Vision-Based Hand Gesture Recognition for Human Computer Interaction: A Review

Reza Hassanpour[1,2] Stephan Wong[1] Asadollah Shahbahrami[1,3]

[1]Computer Engineering Lab, Delft University of Technology, the Netherlands.
[2]Department of Computer Engineering, Cankaya University, Ankara-Turkey.
[3]Department of Computer Engineering, University of Guilan, Rasht, Iran.

reza@dutepp0.et.tudelft.nl {J.S.S.M.Wong,A.Shahbahrami@tudelft.nl}

ABSTRACT

Evolution of user interfaces shapes the change in the human-computer interaction. With the rapid emergence of three-dimensional (3-D) applications; the need for a new type of interaction device arises as traditional devices such as mouse, keyboard, and joystick become inefficient and cumbersome within these virtual environments. Intuitive and naturalness characteristics of 'Ha nd Gestures" in human computer interaction have been the driving force and motivation to develop an interaction device which can replace current unwieldy tools. This study is a survey on the methods of analyzing, modeling and recognizing hand gestures in the context of human-computer interaction. Taxonomy of the methods based on the applications that they have been developed for and the approaches that they have used to represent gestures is presented. Direction of future developments is also discussed.

KEYWORDS

Gesture Recognition, Computer Vision, Human Computer Interaction.

1. INTRODUCTION

Evolution of user interfaces shapes the change in the human-computer interaction devices. One of the most common human-computer interaction devices is the keyboard which has been the ideal choice for text-based user interfaces. Graphical user interfaces brought mouse into the desktops of the users. As three-dimensional (3-D) applications take place the need for a new type of interaction device arises since traditional devices such as mouse, keyboard, joystick, etc become inefficient for interaction within these virtual environments. A better interaction in virtual environments requires a natural and suitable device. "Hand Gesture" concept in human-computer interfacing context which has become popular in recent years can be used to develop such an interaction device. Human hand gestures are a set of movements of the hand and arm which range from the simple action of pointing at something to the complex ones used to communicate with other people. Understanding and interpreting these movements requires modeling them in both spatial and temporal domains. Static configuration of the human hand which is called hand posture and its dynamic activities are vital for human-compute interaction. The procedures and techniques used for acquiring theses configurations and behaviors are among the most determining traits for classifying on-going researches. This survey starts with the most common definitions of hand gesture and its taxonomy. The classification of the methods using major works in each group is presented. We consider the analysis and modeling techniques from computer vision and human computer interaction points of view. The future trends and research directions are also given.

1.1. Gesture Modeling

Hand gestures are motions of human hand(s) and arm(s) which are used as a means to express or emphasize an idea or convey a manipulative command to control an action. This definition does not include unintentional movements. However, it expresses the common feature of all hand gestures which is a mapping from hand motion space to the mental concepts. This mapping is performed through an observer's visual system which can detect and temporally track the movement. The temporal modeling of a gesture is important since human gestures are dynamic processes. Psychological studies show that a hands gesture consists of three phases. These phases are: Preparation, Nucleus, and Retraction.

The preparatory phase is to bring the hand from its resting state to the starting posture of the gesture. This phase sometimes is very short and sometimes it is combined with the retraction phase of the previous gesture. The nucleus contains the main concept and has a definite form. The retraction phase shows the resting movement of the hand after completing the gesture. Retraction may be very short or not present if the gesture is succeeded by another gesture. The preparatory and retraction phases are generally short and the hand movements are faster compared to the nucleus phase. However, identifying the starting and ending points of the nucleus phase is one of the complexities stemming from the temporal variability of the hand gestures in general and preparatory and retraction phases in particular. The above mentioned uncertainty in the temporal movement of the hand and arm during a gesture together with the differences in the shape of the hand and the way each individual person performs a specific gesture, show that parametric stochastic models are more suitable for gesture recognition systems. Different types of physical or appearance features may be used to model a hand gesture. Each posture assumed by the hand during a gesture movement defines a point in the parameter space of the hand model. A posture in this case specifies a sub-space of the hand parameter space that is given by the distribution of the posture parameter values. A gesture is given by a trajectory in this parameter space. The mathematical representation of a gesture recognition system is a mapping from the hand movement space to a trajectory in parameter space can be given as $G=T_pM$, where G is a trajectory in the parameters space, M is the hand movement and T_p is the transform mapping movement to trajectory using the parameters P .

