Vision-Based Gesture Recognition: A Review

Ying Wu, Thomas S. Huang

Beckman Institute
405 N. Mathews
University of Illinois at Urbana-Champaign
Urbana, IL 61801
{yingwu,huang}@ifp.uiuc.edu

Abstract. The use of gesture as a natural interface serves as a motivating force for research in modeling, analyzing and recognition of gestures. In particular, human computer intelligent interaction needs vision-based gesture recognition, which involves many interdisciplinary studies. A survey on recent vision-based gesture recognition approaches is given in this paper. We shall review methods of static hand posture and temporal gesture recognition. Several application systems of gesture recognition are also described in this paper. We conclude with some thoughts about future research directions.

1 Introduction

The evolution of user interface (UI) witnessed the development from text-based UI based on keyboard to GUI based on mice. What will be the counterpart of mouse when we are trying to explore 3D virtual environments (VEs) in Human Computer Intelligent Interaction (HCII) or Perceptual User Interface (PUI)? In current VE applications, keyboards, mice, wands and joysticks are still the most popular and dominant devices. However, they are inconvenient and unnatural. The use of human movements, especially hand gestures, has become an important part of HCII in recent years, which serves as a motivating force for research in modeling, analyzing and recognition of hand gestures. Many techniques developed in HCII can be extended to other areas such as surveillance, robot control and teleconferencing.

Recognizing gestures is a complex task which involves many aspects such as motion modeling, motion analysis, pattern recognition and machine learning, even psycholinguistic studies. There are already several survey papers in human motion analysis [21, 54] and interpretation [35]. Since gesture recognition is receiving more and more attention in recent research, a comprehensive review on various gesture recognition techniques developed in recent years is needed.

This paper surveys recent studies on vision-based gesture recognition techniques. Section 2 discusses several human gesture representation paradigms in psycholinguistic and cognitive studies, since almost all high-level temporal gesture recognition tasks can be represented by those paradigms which serve as a cognitive model for many complicated temporal hand gestures. Some promising application systems are given in Section 3. Since any recognition method needs feature extraction and data collection, Section 4 discusses the gesture features used in current studies, and Section 5 provides a brief overview of tracking techniques which serves as the data collection process for vision-based gesture recognition.

Since meaningful hand gestures can be classified as static hand postures and temporal gestures, Section 6 and Section 7 discuss various techniques for hand posture recognition and temporal gesture recognition respectively. Especially, recognition by modeling dynamics (Section 7.1), recognition by modeling semantics (Section 7.2), recognition in HMM framework (Section 7.3), and many other techniques (Section 7.4) are given in Section 7.

Since sign language recognition is an important task, Section 8 discusses several studies related to it. Some thoughts about future work and the conclusion of this paper are given in Section 9 and Section 10 respectively.

2 Human Gesture Representation

There have been many studies on human gestures in psycholinguistic research. Stokoe [44] represents gestures as four aspects which are *hand shape*, *position*, *orientation* and *movement*. Kendon [26] describes a philology of gesture, which consists of *gesticulation*, *language-like gestures*, *pantomimes*, *emblems*, and *sign language*. *Sign languages* are characterized by a specific set of vocabulary and grammar. *Emblems* are informal gestural expressions in which the meaning depend on convention, culture and lexicon.

According to different application scenarios, hand gestures can be classified into several categories such as *conversational gestures*, *controlling gestures*, *manipulative gestures*, and *communicative gestures*[54]. Sign language is an important case of *communicative gestures*. Since sign languages are highly structural, they are very suitable for acting as a test-bed for vision algorithms. At the same time, they can also be a good way to help the disabled to interact with computers. *Controlling gestures* are the focus of current research in vision-based interface (VBI). Virtual objects can be located by analyzing pointing gestures. Some display- control applications demonstrate the potential of pointing gestures in HCI. Another controlling gesture is the navigating gesture. Instead of using wands, the orientation of hands can be captured as a 3D directional input to navigate the VEs. The manipulative gesture will serve as a natural way to interact with virtual objects. Tele-operation and virtual assembly are good examples of applications. Communicative gestures are subtle in human interaction, which involves psychological studies, however, vision-based motion capturing techniques can help those studies.

