especially for both hands) is still relatively coarse, currently typically only obtaining bounding boxes of hands. Since our goal is recognition oriented toward medium or even larger vocabularies, we need to obtain as complete target and feature descriptions as possible at the visual front-end. Therefore, we attempted a hybrid solution combining block matching and active contour-based occlusion handling [**?**].

We also conducted research on viewpoint-independent sign language recognition in single-camera environments [**?**]. Since sign language has smaller scales, richer detailed variations, more diverse rotations, and more categories compared to human actions, various viewpoint-independent recognition methods introduced above are not easily directly applicable. 3D methods can be attempted, but reliably recovering 3D data through a single camera is very difficult. Using simple invariants contained in 2D image point sets is insufficient to effectively distinguish numerous sign language words. Viewpoint-independent recognition methods using 2D appearance templates from few viewpoints can typically only be applied to human action recognition with fewer categories, larger scales, and mainly considering only horizontal rotation. Methods based on multi-viewpoint 2D appearance templates can be considered, but due to the large variety of sign language, multiplied by viewpoint variations, the space cost for template storage and time cost for matching are very considerable. Achieving effective viewpoint-independent sign language recognition in single-camera environments requires new recognition ideas and methods. Our idea is that since two sign language sequences of the same sign from different viewpoints can be interpreted as being simultaneously captured by a stereo vision system, viewpoint-independent sign language recognition can be achieved by verifying whether two sequences satisfy epipolar geometry constraints. The advantage of this approach is that it does not require 3D data recovery and can achieve effective viewpoint-independent sign language recognition based on templates from only a few viewpoints. Problems to be solved include how to calculate the fundamental matrix between sequences and how to verify it. The main challenge is that when sequences are not aligned, point correspondences are not easily obtained directly. On a fully automatic dataset (with feature points automatically extracted by computer) containing 100 Chinese Sign Language words collected using self-designed colored gloves, the Sample-Consensus recognition method achieved a Top-3 recognition rate of 53.5%. The decreased recognition rate is mainly due to fewer visible feature points and missed feature point detection caused by viewpoint variation, resulting in insufficient feature points for effective viewpoint-independent recognition analysis. This situation is called data missing. Future work will focus on researching more robust viewpoint-independent sign language recog-

nition methods for data missing situations, while also considering introducing statistical theory to integrate multiple viewpoint templates for more effective viewpoint-independent sign language recognition.

Training reliable and accurate non-specific person models requires collecting sufficient data from enough signers, which is relatively easy to achieve for small vocabulary sets. However, the Chinese Sign Language dictionary defines over 5,000 basic vocabulary words, making it practically impossible to collect sufficient data from enough signers. Therefore, training accurate and reliable non-specific person models is not a good solution for this problem. Another approach is to train specific person models for users, which also encounters problems. First, users need to collect sufficient training data before using the system, which is unacceptable for most users and affects system promotion; second, user data may differ across different stages, making models trained on data from one time period likely unsuitable after some time. Therefore, we propose using learning and adaptation to solve these problems, i.e., adopting adaptive sign language recognition technology [**?**] to address the non-specific person problem in sign language recognition.

4 Conclusion

This paper reviews recent research progress in sign language recognition: introducing some representative work (implemented prototype systems), summarizing main recognition technologies, analyzing remaining challenging problems (also current research hotspots), and proposing some possible countermeasures. It should be noted that as a direct "by-product" of sign language recognition research, it has promoted gesture recognition research and applications to some extent (especially in recent years, as industry continues to follow up). Additionally, this paper briefly introduces our progress in related research work.