1.2. Gesture Taxonomy

Several classifications have been considered for hand gestures in the literature. One taxonomy which is more suitable for human computer interaction applications divides hand gestures into three groups [38]. These groups are: communicative gestures, manipulative gestures, and controlling gestures. Communicative gestures are intended to express an idea or a concept. These gestures are either used together with speeches or are a substitute for verbal communications which on the other hand requires a high structured set of gestures such as those defined in sign languages [34],[37]. Manipulative gestures are used for interaction with objects in an environment. These gestures are mostly used for interaction in virtual environments such as teleoperation or virtual assembly systems however; physical objects can be manipulated through gesture controlled robots. Controlling gestures are the group of gestures which are used to control a system or point and locate and object. FingerMouse [28] is a sample application which detects 2D finger movements and controls mouse movements on the computer desktop. Analyzing hand gestures is completely application dependant and involves analyzing the hand motion, modeling hand and arm, mapping the motion features to the model and interpreting the gesture in a time interval.

1.3. Hand Modeling

Understanding and interpreting hand gestures involve determining the posture of the hand and arm during a gesture period. This process might be highly complicated considering the articulated structure of the human hand. However, many physiological constraints are also available in human hand which makes its modeling difficult. Depending on the application, different types of model-based solutions for hand gesture recognition systems have been proposed. A typical vision based hand gesture recognition system consists of a camera(s), a feature extraction module, a gesture classification module and a set of gesture models. In the feature extraction process the necessary features are extracted from the captured frames of the camera(s). These features can be divided into three sub categories:

1. High-level features, generally based on three dimensional models,
2. The image itself as a feature used by view-based approaches,
3. Low-level features measured from the image.

High-level features can be inferred from the joint angles and pose of the palm. For this feature set, generally the anatomic structure of the hand is used as a reference. For precision purposes colorful gloves can be used. View-based approaches are alternatives to the high-level modeling and they model the hands as a set of two dimensional intensity images. Low-level features are based on the thought that the full reconstruction of the hand is not essential for gesture recognition. Therefore, only some cues like the centroid of the hand region, the principle axes defining an elliptical bounding region of the hand, the optical flow/affine flow of the hand region in a scene, etc can be chosen as features. One of the most popular areas is recognition of a local sign language. Liang *et al.* [18] worked on the Taiwanese sign language, Starner *et al.* [34] worked on the American Sign Language and Haberdar [14] studies the Turkish Sign Language for his thesis work. Similarly Gejgus *et al.* [11] worked on the finger alphabet. The general purpose of these applications is either helping the deaf-dumb people at their communication with others or completely translating from a sign language into a normal one. Another type of application about sign languages is human-computer interaction, in other words, using sign languages as input, the information conveyed by the gestures is transferred to the computer via camera(s). Eisenstein and Davis[9] controlled a display in their application. Bretzner [5] developed a prototype system, where the user can control a TV set and a lamp. Robot control is the aim of the works of Ren *et al.* [29], Malima *et al.* [20], Ludovic *et al.* [19], Agrawal and Chaudhuri [1], Starner *et al.* [34], Postigo *et al.* [27]. Malima *et al.* [20], propose an algorithm for automatically recognizing a limited set of gestures from hand images for a robot control application. The algorithm enables the robot to identify a hand pose sign in the input image, as one of five possible commands . The identified command is then used as a controller input for the robot to perform a certain task. Application of Kolsch and Hollerer [15] helps people with a wearable device. Fujisawa *et al*. [10] developed an HID device as an alternative for mouse for physically handicapped persons. Human-Building interaction is considered at Malkawi and Srinivasan [21]. Marschall [24] has an interesting application which provides a visual sculpture. Pedestrian tracking from a moving vehicle is the primary goal of Philomin *et al*. [26]. Mantyla *et al*. [22] developed a system for mobile device users. While some of the works mentioned above uses a complete sign language, some of them uses just a part of a sign language or develops an application-specific sign language for human-computer interaction. The detailed discussion of the hand modeling methods is given in section 2.