Communicative gestures can be decomposed into three motion phases:*preparation*, *stroke*, and *retraction* [26]. Psycholinguistic studies show that *stroke* may be distinguished from other gesture phases, since *stroke* contains the most information. This model is taken from Quek[39]. He also makes a distinction between *presentation* gestures and *repetitive* gestures.

Bobick [4] emphasizes the dynamical part of gestures. He represents gestures as *movement*, *activity* and *action*. *Movements* are typically atomic and are the most primitive form of motion that can be interpreted semantically. *Activity* is a sequence of either movements or of static configurations. Dynamic models may be used to recognize activities. *Actions* are the high-level entities that people typically use to describe what is happening. Time and context becomes fundamental, though how much one has to reason about context is unclear.

3 Application Systems

There have been many implemented application systems in many domains such as virtual environments, smart surveillance, HCII, teleconferencing, sign language translation, etc..

Zeller et al. [57] present a visual environment for very large scale biomolecular modeling application. This system permits interactive modeling of biopolymers by linking a 3D molecular graphics and molecular dynamics simulation program. Hand gestures serve as the input and controlling device of the virtual environment. Pavlovic and Berry [2] integrate controlling gesture into the virtual environment *BattleField*, in which hand gestures are used not only for navigating the VE, but also as an interactive device to select and move the virtual objects in the *BattleField*. Ju et al. [25] develop an automatic system for analyzing and annotating video sequences of technical talks. Speaker's gestures such as pointing or writing are automatically tracked and recognized to provide a rich annotation of the sequence that can be used to access a condensed version of the talk. Davis and Bobick [17] implement a prototype system for a virtual *Personal Aerobics Trainer* (PAT). Their system allows the user to create and personalize an aerobics session to meet the user's needs and desires. Six stretching and aerobic movements are recognized by the system. Quek [39] presents a *FingerMouse* application to recognize 2-D finger movements which are the input to the desktop. Crowley and Coutaz [11] also develop an application *FingerPaint* to use finger as an input device for augmented reality. Triesch and Maslburg [47] develop a person-independent gesture interface on a real robot which allows the user to give simple commands such as how to grasp an object and where to put it. Imagawa et al. [23] implement a bi-directional translation system between Japanese Sign Language (JSL) and Japanese in order to help the hearing impaired communicate with normal speaking people through sign language.

4 Features for Gesture Recognition

Selecting good features is crucial to gesture recognition, since hand gestures are very rich in shape variation, motion and textures. For static hand posture recognition, although it is possible to recognize hand posture by extracting some geometric features such as fingertips, finger directions and hand contours, such features are not always available and reliable due to self-occlusion and lighting conditions. There are also many other non-geometric features such as color, silhouette and textures, however, they are inadequate in recognition. Since it is not easy to specify features explicitly, the whole image or transformed image is taken as the input and features are selected implicitly and automatically by the recognizer

Cui and Weng [14] investigate the difference between the *most discriminating features (MDF)* and the *most expressive features (MEF)*. MEFs are extracted by K-L projection. However, MEFs may not be the best for classification, because the features that describe some major variations in the class are typically irrelevant to how the subclasses are divided. MDFs are selected by multi-class, multivariate discriminate analysis and have a significantly higher capability to catch major differences between classes. Their experiments also showed that MDFs are superior to the MEFs in automatic feature selection for classification.

Recognizing temporal gestures not only needs spatial features, but also require temporal features. It is possible to recognize some gestures by 2D locations of hands, however, it is not general and view-dependent. The most fundamental feature is the 2D location of the interested blob. Wren et.al [53] use a multi-class statistical model of color and shape to obtain a 2D representation of the head of hand in a wide range of viewing conditions in their tracking system *Pfinder*.

In order to achieve spatial invariant recognition, 3D features are necessary. Campbell et al.[9] investigated the 3D invariant features by comparing the recognition performance on ten different feature vectors derived from a single set of 18 T'ai Chi gestures which are used in the *Staying Alive* application developed by Becker and Pentland [1]. Hidden Markov Model (HMM) is taken as the recognizer. They reported that $(dr, d\theta, dz)$ had the best overall recognition rates. At the same time, their experiments highlight the fact that choosing the right set of features can be crucial to the performance.

