International Journal of Scientific & Engineering Research, Volume 4, Issue 6, June 2013
ISSN 2229-5518

Hand Gesture Recognition using Computer Vision

Ashutosh Samantaray, Sanjaya Kumar Nayak, Ashis Kumar Mishra

Abstract-The use of the gesture system in our daily life as a natural human-human interaction has inspired the researchers to simulate and utilize this gift in human-machine interaction which is appealing and can take place the bore interaction ones that existed such as television, radio, and various home appliances as well as virtual reality will worth and deserve its name. This kind of interaction ensures promising and satisfying outcomes if applied in systematic approach, and supports unadorned human hand when transferring the message to these devices which is easiest, comfort and desired rather than the communication that requires frills to deliver the message to such devices. With the rapid emergence of 3d applications and virtual environments in computer system the need for a new type of interaction device arises. This is because the traditional devices such as mouse, keyboard and joystick become inefficient and cumbersome within this virtual environments.in other words evolution of user interfaces shapes the change in the Human- Computer Interaction (HCI).Intuitive and naturalness characteristics of "Hand Gestures" in the HCI have been the driving force and motivation to develop an interaction device which can replace current unwieldy tools. A survey on the methods of analysing, modelling and recognizing hand gestures in the context of the HCI is provided in this paper.

Index Terms-Hand Gesture, Computer Vision, HCI, Gesture Recognition, Hand Posture, Hidden Markov Model

1 INTRODUCTION

Gestures are a powerful means of communication among humans. In fact, gesturing is so deeply rooted in our communication that people often continue gesturing when speaking on the telephone. Hand gestures provide a separate complementary modality to speech for expressing ones ideas. Information associated with hand gestures in a conversation is degree, discourse structure, spatial and temporal structure. So, a natural interaction between humans and computing devices can be achieved by using hand gestures for communication between them.

Ashutosh Samantaray is currently pursuing bachelor's degree program (2009-2013) in computer science and engineering in College of Engineering and Technology (BPUT), India. E-mail: ashu.eng91@gmail.com . His current Field of research being Real time systems.

Sanjaya Kumar Nayak is currently pursuing bachelor's degree program (2009-2013) in computer science and engineering in College of Engineering and Technology (BPUT), India. E-mail: Sanjaya.eng91@gmail.com His current field of research being Real time system and open gl.

Ashis Kumar Mishra has completed his post graduate from KIIT University (2011) and under graduate from Biju Pattnaik University of Technology (2007). He is currently a faculty member in Department of CSE in College of Engineering and Technology, Bhubaneswar, India. E-mail: ashiskumar.misra@gmail.com . He has published 9 numbers of International Journals and a National Conference in fields of Cryptography, Real Time Systems, Evolutionary algorithm, Cloud Computing and machine learning

Vision has the potential of carrying a wealth of information in a non-intrusive manner and at a low cost; therefore it constitutes a very attractive sensing modality for developing perspective user interfaces. Proposed approaches for vision driven interactive user interfaces resort to technologies such as head tracking, face and facial expression recognition, eye tracking and gesture recognition.

The key problem in gesture interaction is how to make hand gestures understood by computers. The approaches present can be mainly divided into "Data-Glove based" and "Vision Based" approaches. The Data-Glove based methods use sensor devices for digitizing hand and finger motions into multi-parametric data. The extra sensors make it easy to collect hand configuration and movement. However, the devices are quite expensive and bring much cumbersome experience to the users [1]. In contrast, the Vision Based methods require only a camera [2], thus realizing a natural interaction between humans and computers without the use of any extra devices. These systems tend to complement biological vision by describing artificial vision systems that are implemented in software and/or hardware. This poses a challenging problem as these systems need to be background invariant, lighting insensitive, person and camera independent to achieve real time performance. Moreover, such systems must be optimized to meet the requirements, including accuracy and robustness.

The purpose of this paper is present a review of Vision based Hand Gesture Recognition techniques for human computer interaction, consolidating the various available approaches, pointing out their general advantages and disadvantages. Although other reviews have been written on the subsets of hand posture and gesture recognition [3], [4], [5], this one specifically relates to the vision based technique and is up-to-date. It is intended to point out the various open research issues as well as act as a starting point for anyone interested in using hand gesture recognition in their interfaces.