1.4. Modeling Shape

The hand gesture is an intentional and meaningful movement of the hand and arm in space therefore it seems necessary to define a spatial model for representing this movement especially when delicate hand movements are to be interpreted by the interfacing computer. Hand shape models can be classified into two groups: Volumetric models and Skeletal models. Volumetric models are used to describe the appearance and shape of the hand. These models are commonly used in computer graphics applications but appearance based gesture extraction systems are also using them. Skeletal models on the other hand are interested in the joint parameter values and represent a hand posture using a set of these values. Huang *et al*. in [38] present a model based tracking system. To work with a diverse group of people, they use a generic model which is sufficiently flexible. To fit the model to an arbitrary hand as well as having accurate surface characteristic, they use cubic B-Splines to represent the surfaces of palm, fingers and the thumb. The model used in their implementation contains 300 control points. The model also includes a total of 23 degrees of freedom based on the anatomical analysis of the hand. Huang *et al*. consider two sets of constraints. Group one includes static constraints such as joint length and finger MCP flexion convergence angle. These constraints are set interactively by the user. Dynamic constraints have to be updated every time a joint is moved. An example of such a constraint is the reduction of the ability to abduct or adduct when the fingers flex downwards. Calibration of the model to a real hand is done visually. Four interactive sessions have been considered. Each section is followed by an automatic fitting stage which accounts for the smooth contours which make up the surface of the final hand model.

2. Classification of the Methods

Hand gesture recognition is a relatively new field for the computer science. Applications for hand gesture recognition in machine learning systems have been developed approximately for 20 years. The methods used in these systems can be categorized into two groups. Generally the earlier systems like Liang *et al*. [19] used gloves for gesture recognition. Their method was unpractical since the gloves were limiting moving abilities of the user. Recent studies like Malima *et al*. [20] concentrated on vision based systems since they provide relatively cost-effective methods to acquire and interpret human hand gestures while being minimally obtrusive to the participant. This survey considers the vision based hand gesture recognition systems only.

2.1. Hand Modeling with High-Level Features

High-level features are extracted by model-based approaches. A typical model-based approach may create a 3D model of a hand by using some kinematics parameters and projecting its edges onto a 2D space. Estimating the hand pose which in this case is reduced to the estimation of the kinematics parameters of the model is accomplished by a search in the parameters space for the best match between projected edges and the edges acquired from the input image. Ueda *et al.* [35] uses a method that estimates all joint angles to manipulate an object in the virtual space. In their method, the hand regions are extracted from multiple images obtained by the multi-viewpoint camera system. By integrating these multi-viewpoint silhouette images, a hand pose is reconstructed as a "voxel model". Then all joint angles are estimated using three dimensional matching between hand model and voxel model. They performed an experiment in which the joint angles were estimated from the silhouette images by the hand-pose simulator. Utsumi *et al.* [36] used multi-viewpoint images to control objects in the virtual world. Eight kinds of commands are recognized based on the shape and movement of the hands. Bray *et al.* [4] proposed a tracker based on 'Stochastic Meta-Descent' for optimizations in such high dimensional state spaces. The algorithm is based on a gradient descent approach with adaptive and parameter-specific step sizes. The Stochastic Meta-Descent tracker facilitates the integration of constraints, and combined with a stochastic sampling technique, can get out of spurious local minima. Furthermore, the integration of a deformable hand model based on linear blend skinning and anthropometrical measurements reinforce the robustness of the tracker. Bettio *et al.* [2] presented a practical approach for developing interactive environments that allows humans to interact with large complex 3D models without having them to manually operate input devices. The system provides support for scene manipulation based on hand tracking and gesture recognition and for direct 3D interaction with the 3D models in the display space if a suitably registered 3D display is used. Being based on markerless tracking of a user's two hands, the system does not require users to wear any input or output devices. In model based approaches the initial parameters have to be close to the solution at each frame and noise is a real problem for fitting process. Another problem is the textureless nature of the human hand which makes it difficult to detect the inner edges of the hand. Davis and Shah [7], Dorner [8] and Lee and Kunii [17] used a glove with markers in order to make the feature extraction process easier. Similarly manual parameter instantiation or placing user hands in a specific position were also used for the ease of initialization process. Processes on these features may be relatively slower than the other feature approaches due to the 3D structure complexity of high-level features.