Features for temporally invariant gesture recognition is hard to specify since it depends on the temporal representation of gestures. However, it can be handled implicitly in some recognition approaches such as finite state machine and HMM, which will discussed in Section 7.

5 Data Collection for Recognition

To collect data for temporal gesture recognition is not a trivial task. The hand has to be localized in the image sequences and segmented from the background. 2-D tracking supplies the localized information such as hand bounding boxes and centroid of hand blobs. Simple 2-D motion trajectories can be extracted from the image sequences. In some cases, these 2-D features are sufficient for gesture recognition. There has been many 2-D tracking algorithms such as color tracking, motion tracking, template matching, blob tracking, and multiple cues integrating.

Although 2-D tracking gives the position information of hand, some recognition applications still need more features such as hand orientation and hand shape. 3-D tracking approaches try to locate the hand in 3-D space by given the 3-D position and orientation of hand. However, since hand can not be treated as a rigid object, it is very hard to estimate the hand orientation. 3-D position of hand can be achieved by stereo camera or model-based approaches.

Since hand is highly articulated and shape depends on viewpoint, hand shape is hard to describe. Several studies try to recover the *state of the hand* which is represented by the set of joint angles, which is full DOF tracking. If the hand configuration can be estimated, recognizing finger spelling may be easier. However, how to estimate the configuration of articulated objects needs more study.

6 Static Hand Posture Recognition

Since hand postures not only can express some concepts, but also can act as special transition states in temporal gestures, recognizing or estimate hand postures or human postures is one of the main topics in gestures recognition.

Cui and Weng [14] use the most discriminating features to classify hand signs by partitioning the MDF space. A manifold interpolation scheme is introduced to generalize to other variations from a limited number of learned samples. Their algorithm can handle complex background.

Triesch and Malsburg [46] employ the *elastic graph matching* technique to classify hand postures against complex backgrounds. Hand postures are represented by labeled graphs with an underlying two-dimensional topology. Attached to the nodes are *jets*, which is a sort of local image description based on Gabor filters. The recognition rate against complex background is 86.2%. This approach can achieve scale-invariant and user-independent recognition, and it does not need hand segmentation. Since using one graph for one hand posture is insufficient, this approach is not view-independent.

Quek and Zhao [40] introduced an inductive learning system which is able to derive rules of disjunctive normal form formulate. Each DNF describes a hand pose, and each conjunct within the DNF constitutes a single rule. Twenty-eight features such as the area of the bounding box, the compactness of the hand, the normalized moments, served as the input feature vector for their learning algorithm. They obtained 94% recognition rate.

Nolker and Ritter [33] detected the 2D location of fingertips by the *Local Linear Mapping* (LLN) neural network, and those 2D locations are mapped to 3D position by the *Parametric Self-Organizing Map* (PSOM) neural network, since PSOM has the ability to perform an associative completion of fragmentary input. By this means, their approach can recognize hand pose under different views.

7 Temporal Gesture Modeling and Recognition

There are some similarities between temporal gestures and speech so that some techniques in speech such as HMM can be applied to gesture. However, temporal gesture is more complicated than speech. Some low-level movements can be recognized using dynamic models. Some gesture semantics can be exploited to recognize high-level activities. Example-based learning methods can also be used. There are also many other techniques developed in recent years.

7.1 Recognition by Modeling the Dynamics

Modeling the low-level dynamics of human motion is important not only for human tracking, but also for human motion recognition. It serves as a quantitative representation of simple movements so that those simple movements can be recognized in a reduced space by the trajectories of motion parameters. However, those low-level dynamics models are not sufficient to represent more complicated human motions. Some low-level motions can be represented by simple dynamic processes, in which *Kalman filter* is often employed to estimate, interpolate and predict the motion parameters. However, this simple dynamic model is not sufficient to model most cases of human motion, and the Gaussian assumption of the Kalman filtering is usually invalid.