A. Overview

The organization of the rest of this paper is as follows.

Section 2 highlights the various computer vision techniques for hand gesture recognition. Section 3 discusses the process of vision based hand gesture recognition system. Section 4 describes various application areas of hand gesture recognition. Section 5 points out the current works and Section 6 conclude the paper.

2 COMPUTER VISION TECHNIQUES FOR HAND GESTURE RECOGNITION

Most of the complete hand interactive systems can be considered to be comprised of three layers: detection, tracking and recognition. The detection layer is responsible for defining and extracting visual features that can be attributed to the presence of hands in the field of view of the camera(s). The tracking layer is responsible for performing temporal data association between successive image frames, so that, at each moment in time, the system may be aware of what is where". Moreover, in model-based methods, tracking also provides a way to maintain estimates of model parameters, variables and features that are not directly observable at a certain moment in time. Last, the recognition layer is responsible for grouping the spatiotemporal data extracted in the previous layers and assigning the resulting groups with labels associated to particular classes of gestures. In this section, research on these three identified sub problems of vision-based gesture recognition is reviewed.

2.1 Detection

The primary step in gesture recognition systems is the detection of hands and the segmentation of the corresponding image regions. This segmentation is crucial because it isolates the task-relevant data from the image background, before passing them to the subsequent tracking and recognition stages. A large number of methods have been proposed in the literature that utilize a several types of visual features and, in many cases, their combination. Such features are skin colour, shape, motion and anatomical models of hands. [6] A comparative study on the performance of some hand segmentation techniques can be found.

2.1.1 Color

Skin color segmentation has been utilized by several approaches for hand detection. A major decision towards providing a model of skin color is the selection of the color space to be employed. Several color spaces have been proposed including RGB, normalized RGB, HSV, YCrCb, YUV, etc. Color spaces efficiently separating the chromaticity from the luminance components of color are typically considered preferable. This is due to the fact that by employing chromaticity-dependent components of color only, some degree of robustness to illumination changes can be achieved. Terrillon et al [7] review different skin chromaticity models and evaluate their performance.

The perceived color of human skin varies greatly across human races or even between individuals of the same race. Additional variability may be introduced due to changing illumination conditions and/or camera characteristics. Therefore, color-based approaches to hand detection need to employ some means for compensating for this variability. In [8], an adaptation technique estimates the new parameters for the mean and covariance of the multivariate Gaussian skin color distribution, based on a linear combination of previous parameters. However, most of these methods are still sensitive to quickly changing or mixed lighting conditions. A simple color comparison scheme is employed in [9], where the dominant color of a homogeneous region is tested as if occurring within a color range that corresponds to skin color variability. More advanced color segmentation techniques rely on histogram matching [10] or employ a simple look-up table approach [11, 12] based on the training data for the skin and possibly its surrounding areas.

In general, color segmentation can be confused by background objects that have a color distribution similar to human skin. However, background subtraction is typically based on the assumption that the camera system does not move with respect to a static background. To solve this problem, some research [13], has looked into the dynamic correction of background models and/or background compensation methods. Skin color is only one of many cues to be used for to hand detection. For example, in cases where the faces also appear in the camera field of view, further processing is required to distinguish hands from faces. Thus, skin color has been utilized in combination with other cues to obtain better performance. Stereoscopic information has been utilized mainly in conjunction with the skin color cue to enhance the accuracy of hand localization.

2.1.2 Shape

The characteristic shape of hands has been utilized to detect them in images in multiple ways. Much information can be obtained by just extracting the contours of objects in the image. If correctly detected, the contour represents the shape of the hand and is therefore not directly dependent on viewpoint, skin color and illumination. On the other hand, the expressive power of 2D shape can be hindered by occlusions or degenerate viewpoints. In the general case, contour extraction that is based on edge detection results in a large number of edges that belong to the hands but also to irrelevant background objects. Therefore, sophisticated post-processing approaches are required to increase the reliability of such an approach. In this spirit, edges are often combined with skin color and background subtraction/motion cues.