References

[1] Gao Wen, Chen Xilin, Ma Jiyong, Wang Zhaoqi. A multimodal interface technology-based communication system for deaf and normal-hearing people. Chinese Journal of Computers, 2000, 23(12): 1253−1260
[2] S. C.W. Ong, S. Ranganath. Automatic sign language analysis: a survey and the future beyond lexical meaning. IEEE TPAMI, 2005, 27(6)
[3] K.H. Jo, Y. Kuno, and Y. Shirai. Manipulative Hand Gesture Recognition Using Task Knowledge for Human Computer Interaction. The Third IEEE International Conference on Automatic Face and Gesture Recognition, Los Alamitos, USA, 1998: 468-473
[4] B. Salem, R. Yates, and R. Saatchi. Current Trends in Multimodal Input Recognition. IEEE Colloquium Virtual Reality: Personal, Mobile and Practical Applications, London, UK, 1998: 1-6
[5] W.T. Freeman and C.D. Weissman. Television control by hand gesture. The International Workshop on Automatic Face-and Gesture-Recognition, Zurich,

Switzerland, 1995: 179-183
[6] C.P. Tung and A.C. Kak. Automatic learning of assembly tasks using a dataglove system. IEEE International Conference on Intelligent Robots and Systems, 1995: 1-8
[7] G.J. Grimes. Digital data entry glove interface device. Technical Report US Patent 4,414,537, Bell Telephone Laboratories, November 1983
[8] H. Brashear, T. Starner, P. Lukowicz, H. Junker. Using multiple sensors for mobile sign language recognition. In: Proc. IEEE International Symposium on Wearable Computers (ISWC), 2003, 45-52
[9] V.R. Culver. A hybrid sign language recognition system. In: Proc. ISWC' 04, 2004, 30-33
[10] T.G. Zimmerman and J. Lanier. Hand gesture interface device. The Human Factors and Computing Systems and Graphics Interface, 1987: 189-192
[11] J. LaViola. A Survey of Hand Posture and Gesture Recognition Techniques and Technology. Technical Report CS-99-11, Brown University, Department of Computer Science, June, 1999
[12] J. Kramer and L. Leifer. The Talking Glove: An Expressive and Receptive 'Verbal' Communication Aid for the Deaf, Deaf-blind, and Non-vocal. Technical Report, Department of Electrical Engineering, Stanford University, 1989
[13] M. Zhao, F.K.H. Quek, and X. Wu. RIEVL: recursive induction learning in hand gesture recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1998, 20(11): 1174-85
[14] S. Gutta, I. Imam, and H. Wechsler. Hand gesture recognition using ensembles of radial basis function (RBF) networks and decision trees. International Journal of Pattern Recognition and Artificial Intelligence. 1997, 11(6): 845-872
[15] J.S. Kim, C.S. Lee, W. Jang, and Z. Bien. Online dynamic hand gesture recognition system for the Korean sign language. Journal of the Korean Institute of Telematics and Electronics. 1997, 34C(2):
[16] T. Takahashi and F. Kishino. A hand gesture recognition method and its application. Systems and Computers in Japan. 1992, 23(3): 38-48
[17] C. Lee and Y. Xu. Online. Interactive Learning of Gestures for Human/Robot Interfaces. IEEE International Conference on Robotics and Automation, Minneapolis, MN, 1996: 2982-2987
[18] S.S. Fels and G.E. Hinton. Glove-talk: a neural network interface between a data-glove and a speech synthesizer. IEEE Trans. Neural Networks. 1993, 4(1): 2-8
[19] S.S. Fels and G.E. Hinton. Glove-TalkII: A neural network interface which maps gestures to parallel formant speech synthesizer controls. IEEE Transactions on Neural Networks. 1998, 9(1): 205-212
[20] J.L. Hernandez-Rebollar, N. Kyriakopoulos, R.W. Lindeman. A New Instrumented Approach For Translating American Sign Language Into Sound And Text. Six IEEE Int' l Conf. Automatic Face and Gesture Recognition (FG' 04), Korean, 2004: 547-552
[21] R.H. Liang and M. Ouhyoung. A real-time continuous gesture recognition system for sign language. The Third International Conference on Automatic Face and Gesture Recognition, Nara, Japan, 1998:

[22] H. Sagawa and M. Takeuchi. A method for recognizing a sequence of sign language words represented in a Japanese sign language sentence. The Fourth International Conference on Automatic Face and Gesture Recognition (FG ' 00), Grenoble, France, 2000: 434-439
[23] H. Sagawa and M. Takeuchi. Development of an information kiosk with a sign language recognition system. International Conference on Universal Usability, ACM Press, 2000: 149–150
[24] H. Sagawa and M. Takeuchi. The Structure and valuation of an Information Kiosk with a sign-Language Recognition System. International Conference on Intelligent User Interfaces, New Mexico, USA, 2001
[25] V.I. Pavlovic, R. Sharma, and T.S. Huang. Visual interpretation of hand gestures for human-computer interaction: a review. IEEE Trans. on Pattern Analysis and Machine Intelligence. 1997, 19(7):
[26] Y. Wu and T.S. Huang. Vision-based gesture recognition: a review. The International Gesture Workshop (GW' 99), Gif-sur-Yvette, France, 1999: 103-115
[27] Charayaphan C, Marble A. Image processing system for interpreting motion in American Sign Language. Journal of Biomedical Engineering, 1992, 14(15): 419-425
[28] Y. Cui and J. Weng. Appearance-Based Hand Sign Recognition from Intensity Image Sequences. Computer Vision and Image Understanding. 2000, 78(2): 157-176
[29] Y. Cui and J. Weng. A Learning-Based Prediction-and-Verification Segmentation Scheme for Hand Sign Image Sequences. IEEE Trans. Pattern Analysis and Machine Intelligence. 1999, 21(8): 798-804
[30] M.H. Yang, N. Ahuja and M. Tabb. Extraction of 2D Motion Trajectories and Its Application to Hand Gesture Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2002, 24(8):
[31] K. Grobel and M. Assan. Isolated Sign Language Recognition using Hidden Markov Models. IEEE International Conference of System, Man and Cybernetics, Orlando, USA, 1997: 162-167
[32] A. Shamaie and A. Sutherland. Accurate Recognition of Large Number of Hand Gestures. The Second Iranian Conference on Machine Vision and Image Processing, Tehran, Iran, 2003
[33] J.W. Deng and H. T. Tsui. A PCA/MD A Scheme for Hand Posture Recognition. The 5th International Conference on Automatic Face and Gesture Recognition, Washington, DC, USA, 2002: 294-299
[34] J.W. Deng and H.T. Tsui. A Two-step Approach based on PaHMM for the Recognition of ASL. The Fifth Asian Conference on Computer Vision, Melbourne, Australia, 2002: 126-131
[35] Starner T, Weaver J, Pentland A. Real-time American Sign Language recognition using desk and wearable computer based video. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(12): 1371-1375
[36] B. Bauer and H. Hienz. Relevant features for video-based continuous sign language recognition. The Fourth International Conference on Automatic Face and Gesture Recognition, Grenoble, France, 2000:

[37] B. Bauer and K.F. Kraiss. Towards an automatic sign language recognition system using subunits. The International Gesture Workshop (GW' 01), London, UK, 2001: 64-75
[38] B. Bauer, K.F. Kraiss. Video-Based Sign Recognition using Self-Organizing Subunits. The 16th International Conference on Pattern Recognition(ICPR), 2002: 434-437
[39] C. Vogler and D. Metaxas. ASL recognition based on a coupling between HMMs and 3D motion analysis. IEEE International Conference on Computer Vision, Mumbai, India, 1998: 363-369
[40] C. Vogler and D. Metaxas. Toward scalability in ASL recognition: breaking down signs into phonemes. The International Gesture Workshop (GW' 99), Gif-sur-Yvette, France, 1999: 400-404
[41] C. Vogler and D. Metaxas. A framework for recognizing the simultaneous aspects of American sign language. Computer Vision and Image Understanding. 2001, 81(3): 358-384
[42] Gao Wen, Wang Shuanglin. Capture and recognition of gestures in complex backgrounds. Pattern Recognition and Artificial Intelligence. 1995, 8: 93-100
[43] Ren Haibing, Zhu Yuanxin, Xu Guangyou, Lin Xueyin, Zhang Xiaoping. Spatiotemporal appearance modeling and recognition of continuous dynamic gestures. Chinese Journal of Computers. 2000, 23(8): 824-828
[44] Zhu Yuanxin, Xu Guangyou, Huang Yu. Appearance-based dynamic isolated gesture recognition. Journal of Software. 2000, 11(1): 54-61
[45] P. Vamplew, A. Adams. Recognition of sign language gestures using neural networks. Australian Journal of Intelligent Information Processing Systems, 1998, 5(2): 94-102
[46] S. Akyol and U. Canzler. An Information Terminal using Vision Based Sign Language Recognition. ITEA Workshop on Virtual Home Environments, 2002: 61-68
[47] S.C.W. Ong and S. Ranganath, "Deciphering gestures with layered meanings and signer adaptation," Proc. Int' l Conf. Automatic Face and Gesture Recognition, 2004: 559-564
[48] U. V. Agris, D. Schneider, J. Zieren, K. F. Kraiss, "Rapid signer adaptation for isolated sign language recognition. In Proc. CVPR' 06 Workshop, 2006: 159
[49] Y. Wu and T. S. Huang," View-independent recognition of hand postures," in Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2000: 88-94
[50] Richard Bowden, David Windridge and et al. , "A Linguistic Feature vector for the Visual Interpretation of sign Language," In Proceedings of The 8th European Conference on Computer Vision, 2004: 390-401
[51] C. Rao, A. Gritai, M. Shah, and T. Syeda-Mahmood, "View invariant alignment and matching of video sequences," In Proceedings of the 9th IEEE International Conference of Computer Vision, 2003: 939–945
[52] W. Gao, J.Y. Ma, J.Q. Wu, and C.L. Wang. Sign language recognition based on HMM/ANN/DP. International Journal of Pattern Recognition and Artificial Intelligence. 2000, 14(5): 587-602
[53] Wu Jiangqin. Research and Practice of Chinese Sign Language Recognition Algorithms. PhD Dissertation, Harbin Institute of Technology. 2000