Black and Jepson [3] extended the *Condensation* algorithm to recognize temporal trajectories. Since a *sampling* technique is used to represent the probability density in the *Condensation* algorithm, their approach avoids some difficulties of Kalman filtering. Gesture recognition is achieved by matching of input motion trajectories and model trajectories using *Dynamic Time Warping* (DTW).

Pentland and Liu [36] try to represent human behavior by a complex, multistate model. They used several alternative models to represent human dynamics, one for each class of response. Model switching is based on the observation of the state of the dynamics. This approach produces a generalized maximum likelihood estimate of the current and future values of the state variables. Recognition is achieved by determining which model best fits the observation.

Rittscher and Blake [42] push the technique of combining the idea of model switching and *Condensation*. They use mixed discrete/continuous states to couple perception with classification, in which the continuous variable describes the motion parameters and the discrete variable labels the class of the motion. An ARMA model is used to represent the dynamics. This approach can achieve automatic temporal sequence segmentation.

There is also some work dealing with specific gestures. Cohen et.al [10] use dynamic model to represent circle and line gesture to generate and recognize basic oscillatory gestures such as crane control gestures.

7.2 Recognition by Modeling the Semantics

Many applications need to recognize more complex gestures which include semantic meaning in the movements. Modeling the dynamics alone is not sufficient in such tasks.

The *Finite State Machine* is a usually employed technique to handle this situation. Davis and Shah [19] use this technique to recognize simple hand gestures. Jo, Kuno and Shirai [24] take this approach to recognize manipulative hand gestures such as grasping, holding and extending. The task knowledge is represented by a state transition diagram, in which each state indicates possible gesture states at the next moment. By using a rest state, all unintentional actions can be ignored. Pavlovic and Berry [2] also take this approach.

Another approach is rule-based modeling. Quek [39] uses *extended variable-valued logic* and rule-based induction algorithm to build a inductive learning system to recognize 3-D gestures. Cutler and Turk [15] build a set of simple rules to recognize gestures such as waving, jumping, marching etc.

Pinhanez and Bobick [37] develop a new representation for temporal gestures, a 3-valued domain {past, now, fut}(PNF) network. The occurrence of an action is computed by minimizing the domain of its PNF-network, under constraints imposed by the current state of the sensors and the previous states of the network.

Another promising approach to modeling the semantics of temporal gestures is the *Bayesian Network* and the *Dynamic Bayesian Network*. Pavlovic [34] push this idea forward recently.

7.3 Gesture Recognition in the HMM Framework

HMM is a type of statistical model. A HMM λ consists of N states and a transition matrix. Each state has assigned an output probability distribution function $b_i(O)$, which gives the probability of the state S_i generating observation O under the condition that the system is in S_i. There are three basic problems in HMMs. The first problem is evaluation $P(O|\lambda)$, which can be solved by forward-backward algorithm. The second problem is to find the most likely state sequence S, given an observation and a HMM model, i.e. $maxP(S|O, \lambda)$. The Viterbi algorithm is used to solve it. The third problem is to train the HMM. Baum-Welch algorithm is used to solve it.

Pentland and Liu [36] use HMM to model the state transitions among a set of dynamic models. Bregler [8] takes the same approach. HMM has the capacity for not only modeling the low-level dynamics, but also the semantics in some gestures. Stoll and Ohya [45] employ HMM to model semantically meaningful human movements, in which one HMM is learned for each motion class. The data used for modeling the human motions is an approximate pose derived from an image sequence. Nam and Wohn [32] present a HMM-based method to recognize some controlling gestures. Their approach takes into account not only hand movement, but also hand postures and palm orientations.

There are also many variations of HMM. Yang et al. [55] model the gesture by employing a multi-dimensional HMM, which contains more than one observation symbol at each time. Their approach is able to model multi-path gestures and provides a means to integrate multiple modality to increase the recognition rate.

Since the output probability of feature vectors of each state in HMM is unique, HMM can handle only piecewise stationary processes which are not adequate in gesture modeling. Kobayashi and Haruyama [28] introduce *Partly-Hidden Markov Model* (PHMM) for temporal matching. Darrell and Pentland

[16] introduce a hidden-state reinforcement learning paradigm based on the *Partially Observable Markov Decision Process* to gesture recognition by which an active camera is guided.