2.2. View-based Approaches

These approaches are also called appearance-based approaches. These approaches model the hand by a collection of 2D intensity images. At the same time, gestures are modeled as a sequence of views. Eigenspace approaches are used within the view-based approaches. They provide an efficient representation of a large set of high dimensional points using a small set of orthogonal basis vectors. These basis vectors span a subspace of the training set called the eigenspace and a linear combination of these images can be used to approximately reconstruct any of the training images. These approaches were used in many of the face recognition approaches. Their success in face recognition made them attractive for other recognition applications like hand gesture recognition. (e.g. Gupta *et al.* [13] and Black [3]) Black [3] demonstrated their approach by tracking four hand gestures with 25 basis images and provided three major improvements to the original eigenspace approach formulation: A large invariance to occlusions Some invariance to differences in background from the input images and the training images The ability to handle both small and large affine transformations of the input image with respect to the training images Zahedi *et al.* [40] showed how appearance-based features can be used for the recognition of words in American Sign Language from a video stream. The features are extracted without any segmentation or tracking of the hands or head. Experiments are performed on a database that consists of 10 words in American Sign Language with 110 utterances in total. The video streams of two stationary cameras are used for classification. Hidden Markov Models (HMM) and the leaving one out method are employed for training and classification. Using the appearance-based features, they achieved an error rate of 7%. About half of the remaining errors are due to words that are visually different from all other utterances. Although these approaches may be sufficient for a small set of gestures, with a large gesture space collecting adequate training sets may be problematic. Another problem is the loss of compactness in the subspace required for efficient processing.

2.3. Low-Level Features

Starner *et al.* [34] noticed that prior systems could recover relatively detailed models of the hands from video images when given some constraints. However, many of those constraints conflicted with recognizing American Sign Language in a natural context, either by requiring simple, unchanging backgrounds; not allowing occlusion; requiring carefully labeled gloves; or being difficult to run in real time. Therefore they presented such a new and relatively simple feature space that assumes detailed information about hand shape is not necessary for humans to interpret sign language. They found that all human hands have approximately the same hue and saturation, and vary primarily in their brightness. By using this color cue they used the low-level features of hand's x and y position, angle of axis of least inertia, and eccentricity of the bounding ellipse. This feature set is one of the first low-level features in the literature for hand gesture concept of computer vision. They combined the low-level feature set by HMM network and achieved the accuracy of 97% per word on a 40 word lexicon. Gknar and Yldrm [12] presented a hand gesture recognition system using an inexpensive camera with fast computation time. They used skin tone density and eccentricity of the bounding ellipse low-level features and Multilayer Perceptron and Radial Basis Function neural networks for classification. They achieved the success of 78.3% on 3 layered structures and 80% for 4 layered structures. Lee [16] used low-level feature, the distance from the centroid of the hand region to the contour boundary. The method obtains the image through subtract one image from another sequential image, measuring the entropy, separating hand region from images, tracking the hand region and recognizing hand gestures. Through entropy measurement, they have got color information that have near distribution in complexion for region that have big value and extracted hand region from input images. They could draw hand region adaptively in change of lighting or individual's difference because entropy offer color information as well as motion information at the same time. Detected contour using chain code for hand region that is extracted, and present centroidal profile method that is improved little more and recognized gesture of hand. In the experimental results for 6 kinds of hand gesture, the recognition rate was found more than 95%. Malima *et al.* [20] proposed a fast algorithm for automatically recognizing a limited set of gestures from hand images for a robot control application. They considered a fixed set of manual commands and a reasonably structured environment, and developed a procedure for gesture recognition. The algorithm is invariant to translation, rotation, and scale of the hand. The low-level feature used in the algorithm is the center of the gravity and the distance from the most extreme point in the hand to the center which is the farthest distance from centroid to tip of the longest active finger in the particular gesture. Yang [39] presented an algorithm for extracting and classifying two-dimensional motion in an image sequence based on motion trajectories. First, a multi-scale segmentation is performed to generate homogeneous regions in each frame. Regions between consecutive frames are then matched to obtain two-view correspondences. Affine transformations are computed from each pair of corresponding regions to define pixel matches. Pixels matches over consecutive image pairs are concatenated to obtain pixel-level motion trajectories across the image sequence. Motion patterns are learned from the extracted trajectories using a time-delay neural network. They applied the proposed method to recognize 40 hand gestures of American Sign Language. They approximated the human head and hand shapes by ellipses. Roy and Jawahar [31] presented a feature selection method for hand geometry based person authentication system. They used lengths of four fingers and widths at five equidistant points on each finger as raw features. Since the localization of hands in arbitrary scenes is difficult, one of the major difficulties associated with low level features is that the hand has to be localized before feature extraction.