In the 2D/3D drawing systems of [13], the user's hand is directly extracted as a contour by assuming a uniform background and performing real-time edge detection in this image. Examples of the use of contours as features are found in both model [14] and appearance based techniques [15]). In [16], finger and arm link candidates are selected through the clustering of the sets of parallel edges. In a more global approach [17], hypotheses of hand 3D models are evaluated by first synthesizing the edge image of a 3D model and comparing it against the acquired edge image. The fingertip of the user was detected in both images of a calibrated stereo pair. In these images, the two points at which this tip appears establish a stereo correspondence, which is utilized to estimate the fingertip's position is 3D space. In turn, this position is utilized by the system to estimate the distance of the finger from the desk and, therefore, determine if the user is touching it.

2.1.3 3D model-based detection

A category of approaches utilize 3D hand models for the detection of hands in images. One of the advantages of these methods is that they can achieve view independent detection. The employed 3D models should have enough degrees of freedom to adapt to the dimensions of the hand(s) present in an image. Different models require different image features to construct feature-model correspondences. Point and line features are employed in kinematic hand models to recover angles formed at the joints of the hand [18]. Hand postures are then estimated provided that the correspondences between the 3D model and the observed image features are well established. Various 3D hand models have been proposed in the literature. In [19], a full hand model is proposed which has 27 degrees of freedom (DOF) (6 DOF for 3D location/orientation and 21 DOF for articulation). In [19], edge features in the two images of a stereoscopic pair are corresponded to extract the orientation of in-between joints of fingers. These are subsequently utilized for model based tracking of the hands. Some approaches utilize a deformable model framework to fit a 3D model of the hand to image data. The fitting is guided by forces that attract the model to the image edges, balanced by other forces that tend to preserve continuity and evenness among surface points. The process is enhanced with anatomical data of the human hand that are incorporated into the model. Also, to fit the hand model to an image of a real hand, characteristic points on the hand are identified in the images, and virtual springs are implied which pull these characteristic points to goal positions on the hand model.

2.1.4 Motion

Motion is a cue utilized by a few approaches to hand detection. The reason is that motion-based hand detection demands for a very controlled setup, since it assumes that the only motion in the image is due to hand movement. Indeed, early works assumed that hand motion is the only motion occurring in the imaged environment. In more recent approaches, motion information is combined with additional visual cues. In the case of static cameras, the problem of motion estimation reduces to that of background maintenance and subsequent subtraction. For example in [20] such information is utilized to distinguish hands from other skin-colored objects and cope with lighting conditions imposed by colored lights. The difference in luminance of pixels from two successive images is close to zero for pixels of the background. By choosing and maintaining an appropriate threshold, moving objects are detected within a static scene.

In [21], a novel feature, based on motion residue, is proposed. Hands typically undergo non-rigid motion, because they are articulated objects. Consequently, hand detection capitalizes on the observation that for hands, inter-frame appearance changes are more frequent than for other objects such as clothes, face, and background.

2.2 Tracking

Tracking, or the frame-to-frame correspondence of the segmented hand regions or features, is the second step in the process towards understanding the observed hand movements. The importance of robust tracking is twofold. First, it provides the inter-frame linking of hand/finger appearances, giving rise to trajectories of features in time. These trajectories convey essential information regarding the gesture and might be used either in a raw form (e.g. in certain control applications like virtual drawing the tracked hand trajectory directly guides the drawing operation) or after further analysis (e.g. recognition of a certain type of hand gesture). Second, in model-based methods, tracking also provides a way to maintain estimates of model parameters variables and features that are not directly observable at a certain moment in time.

2.2.1 Template based tracking

This class of methods exhibits great similarity to methods for hand detection. Members of this class invoke the hand detector at the spatial vicinity that the hand was detected in the previous frame, so as to drastically restrict the image search space. The implicit assumption for this method to succeed is that images are acquired frequently enough.