When Markov condition is violated, conventional HMMs fails. HMMs are ill-suited to system that have compositional states. Brand et.al. [7] presented an algorithm for coupling and training HMMs to model interactions between processes that may have different state structures and degrees of influence on each other. These problems often occur in vision, speech, or both–coupled HMMs are well suited to applications requiring sensor fusion across modalities.

Wilson and Bobick [52] extended the standard HMM method to include a global parametric variation in the output probabilities of the HMM to handle parameterized movements such as musical conducting and driving by EM algorithm. They presented results on two different movements – a size gesture and a point gesture – and show robustness with respect to noise in the input features.

7.4 Other techniques

There are also many statistical learning techniques applied to gesture recognition. As we describe before, Cui and Weng [12] use the multiclass, multidimensional discriminant analysis to automatically select the most discriminating features for gesture recognition. Polana and Nelson [38] attempt to recognize motion by low-level statistical features of image motion information. A simple nearest centroid algorithm serves as the classifier. Their experiments show their approach is suitable for repetitive gesture recognition. Watanabe and Yachida [51] introduce an eigenspace which is constructed from multi input image sequences to recognize gestures. Since this eigenspace represents the approximate 3-D information for gestures, their approach can handle self-occlusion.

Bobick and Ivanov [4] model the low-level temporal behaviors by HMM techniques. The outputs of HMM serve as the input stream of a stochastic context-free grammar parsing system. The grammar and parser provide longer range temporal constraints. The uncertainty of low level movement detection is disambiguated in the high level parser which include a priori knowledge about the structure of temporal actions.

Yang and Ahuja [56] use *Time-Delay Neural Networks* (TDNN) to classify motion patterns. TDNN is trained with a database of more than ASL signs. The input of the TDNN is the motion trajectories extracted by multi-scale motion segmentation.

8 Sign Language Recognition

Unlike general gestures, sign languages are highly structured so that it provides an appealing test bed for understanding more general principles. However, there are no clear boundaries between individual signs, recognition of sign languages are still very difficult. Speech recognition and sign language recognition are parallels. Both are time-varying processes, which show statistical variations, making

HMMs a plausible choice for modeling the processes. And both must devise ways to cope with context and co-articulation effects. HMMs provide a framework for capturing the statistical variations in both position and duration of the movement. In addition, it can segment the gesture stream implicitly.

There are two kinds of gestures to be recognized, one is isolated gesture, and the other is continuous gesture. The presence of silence makes the boundaries of isolated gestures easy to spot. Each sign can be extracted and presented to the trained HMMs individually. Continuous sign recognition, on the other hand, is much harder since there is no silence between the signs. Here HMMs offer the compelling advantage of being able to segment the streams of signs automatically with the Viterbi algorithm. Co-articulation is difficult to handle in continuous recognition, since it results in the insertion of an extra movement between the two signs.

Starner et al.[43] employ HMM to recognize American Sign Language (ASL). They assume that detailed information about hand shape is not necessary for humans to interpret sign language, so a coarse tracking system is used in their studies.

There are several possible approaches to deal with the co-articulation problem. One is to use context-dependent HMMs, and the other is modeling the co-articulation. The idea of context- dependent HMMs is to train bi-sign or even tri-sign context dependent HMMs. However, this method can not work well. Vogler and Metaxas [49, 50] study the co-articulation in sign language recognition. They propose an unsupervised clustering scheme to obtain the necessary classes of “phonemes” for modeling the movements between signs. Recently, they use phonemes instead of whole signs as the basic units so that the ASL signs can be broken into phonemes such as movements and holds, and HMMs are trained to recognize the phonemes [50]. Since the number of phonemes is limited, it is possible to use HMMs to recognize large-scale vocabularies.

Liang and Ouhyoung [30] also take the HMM approach to the recognition of continuous Taiwanese Sign Language with a vocabulary of 250 signs. The temporal segmentation is performed explicitly based on the discontinuity of the movements according to 4 gesture parameters such as posture, position, orientation and motion.

9 Future Directions

Current static hand posture recognition techniques seldom try to achieve rotation-invariant and view-independent recognition. One approach is to extract some 3-D features or to estimate the hand configuration. Another approach is based on learning. These two approaches need more investigation in hand gestures.