3. Gesture Classification

The hand gesture classification approaches in the literature can be categorized into two main categories: rule-based approaches, in which the gestures are classified according to manually encoded rules and machine learning based approaches those are using a set of exemplars to infer models of gestures.

3.1. Rule-Based Approaches

In these approaches features of the input features are compared to the manually encoded rules. If any of the features or feature sets matches a rule, the related gesture will be given as output. As an example Cutler and Turk [6] used a rule-based technique to identify an action based on a set of conditions in their view-based approach to gesture recognition. They defined six motion rules for corresponding six gestures. When the hands trace a motion path like a predefined rule, corresponding gesture is selected as output.

3.2. Learning Based Approaches

As indicated in the previous section, the rule-based approaches depend on the ability of humans to find rules to classify the gestures. Learning-based approaches are alternative solutions to this problem when finding rules between features is not applicable. In this approach mappings between high-dimensional feature sets and gestures are done by machine learning algorithms. The most popular method for this approach is using HMMs in which gestures are treated as the output of a stochastic process. Many of the recent works Nair and Clark [25], Starner *et al*. [34], and Marcel [23] focused on HMMs for gesture recognition. Russell and Norvig [32] defines the HMM as “a temporal probabilistic model in which the state of the process is described by a single discrete random variable.” The possible values of the variable are the possible states of the world. Haberdar [14] used HMM for gesture recognition in his thesis study which is about Turkish Sign Language recognition. In the study 172 signs are used. Model parameters are determined by training 6 different HMM and using training data. For recognition of an observation sequence, forward and backward algorithms are used by calculating the probabilities of HMM.

4. Conclusions

The visual analysis of human hand gestures has major applications in human computer interaction and understanding human activities. A number of these applications could be virtual reality systems, understanding and synthesizing sign languages, advanced human interfacing and controlling robots and other machineries. The scope of this survey is limited to the analysis of human hand gestures and the models developed for this analysis. A taxonomy of the methods is introduced. Three main approaches are discussed: 2-D approaches without explicit shape models, 2-D approaches with explicit shape models, and 3-D approaches. The suitability of each method depends largely on the problem at hand. Although a large amount of work has been already performed in this topic, many more issues which are real challenges in front of the researchers should be considered. In most of the works done so far, segmentation has been skipped and the works have concentrated on posture or gesture modeling by assuming a uniform and static background. These assumptions hinder hand gesture applications from entering into the real life. Highly sophisticated models are not applicable and the computation cost is a prohibiting factor. More flexible modeling methods with less complexity are necessary. Multiple hands and occlusion are among the challenges. Hand pose recovery depends on initialization and clues such as current viewpoint provided by the user. Lack of ground truth data to measure system performance is another major challenge. Complex gestures are too difficult to extract and the available methods rely on a limited range of feasible postures. The general trend in current approaches is using single camera systems. However, there exist an inevitable tendency to avoid occlusions by using multiple camera systems and exploring 3D features. Although these systems are more expensive, they can provide better ways to handle occlusions and can lead to more accurate hand tracking systems for advanced tasks such as virtual object manipulation. A higher level of functionality can be achieved by developing a generic set of hand postures/gestures and interpreting them symbolically after acquisition. Integrating hand gesture recognition systems with the information of the context in which they are used is also an important trend of future works.

REFERENCES

[1] T. Agrawal, S. Chaudhuri, (2003), *Gesture Recognition Using Position and Appearance Features*, *ICIP* pp. 109–112.