[54] W. Gao, J.Y. Ma, X.L. Chen et al. HandTalker: a multimodal dialog system using sign language and 3-D virtual human. The 3rd International Conference on Multimodal Interface, Beijing, China, 2000:
[55] Ma Jiyong. Research on Statistical Models for Sign Language Understanding. Postdoctoral Research Report, Institute of Computing Technology, Chinese Academy of Sciences. 2001
[56] Wang Chunli. A Large-Vocabulary Continuous Chinese Sign Language Recognition System. PhD Dissertation, Dalian University of Technology. 2003
[57] W Gao, Y Q Chen, G L Fang, C S Yang, D L Jiang, C B Ge, C L Wang, HandTalker II: A Chinese Sign language Recognition and Synthesis System, The Eighth International Conference on Control, Automation, Robotics and Vision, Kunming, China, Dec.06-09, 2004
[58] Fang Gaolin. Research on Statistical Models for Sign Language Recognition. PhD Dissertation, Harbin Institute of Technology. 2004
[59] Wang Zhaoqi, Gao Wen. A Chinese Sign Language Synthesis Method Based on Virtual Human Synthesis Technology. Journal of Software. 2002, 13(10): 2051-2056
[60] Chen Yiqiang, Gao Wen, Wang Zhaoqi, Jiang Dalong. A Machine Learning-Based Speech-Driven Facial Animation Method. Journal of Software. 2003, 14(2): 215-221
[61] L.-G. Zhang, X. Chen, C. Wang, W. Gao. Robust automatic tracking of skin-colored objects with level set based occlusion handling. In Proc. The 8th International Gesture Workshop (GW' 09), Bielefeld, Germany, Feb 25-27, 2009
[62] Q. Wang, X. Chen, L.-G. Zhang, C. Wang, and W. Gao. Viewpoint invariant sign language recognition. Computer Vision and Image Understanding, 2007, 108: 87-97
[63] Y. Zhou, W. Gao, X. Chen, L.-G. Zhang, C. Wang. Signer Adaptation Based on Etyma for Large Vocabulary Chinese Sign Language Recognition, 8th Pacific-Rim Conference on Multimedia (PCM 2007), Hong Kong, China, Dec.11-14, 2007: 458-461

Author Biographies:

Zhang Liangguo: PhD candidate, Institute of Computing Technology, Chinese Academy of Sciences
Chen Xilin: Professor and PhD supervisor, Institute of Computing Technology, Chinese Academy of Sciences

Note: Figure translations are in progress. See original paper for figures.