The representation for temporal gesture is crucial to recognition. In low-level movement recognition and tracking, automatic switching among different motion models should be considered more in future studies. Most current gesture applications only look into symbolic gesture commands. Automatic segmentation of temporal gestures plays an important role in extracting or segmenting these gesture commands in continuous movements. However, it is still an open problem and it should receive more attention. Although HMM can handle segmentation in some cases, it may fail in the presence of co-articulation. Two-handed gestures not only make the tracking more difficult, but also make the interpretation of gesture harder. These problems should be investigated in future research. Since speech and gestures are coupled, it is natural to consider combing gesture and speech to a multi-modality system.

10 Conclusion

In this paper, we report the recent development on the research of hand gesture recognition with focus on various recognition techniques. Feature selection, which can be specified explicitly or implicitly by the recognizer, is crucial to the recognition algorithms. Data collection for visual gesture learning is not a trivial task. Various algorithms on static hand posture recognition and temporal gesture recognition are surveyed in this paper. HMM and its variants can be used in sign language recognition. Due to the complexity of gesture, machine leaning techniques seems promising in this task.

Overall, gesture recognition is still in its infancy. It involves the cooperation of many disciplines. In order to understand hand gestures, not only for machines, but also for humans, substantial research efforts in computer vision, machine learning and psycholinguistics will be needed.

Acknowledgements

This work was supported in part by National Science Foundation Grants CDA-96-24396 and IRI- 96-34618.

References

1. Becker,D.: Sensei: A Real-Time Recognition, Feedback and Training System for Tai Chi Gestures, *MIT Media Lab, MS thesis* (1997)
2. Berry,G.: Small-wall: A Multimodal Human Computer Intelligent Interaction Test Bed with Applications, *Dept. of ECE, University of Illinois at Urbana-Champaign, MS thesis* (1998)
3. Black,M., Jepson,A.: Recognition Temporal Trajectories using the Condensation Algorithm, *Int'l Conf. on Automatic Face and Gesture Recognition*, Japan, pp.16-21 (1998)
4. Bobick,A., Ivanov,Y.: Action Recognition using Probabilistic Parsing, *IEEE Int'l Conf. on Computer Vision and Pattern Recognition* (1998)
5. Bobick, A., Wilson,A.: A State-Based Approach to the Representation and Recognition of Gesture, *IEEE trans. PAMI*, Vol.19, No.12, Dec., pp1325-1337 (1997)
6. Bradski,G., Yeo,B., Yeung,M.: Gesture and Speech for Video Content Navigation, *Proc. Workshop on Perceptual User Interfaces* (1998)
7. Brand,M., Oliver,N., Pentland,A.: Coupled Hidden Markov Models for Complex Action Recognition, *Proc. IEEE Int'l Conf. on Computer Vision and Pattern Recognition* (1997)
8. Bregler,C.: Learning and Recognizing Human Dynamics in Video Sequences, *Proc. IEEE Int'l Conf. on Computer Vision and Pattern Recognition* (1997)
9. Campbell,L., et al.: Invariant Features for 3-D Gesture Recognition, *Int'l Conf. on Automatic Face and Gesture Recognition*, Killington, pp.157-162. (1996)
10. Cohen,C., Conway,L., Koditschek,D.: Dynamical System Representation, Generation, and Recognition of Basic Oscillatory Motion Gestures, *Int'l Conf. on Automatic Face and Gesture Recognition* , Killington (1996)