Correlation-based feature tracking is directly derived from the above approach. In [22] correlation-based template matching is utilized to track hand features across frames. Once the hand(s) have been detected in a frame, the image regions in which they appear is utilized as the prototype to detect the hand in the next frame. Again, the assumption is that hands will appear in the same spatial neighbourhood. This technique is employed for a static camera in [23], to obtain characteristic patterns of gestures, as seen from a particular view. Real-time performance is achieved by pre-computing "motion templates" which are the product of the spatial derivatives of the reference image to be tracked and a set of motion fields. Some approaches detect hands as image blobs in each frame and temporally correspond blobs that occur in proximate locations across frames. Approaches that utilize this type of blob tracking are mainly the ones that detect hands based on skin color, the blob being the correspondingly segmented image region. Blob-based approaches are able to retain tracking of hands even when there are great variations from frame to frame.

2.2.2 Optimal estimation techniques

Feature tracking has been extensively studied in computer vision. In this context, the optimal estimation framework provided by the Kalman Filter has been widely employed in turning observations (feature detection) into estimations (extracted trajectory). The reasons for its popularity are real-time performance, treatment of uncertainty, and the provision of predictions for the successive frames.

Treating the tracking of image features within a Bayesian framework has been long known to provide improved estimation results. In [24], a system tracks a single person by color-segmentation of the image into blobs and then uses prior information about skin color and topology of a person's body to interpret the set of blobs as a human figure. In [25], a method is proposed for tracking human motion by grouping pixels into blobs based on coherent motion, color and temporal support using an expectation-maximization (EM) algorithm. Each blob is subsequently tracked using a Kalman Filter. Finally, in [26] the contours of blobs are tracked across frames by a combination of the Iterative Closed Point (ICP) algorithm and a factorization method to determine global hand pose.

2.2.3 Tracking based on the Mean Shift algorithm

The Mean Shift algorithm [27] is an iterative procedure that detects local maxima of a density function by shifting a kernel towards the average of data points in its neighbourhood. The algorithm is significantly faster than exhaustive search, but requires appropriate initialization. The Mean Shift algorithm has been utilized in the tracking of moving objects in image sequences. It characterizes the object of interest through its color distribution as this appears in the acquired image sequence and utilizes the spatial gradient of the statistical measurement towards the most similar (in terms of color distribution similarity) image region. Mean-Shift tracking is robust and versatile for a modest computational cost. It is well suited for tracking tasks where the spatial structure of the tracked objects exhibits such a great variability that trackers based on a space-dependent appearance reference would break down very fast. On the other hand, highly cluttered background and occlusions may distract the mean-shift trackers from the object of interest. The reason appears to be its local scope in combination with the single-state appearance description of the target.

2.2.4 Particle Filtering

Particle filters have been utilized to track the position of hands and the configuration of fingers in dense visual clutter. In this approach, the belief of the system regarding the location of a hand is modelled with a set of particles. A disadvantage of particle filters is that for complex models (such as the human hand) many particles are required, a fact which makes the problem intractable especially for high-dimensional models. Therefore, other assumptions are often utilized to reduce the number of particles.

2.3 Recognition

The overall goal of hand gesture recognition is the interpretation of the semantics that the hand(s) location, posture, or gesture conveys. Basically, there have been two types of interaction in which hands are employed in the user's communication with a computer. The first is control applications such as drawing, where the user sketches a curve while the computer renders this curve on a 2D canvas. Methods that relate to hand-driven control focus on the detection and tracking of some feature (e.g. the fingertip, the centroid of the hand in the image etc.) and can be handled with the information extracted through the tracking of these features. The second type of interaction involves the recognition of hand postures, or signs, and gestures. Naturally, the vocabulary of signs or gestures is largely application dependent. Typically, the larger the vocabulary is, the hardest the recognition task becomes.

The recognition of postures is of topic of great interest on its own, because of sign language communication. Moreover, it also forms the basis of numerous gesture-recognition methods that treat gestures as a series of hand postures. Besides the recognition of hand postures from images, recognition of gestures includes an additional level of complexity, which involves the parsing, or segmentation, of the continuous signal into constituent elements.