[2] F. Bettio, A. Giachetti, E. Gobbetti, F. Marton, G. Pintore(2007). *A Practical Vision Based Approach to Unencumbered Direct Spatial Manipulation in Virtual Worlds*, *In Eurographics Italian Chapter Conf.*

[3] M. J. Black (1996). *EigenTracking: Robust Matching and Tracking of Articulated Objects Using a View Based Representation*, *International Journal of Computer Vision*, 26(1), pp. 63–84, 1998.

[4] M. Bray, E. Koller Meier, P. Muller, L. Van Gool, N.N. Schraudolph, (2004), *3D Hand Tracking By Rapid Stochastic Gradient Descent Using A Skinning Model*, Visual Media Production, 1st European Conf. pp. 59–68.

[5 L. Bretzner, I. Laptev, T. Lindeberg, S. Lenman, Y. Sundblad (2001), *A Prototype System for Computer vision based Human Computer Interaction*, Technical report, Stockholm, Sweden.

[6] R. Cutler, M. Turk. *View based Interpretation of Real-time Optical Flow for Gesture Recognition*, *3rd IEEE Conf. on Face and Gesture Recognition*, Nara, Japan, April 1998.

[7] J. Davis, M. Shah (1994), *Visual Gesture Recognition*, *Vision, Image and Signal Precessing*. 141(2) pp. 101–106.

[8] B. Dorner (1994),*Chasing the Colour Glove*, MSc. Thesis. Burnaby, BC, Canada: School of Computer Science,
Simon Fraser University.

[9] J. Eisenstein, R. Davis (2003), *Natural Gesture in Descriptive Monologues*, *Proc. ACM Symp. User Interface Software and Technology*, ACM Press , pp. 69–70.

[10] S. Fujisawa, K. Ohkubo, T. Nishi, T. Yoshida, Y. Shidama, H. Yamaura (1997), *Fundamental research on human interface devices for physically handicapped persons*, *23rd Int. Conf. IECON* , New Orleans.

[11] P. Gejgus, J. Placek, M. Sperka (2004), *Skin color segmentation method based on mixture of Gaussians and its application in Learning System for Finger Alphabet*, *Proc. Comp. Sys. Tech*.,

[12] G. Gonar, T. Yildirim, (2005), Hand *Gesture Recognition Using Artificial Neural Networks*, *Signal Processing and Communications Applications Conference*, *Proc. of the 13th IEEE* Vol. , pp: 210–213.

[13] N. Gupta, P. Mittal, R. Sumantra Dutta, S. Chaudhury, S. Banerjee (2002). *Developing a gesture based inter- face*, *IETE, Journal of Research: Special Issue on Visual Media Processing*.

[14] H. Haberdar (2005). *Real Time Isolated Turkish Sign Language Recognition from Video using Hidden Markov Models with Global Features*, MSc. thesis, Istanbul, Computer Engineering, Yildiz Teknik University.

[15] M. Kolsch , T.T. Hollerer (2004), *Vision Based Interfaces for Mobility*, In Proc. *1st IEEE Int. Conf. on Mobile and Ubiquitous Systems: Networking and Services*, pp. 86–94, Boston, MA.

[16] J.S. Lee (2004), *Hand region extraction and Gesture recognition from video stream with complex background through entropy analysis Engineering in Medicine and Biology Society*, *26th Annual Int. Conf. IEEE*, pp 1513–1516.

[17] J. Lee, T. Kunii (1995), *Model Based Analysis of Hand Posture* ,*IEEE Computer Graphics and Applications*.Vol.15 No. 5 pp. 77-86

[18] R.H. Liang, M. Ouhyoung (1996), *A Sign Language Recognition System Using Hidden Markov Model and Context Sensitive Search, Proc. ACM Symp. on Virtual Reality Software and Technology*, pp. 59–66, HongKong.

[19] L. Brethes, P. Menezes, F. Lerasle, J. Hayet (2004). *Face Tracking and hand gesture recognition for human robot interaction, In Int. Conf. on Robotics and Automation*, Vol. 2, pp. 1901–1906 New Orleans.

[20] A. Malima, E. Ozgur, M. Cetin (2006), *A fast algorithm for vision based hand gesture recognition for robot control, 14th IEEE Int. Conf. on Signal Processing and Communications Applications*, Antalya, Turkey.