11. Crowley,J., Berard,F., Coutaz,J.: Finger Tracking as An Input Device for Augmented Reality, *Int.Workshop on Automatic Face and Gesture Recognition*, Zurich, pp.195-200. (1995)
12. Cui,Y, Weng,J.: Hand Sign Recognition from Intensity Image Sequences with Complex Background, *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp.88-93. (1996)
13. Cui,Y., Weng,J.: Hand Segmentation Using Learning-Based Prediction and Verification for Hand Sign Recognition, *Int'l Conf. on Automatic Face and Gesture Recognition* , Killington (1996)
14. Cui,Y., Swets,D., Weng,J.: Learning-Based Hand Sign Recognition Using SHOSLIF-M, *Int. Workshop on Automatic Face and Gesture Recognition*, Zurich, pp.201-206. (1995)
15. Cutler,R., Turk,M.: View-based Interpretation of Real-Time Optical Flow for Gesture Recognition, *IEEE Int. Conf. on Automatic Face and Gesture Recognition*, Japan. (1998)
16. Darrell,T., Pentland,A.: Active Gesture Recognition Using Partially Observable Markov Decision Processes, *IEEE Int'l Conf. on Pattern Recognition* (1996)
17. Davis,J., Bobick,A.: Virtual PAT: A Virtual Personal Aerobic Trainer, *Proc. Workshop on Perceptual User Interfaces*, pp.13-18. (1998)
18. Davis, J., Bobick, A.: The Representation and Recognition of Action Using Temporal Templates, *IEEE CVPR*, pp.928-934. (1997)
19. Davis, J., Shah, M.: Visual Gesture Recognition, *Vision, Image and Signal Processing*, 141(2), pp.101-106. (1994)
20. Fernandez, R.: Stochastic Modeling of Physiological Signals with Hidden Markov Models: A Step Toward Frustration Detection in Human-Computer Interfaces, *MIT Media Lab, MS thesis.* (1997)
21. Gavrila, D.: The Visual Analysis of Human Movement: A Survey, *Computer Vision and Image Understanding*, Vol.73, No.1, Jan, pp.82-98. (1999)
22. Goncalves, L., Bernardo, E., Perona, P.: Reach Out and Touch Space, *IEEE Int. Conf. on Automatic Face and Gesture Recognition*, Japan.(1998)
23. Imagawa, K., Lu, S., Igi, S.: Color-Based Hand Tracking System for Sign Language Recognition, *IEEE Int. Conf. on Automatic Face and Gesture Recognition*, Japan. (1998)
24. Jo, K., Kuno, Y., Shirai, Y.: Manipulative Hand Gestures Recognition Using Task Knowledge for Human Computer Interaction, *IEEE Int. Conf. on Automatic Face and Gesture Recognition*, Japan. (1998)
25. Ju, S., Black, M., Minneman, S., Kimber, D.: Analysis of Gesture and Action in Technical Talks for Video Indexing, *IEEE Conf. on Computer Vision and Pattern Recognition, CVPR97* . (1997)
26. Kendon, A.: urrent Issues in the Study of Gesture *The Biological Foundation of Gestures: Motor and Semiotic Aspects*, pp.23-47, Lawrence Erlbaum Associate, Hillsdale, NJ, (1986)
27. Kjeldsen, R., Kender, J.: Interaction with On-Screen Objects using Visual Gesture Recognition, *Proc. IEEE CVPR97*, (1997)
28. Kobayashi, T., Haruyama,S.: Partly-Hidden Markov Model and Its Application to Gesture Recognition, *IEEE Proceedings of ICASSP97*, Vol. VI, pp.3081-84. (1997)
29. Kurita, T., Hayamizu, S.: Gesture Recognition using HLAC Features of PARCOR Images and HMM based Recognizer, *IEEE Int. Conf. on Automatic Face and Gesture Recognition* , Japan. (1998)
30. Liang, R., Ouhyoung, M.: A Real-time Continuous Gesture Recognition System for Sign Language, *IEEE Int. Conf. on Automatic Face and Gesture Recognition*, Japan. (1998)
31. McNeil, D.: Hand and Mind, University of Chicago Press, Chicago. (1992)
32. Nam, Y., Wohn, K.: Recognition of Space-Time Hand-Gestures using Hidden Markov Mdel, *ACM Symposium on Virtual Reality Software and Technology*, HongKong, pp. 51-58. (1996)
33. Nolker, C., Ritter, H.: Illumination Independent Recognition of Deictic Arm Postures, *Proc. 24th Annual Conf. of the IEEE Industrial Electronics Society*, Germany, pp. 2006- 2011. (1998)
34. Pavlovic,V.: Dynamic Bayesian Networks for Information Fusion with Applications to Human–Computer Interfaces, *Dept. of ECE, University of Illinois at Urbana-Champaign, Ph.D. Dissertation*, (1999)
35. Pavlovic, V., Sharma, R., Huang, T.: Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review, *IEEE trans. PAMI*, Vol.19, No.7, July, pp677-695, (1997)
36. Pentland, A., Liu, A.: Modeling and Prediction of Human Behavior, *IEEE Intelligent Vehicles*, (1995)
37. Pinhanez, C. Bobick, A.: Human Action Detection Using PNF Propagation of Temporal Constraints, *IEEE ICCV*, (1998)
38. Polana, R. Nelson, R.: Low Level Recognition of Human Motion, *IEEE Workshop on Motion of Non-Rigid and Articulated Objects*, Austin, pp77-82. (1994)
39. Quek, F.: Unencumbered Gestural Interaction, *IEEE Multimedia*, Vol.3, No.4, pp.36-47, (1997)
40. Quek, F., Zhao, M.: Inductive Learning in Hand Pose Recognition, *IEEE Automatic Face and Gesture Recognition*, (1996)
41. Rohr, K.: Towards Model-Based Recognition of Human Movements in Image Sequences, *CVGIP:Image Understanding*, Vol.59, No.1, Jan, pp.94-115, (1994)