The fact that even hand posture recognition exhibits considerable levels of uncertainty casts the above processing computationally complex or error prone. Several of the reviewed works indicate that lack of robustness in gesture recognition can be compensated by addressing the temporal context of detected gestures. This can be established by letting the gesture detector know of the grammatical or physical rules that the observed gestures are supposed to express. Based on these rules, certain candidate gestures may be improbable. In turn, this information may disambiguate candidate gestures, by selecting to recognize the most likely candidate. The framework of Hidden Markov Models (HMMs) provides a suitable framework for modelling the context-dependent reasoning of the observed gestures.

2.3.1 Template matching

Template matching, a fundamental pattern recognition technique, has been utilized in the context of both posture and gesture recognition. In the context of images, template matching is performed by the pixel-by-pixel comparison of a prototype and a candidate image. The similarity of the candidate to the prototype is proportional to the total score on a preselected similarity measure. For the recognition of hand postures, the image of a detected hand forms the candidate image which is directly compared with prototype images of hand postures. The best matching prototype (if any) is considered as the matching posture. Clearly, because of the pixel-by-pixel image comparison, template matching is not invariant to scaling and rotation. Template matching was one of the first methods employed to detect hands in images.

To cope with the variability due to scale and rotation, some authors have proposed scale and rotational normalization methods. In [28], the image of the hand is normalized for rotation based on the detection of the hands main axis and, then, scaled with respect to hand dimensions in the image. Therefore, in this method the hand is constrained to move on a planar surface that is front of parallel to the camera. Edge detection is performed on the image of the isolated hand and edge orientations are computed. The histogram of these orientations is used as the feature vector.

2.3.2 Boosting

A real-time gesture recognition system is presented. Their method which is based on skin-color segmentation is facilitated by a boosting algorithm [29] for fast classification. To normalize for orientation, the user is required to wear a wristband so that the hand shape can be easily mapped to a canonical frame. In [30], a classification approach was proposed, together with parameter interpolation to track hand motion. Image intensity data was used to train a hierarchical nearest neighbour classifier, classifying each frame as one of 360 views, to cope with viewpoint variability. This method can handle fast hand motion, but it relies on clear skin color segmentation and controlled lighting conditions.

2.3.3 Model-based recognition methods

Most of the model-based gesture recognition approaches employ successive approximation methods for the estimation of their parameters. Since gesture recognition is required to be invariant of relative rotation, intrinsic parameters such as joint angles are widely utilized. The strategy of most methods in this category is to estimate the model parameters, e.g. by inference or optimization, so that the extracted features match a model.

In an early approach [31], the 3D trajectory of hands was estimated in the image, based on optical flow. The extremal points of the trajectory were detected and used as gesture classification features. The 3D trajectories of hands are acquired by stereo vision and utilized for HMM-based learning and recognition of gestures. Different feature vectors were evaluated as to their efficacy in gesture recognition. The results indicated that choosing the right set of features is crucial to the obtained performance. In particular, it is observed that velocity features are superior to positional features, while partial rotational invariance is also a discriminative feature.

2.3.3.1 HMMs

A Hidden Markov Model (HMM)[32] is a statistical model in which a set of hidden parameters is determined from a set of related, observable parameters. In a HMM, the state is not directly observable, but instead, variables influenced by the state are. Each state has a probability distribution over the possible output tokens. Therefore, the sequence of tokens generated by an HMM provides information about the sequence of states. In the context of gesture recognition, the observable parameters are estimated by recognizing postures (tokens) in images. For this reason and because gestures can be recognized as a sequence of postures, HMMs have been widely utilized for gesture recognition. In this context, it is typical that each gesture is handled by a different HMM. The recognition problem is transformed to the problem of selecting the HMM that matches best the observed data, given the possibility of a state being observed with respect to context. This context may be spelling or grammar rules, the previous gestures, cross-modal information (e.g. audio) and others.

In [33], the 3D locations that result from stereo multiple-blob tracking are input to a HMM that integrates a skeletal model of the human body. Based on the 3D observations, the approach attempts to infer the posture of the body.