[21] Malkawi A.M., Srinivasan R.S. (2005), *A new paradigm for Human-Building Interaction: the use of CFD and Augmented Reality, Automation in Construction Journal* 14 (1), pp. 71–84.

[22] V.M. Mantyla, J. Miintyjiirvi, T. Seppben, E. Tuulari (2000) . *Hand Gesture Recognition of a Mobile Device User, in IEEE Int. Conf. on Multimedia and Expo*, Vol. 1, pp. 281–284, NewYork.

[23] S. Marcel (2000). *Hand Gesture Recognition using Input-Output Hidden Markov Models*, *Proc. 4th IEEE Int. Conf. on Automatic Face and Gesture Recognition*, pp 456.

[24] M. Marschall (2004) , *Virtual Sculpture Gesture Controlled System for Artistic Expression.* Int. Symp. on Gesture Interfaces for Multimedia Systems, Leeds, UK.

[25] V. Nair, J. Clark (2002) . *Automated Visual Surveillance Using Hidden Markov Models*, In VI02, pp 88.

[26] V. Philomin, R. Duraiswami, L. Davis (2000), *Pedestrian tracking from a moving vehicle*, *IEEE Intelligent Vehicles Symp*.

[27] J. F. Postigo, V. A. Mut, R. O. Carelli, L. A. Baigorria, B. R. Kuchen (2000), *Hand controller for bilateral teleoperation of robots*, *ROBOTICA*, 18, 2000, pp. 677–686, UK.

[28] F. Quek, *Unencumbered gesture interaction*, *IEEE Multimedia*, Vol. 3. no. 3, pp. 36-47, 1996.

[29] X. Ren, A. C. Berg, J. Malik (2005), *Recovering human body configurations using pairwise constraints between parts*, In ICCV, Vol. 1, pages 824–831.

[30] B. J. Rhodes, N. Minar, J. Weaver (1999), *Wearable Computing Meets Ubiquitous Computing: Reaping the Best of Both Worlds. Proc. 3rd Int. Symp. on Wearable Computers*, pp. 141–149.

[31] V. Roy, C.V. Jawahar (2005), *Feature Selection for Hand-Geometry based Person Authentication, in Proc. Of Int. Conf. on advanced computing and communication*.

[32] S. Russell , P. Norvig (1995), *Artificial Intelligence: A Modern Approach*. Prentice Hall, Englewood Cliffs, NJ.

[33] A. Sorrentino, F. Cuzzolin, R. Frezza(1997), *Using Hidden Markov Models and Dynamic Size Functions for Gesture Recognition*, BMVC.

[34] T. Starner , J. Weaver, A. Pentland , *A wearable computer based American sign language recognizer*, *Proc. IEEE Int. Symp. Wearable Computing*, Oct. 1997, pp. 130–137.

[35] E. Ueda, (2003), *A Hand Pose Estimation for Vision-Based Human Interfaces, IEEE Transactions on Indus trial Electronics*, Vol. 50, No. 4, pp. 676–684.

[36] A. Utsumi and J. Ohya (1999), *Multiple Hand Gesture Tracking using Multiple Cameras, Proc. Int. Conf.* on

Computer Vision and Pattern Recognition, pp.473–478.
[37] C. Vogler and D. Metaxas, *ASL recognition based on a coupling between HMMs and 3D motion analysis, Proc. IEEE Int. Conf. Computer Vision*, Mumbai, India, Jan. 1998, pp. 363–369.
[38] Y. Wu and T. S. Huang (1999), *Vision-Based Gesture Recognition: A Review, Lecture Notes in Computer Science*, Vol. 1739, pp. 103–110.
[39] M.H. Yang (2002), *Extraction of 2D Motion Trajectories and Its Application to Hand Gesture Recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence archive*, Vol. 24, No. 8, pp. 1061–1075.
[40] M. Zahedi, D. Keysers, H. Ney (2005). *Appearance-Based Recognition of Words in American Sign Language, 2nd Proc. Conf. on Pattern Recognition and Image Analysis,* Vol. 3522, pp511–519, Estoril, Portugal.