42. Rittscher, J., Blake, A.: Classification of Human Body Motion, *IEEE Int'l Conf. on Computer Vision*, (1999)
43. Starner, T., Weaver, J., Pentland, A.: Real-Time American Sign Language Recognition Using Desk and Wearable Computer Based Video, *IEEE trans. PAMI*, (1998)
44. Stokoe, W.: Sign Language Structure, University of Buffalo Press, (1960)
45. Stoll, P., Ohya, J.: Applications of HMM Modeling to Recognizing Human Gestures in Image Sequences for a Man-Machine Interface, *IEEE Intl Workshop on Robot and Human Communication*, (1995)
46. Triesch, J., Malsburg, C.: Robust Classification of Hand Postures Against Complex Background, *Intl Conf. On Automatic Face and Gesture Recognition*, (1996)
47. Triesch, J., Malsburg, C.: A Gesture Interface for Human-Robot-Interaction, *Intl Conf. On Automatic Face and Gesture Recognition*, (1998)
48. Utsumi, A., Miyasato, T., Kishino, F., Nakatsu, R.: Hand Gesture Recognition System Using Multiple Cameras, *IEEE ICPR*, (1996)
49. Vogler, C., Metaxas, D.: ASL Recognition Based on A Coupling Between HMMs and 3D Motion Analysis, *IEEE ICCV*, (1998)
50. Vogler, C., Metaxas, D.: Toward Scalability in ASL Recognition: Breaking Down Signs into Phonemes, *IEEE Gesture Workshop*, (1999)
51. Watanabe, T., Yachida, M.: Real Time Gesture Recognition Using Eigenspace from Multi Input Image Sequences, *Intl Conf. On Automatic Face and Gesture Recognition* , Japan.(1998)
52. Wilson, A., Bobick, A.: Recognition and Interpretation of Parametric Gesture, *IEEE Intl Conf. Computer Vision*, (1998)
53. Wren, C., Pentland, A.: Dynamic Modeling of Human Motion, *IEEE Intl Conf. Automatic Face and Gesture Recognition*, (1997)
54. Wu, Y., Huang, T.: Human Hand Modeling, Analysis and Animation in the Context of HCI, *IEEE Intl Conf. Image Processing*, (1999)
55. Yang, J., Xu, Y., Chen, C.: Gesture Interface: Modeling and Learning, *Proc. IEEE Int. Conf. on Robotics and Automation*, Vol. 2, pp.1747-1752. (1994)
56. Yang, M., Ahuja, N.: Extraction and Classification of Visual Motion Patterns for Hand Gesture Recognition, *IEEE Int'l Conf. on Computer Vision and Pattern Recognition*, (1998)
57. Zeller, M., et al.: A Visual Computing Environment for Very Large Scale Biomolecular Modeling, *Proc. IEEE Int. Conf. on Application-specific Systems, Architectures and Processors (ASAP)*, Zurich, pp. 3-12. (1997)