2.4 Complete gesture recognition systems

Systems that employ hand driven human-computer communication, interpret the actions of hands in different modes of interaction depending on the application domain. In some applications the hand or finger motion is tracked to be replicated in some kind of 2D or 3D manipulation activity. For example, in a painting application the finger may sketch a figure in thin air, which however, is to be replicated as a drawing on the computer's screen. In other cases, the posture, motion, and/or gesture of the user must be interpreted as a specific command to be executed or a message to be communicated. Such a specific application domain is sign language understanding for the hearing impaired. Most of the systems presented in this subsection fall in these categories, however, there are some that combine the above two modes of interaction. Finally, a few other applications focus on gesture recognition for understanding and annotating human behaviour, while others attempt to model hand and body motion for physical training.

3 VISION BASED HAND RECOGNITION PROCESS

In vision based hand gesture recognition system, the movement of the hand is recorded by video camera(s). This input video is decomposed into a set of features taking individual frames into account. Some form of filtering may also be performed on the frames to remove the unnecessary data, and highlight necessary components. For example, the hands are isolated from other body parts as well as other background objects. The isolated hands are recognized for different postures. Since, gestures are nothing but a sequence of hand postures connected by continuous motions, a recognizer can be trained against a possible grammar. With this, hand gestures can be specified as building up out of a group of hand postures in various ways of composition, just as phrases are building up by words. The recognized gestures can be used to drive a variety of applications (Fig.1).

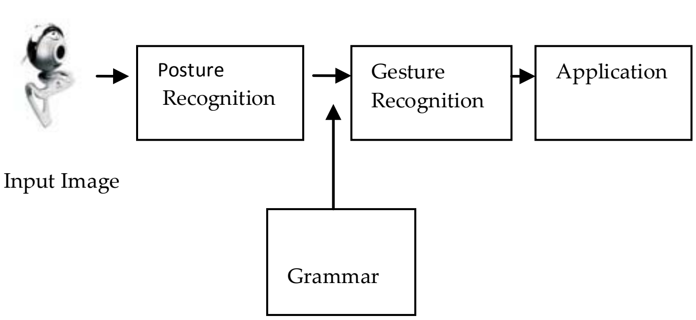

Fig.1. Hand Gesture Recognition Process

4 APPLICATION AREAS

There are many applications for vision-based system which interfere with our daily life; we brief some of these usages herein:

I- Human-computer or human-robot interaction.
II- Sign language recognition especially for hearing impaired people.
III- Hand and finger annotation for pianist player and this annotation can be used for remote education and for facilitating the producing of music sheets [34].
IV- Control of consumer electronics equipment and mechanical systems [35].
V- The interaction that done with visualization systems [35].
VI- Interactive games [35], interactive environment [36], and intelligent rooms [37].
VII- For mobile or hand held devices [36], the use of keyboard or touch screen become encumbered due to the limited size of such devices [36].
VIII- Main component of body language in linguistics [36].

5 CURRENT WORKS

We were working on a view based representation of the hand, including both color and shape cues. The system tracks and recognizes the hand poses based on a combination of multi-scale color feature detection, view-based hierarchical hand models and particle filtering. The hand poses, or hand states, are represented in terms of hierarchies of color image features at different scales, with qualitative inter-relations in terms of scale, position and orientation. These hierarchical models capture the coarse shape of the hand poses. In each image, detection of multi-scale color features is performed. The hand states are then simultaneously detected and tracked using particle filtering, with an extension of layered sampling referred to as hierarchical layered sampling. The particle filtering allows for the evaluation of multiple hypotheses about the hand position, state, orientation and scale, and a likelihood measure determines what hypothesis to choose. To improve the performance of the system, a prior on skin color is included in the particle filtering step.

A detailed description of the algorithms is given in [38].
As the coarse shape of the hand is represented in the feature hierarchy, the system is able to reject other skin colored objects that can be expected in the image (the face, arm, etc.). The hierarchical representation can easily be further extended to achieve higher discrimination to complex backgrounds, at the cost of a higher computational

complexity. An advantage of the approach is that it is to a large extent user and scale (distance) invariant. To some extent, the chosen qualitative feature hierarchy also shows view invariance for rotations out of the image plane (up to approx. 20-30 degrees for the chosen gestures).

6 CONCLUSION

In today's digitized world, processing speeds have increased dramatically, with computers being advanced to the levels where they can assist humans in complex tasks. Yet, input technologies seem to cause a major bottleneck in performing some of the tasks, under-utilizing the available resources and restricting the expressiveness of application use. Hand Gesture recognition comes to rescue here. Computer Vision methods for hand gesture interfaces must surpass current performance in terms of robustness and speed to achieve interactivity and usability. Considering the relative infancy of research related to vision based gesture recognition remarkable progress has been made. To continue this momentum it is clear that further research in the areas of feature extraction, classification methods and gesture representation are required to realize the ultimate goal of humans interfacing with machines on their own natural terms.

ACKNOWLEDGMENT

We would like to thank our teacher for his great efforts of supervising and leading us, to accomplish this fine work. To our friends and families, they were a great source of support and encouragement. We thank every person who gave us something to light our pathway; we thank them for believing in us.

REFERENCES

[1] A. Mulder, "Hand gestures for HCI", *Technical Report 96-1*, vol. Simon Fraster University, 1996

[2] F. Quek, "Towards a Vision Based Hand Gesture Interface", pp. 17-31, in *Proceedings of Virtual Reality Software and Technology*, Singapore,1994.

[3] Ying Wu, Thomas S Huang, "Vision based Gesture Recognition: A Review", Lecture Notes In Computer Science; Vol. 1739, *Proceedings of the International Gesture Workshop on Gesture-Based Communication in Human-Computer Interaction*, 1999

[4] K G Derpains, "A review of Vision-based Hand Gestures", Internal Report, Department of Computer Science. York University, February 2004

[5] Richard Watson, "A Survey of Gesture Recognition Techniques", Technical Report TCD-CS-93-11, Department of Computer Science, Trinity College Dublin, 1993.

[6] M. Cote, P. Payeur, and G. Comeau. Comparative study of adaptive segmentation techniques for gesture analysis in unconstrained environments. In *IEEE Int. Workshop on Imagining Systems and Techniques*, pages 28-33, 2006.

[7] J. Terrillon, M. Shirazi, H. Fukamachi, and S. Akamatsu. Comparative performance of different skin chrominance models and chrominance spaces for the automatic detection of human faces in color images. In *Proc. International Conference on Automatic Face and Gesture Recognition (FG)*, pages 54-61, 2000.

[8] J. Yang, W. Lu, and A. Waibel. Skin-color modeling and adaptation. In *ACCV*, pages 687-694, 1998, 1998.

[9] S. M. Dominguez, T. Keaton, and A. H. Sayed. A robust finger tracking method for wearable computer interfacing. *IEEE Transactions on Multimedia*, 8(5):956-972, 2001.

[10] S. Ahmad. A usable real-time 3D hand tracker. In *Asilomar Conference on Signals, Systems and Computers*, pages 1257-1261, Pacific Grove, CA, 1994.

[11] R. Kjeldsen and J. Kender. Finding skin in color images. In *IEEE Int. Conf. Automatic Face and Gesture Recognition*,pages 312-317, Killington, VT, 1996.

[12] F. Quek, T. Mysliwiec, and M. Zhao. Finger mouse: A freehand pointing interface. In *IEEE Int. Workshop on Automatic Face and Gesture Recognition*, pages 372-377, Zurich, Switzerland, 1995.

[13] A. Utsumi and J. Ohya. Image segmentation for human tracking using sequential-image-based hierarchical adaptation. In *Proc. IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 911-916, 1998.

[14] J. Kuch and T. Huang. Vision based hand modelling and tracking for virtual teleconferencing and telecollaboration. In *Proc. International Conference on Computer Vision (ICCV)*, pages 666-671, 1995.

[15] D. Gavrila and L. Davis. 3-D model-based tracking of humans in action: a multi-view approach. In *Proc. IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 73-80, 1996, 1996.

[16] A. Downton and H. Drouet. Image analysis for model-based sign language coding. In *Int. Conf. Image Analysis and Processing*, pages 637-644, 1991.

[17] D. Gavrila and L. Davis. Towards 3D model-based tracking and recognition of human movement: A multi-view approach. In *Int. Workshop on Automatic Face and Gesture Recognition*, pages 272-277, Zurich, Switzerland, 1995.

[18] J. Rehg and T. Kanade. Model-based tracking of self-occluding articulated objects. In *Proc. International Conference on Computer Vision (ICCV)*, pages 612-617, 1995.

[19] J. Rehg and T. Kanade. Digiteyes: Vision-based hand tracking for human-computer interaction. In *Workshop on Motion of Non-Rigid and Articulated Bodies*, pages 16-24, Austin Texas, November 1994.

[20] R. Cutler and M. Turk. View-based interpretation of real-time optical flow for gesture recognition. In *Proc. International*

Conference on Face and Gesture Recognition, pages 416{421, Washington, DC, USA, 1998. IEEE Computer Society.

[21] Q. Yuan, S. Sclaroff, and V. Athitsos. Automatic 2D hand tracking in video sequences. In *IEEE Workshop on Applications of Computer Vision*, pages 250-256, 1995.

[22] J. Crowley, F. Berard, and J. Coutaz. Finger tracking as an input device for augmented reality. In *International Workshop on Gesture and Face Recognition*, Zurich, June 1995.

[23] T. Darrell, I. Essa, and A. Pentland. Task-specific gesture analysis in real-time using interpolated views. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 18(12):1236-1242, 1996.

[24] C. R. Wren, A. Azarbayejani, T. Darrell, and A. Pentland. PFinder: Real-time tracking of the human body. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(7):780-785, 1997.

[25] C. Bregler. Learning and recognizing human dynamics in video sequences. In *Proc. IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 568-574, Puerto Rico, 1997.

[26] J. MacCormick and A. Blake. A probabilistic exclusion principle for tracking multiple objects. In *Proc. International Conference on Computer Vision (ICCV)*, pages 572-578, Corfu, Greece, 1999.

[27] Y. Cheng. Mean shift, mode seeking, and clustering. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 17(8):790-799, 1995.

[28] H. Birk, T. B. Moeslund, and C. B. Madsen. Real-time recognition of hand alphabet gestures using principal component analysis. In *Proc. Scandinavian Conference on Image Analysis*, Lappeenranta, Finland, June 1997.

[29] Y. Freund and R. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119-139, 1997.

[30] C. Tomasi, S. Petrov, and A. Sastry. 3D tracking = classification + interpolation. In *Proc. International Conference on Computer Vision (ICCV)*, volume 2, pages 1441-1448, Nice, France, 2003.

[31] F. Quek. Eyes in the interface. *Image and Vision Computing*, 13(6):511-525, 1995.

[32] A. Wilson and A. Bobick. Learning visual behaviour for gesture analysis. In *IEEE Symposium on Computer Vision*, Coral Gables, FL, 1995.

[33] C. Wren and A. Pentland. Dynamic models of human motion.In *IEEE Intl Conf. Automatic Face and Gesture Recognition*, Pages 22-27, Nara, Japan, 1997.

[34] Dimitry Gorodnichy, A. Yogeswaran, "Detection and Tracking of Pianist Hands and Fingers", Third Canadian Conference on Computer and Robot Vision (CRV 2006) Canada, June 2006, doi: 10.1109/CRV.2006.26

[35] Lars Bretzner, Ivan Laptev, and Tony Lindeberg, "Hand Gesture Recognition using Multi-Scale Colour Features, Hierarchical Models and Particle Filtering", Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition (FGR.02), Washington DC, USA, May 2002, pp. 423-428, doi: 10.1109/AFGR.2002.1004190.

[36] Yikai Fang, Kongqiao Wang, Jian Cheng and Hanqing Lu, "A Real-Time Hand Gesture Recognition Method", In Proceedings of IEEE International Conference on Multimedia and Expo (ICME 2007), Beijing, China, July 2007, pp. 995-998, doi: 10.1109/ICME.2007.4284820

[37] Hyunjin An, Daijin Kim, "Hand Gesture Recognition using 3D Depth Data", Department of Computer Science and Engineering, Pohang University of Science and Technology, Korea, 2009.

[38] Bretzner, L., Laptev, I., & Lindeberg, T. (2002) Hand Gesture Recognition using Multi-Scale Colour Features, Hierarchical Models and Particle Filtering. To appear in the *5th International Conference on Automatic Face and Gesture Recognition*, Washington, D.C: May 2002.