Vision Based Hand Gesture Recognition

Yanmin Zhu [1], Zhibo Yang [2], Bo Yuan [3]
Intelligent Computing Lab, Division of Informatics
Graduate School at Shenzhen, Tsinghua University
Shenzhen 518055, P. R. China
[1] zhuym111@gmail.com; [2] yangzhibo450@gmail.com; [3] yuanb@sz.tsinghua.edu.cn

***Abstract*—Hand gesture recognition has attracted much attention from academia and industry in recent years due to its apparent superiority over traditional techniques in human-computer interaction in terms of convenience. This domain has been investigated from different perspectives among which vision based approaches provide the most natural and intuitive interfaces. This paper presents a comprehensive review on vision based hand gesture recognition, with an emphasis on dynamic hand gestures. First, a brief introduction of the basic concepts and the classification of hand gesture recognition techniques are given. Then, a number of popular related technologies and interesting applications are reviewed. Finally, we give some discussion on the current challenges and open questions in this area and point out a list of possible directions for future work.**

Keywords-vision based hand gesture recognition; HCI; detection and segmentation; hand tracking; classification

I. INTRODUCTION

Computerized hand gesture recognition has received much attention from academia and industry in recent years, largely due to the development of human-computer interaction (HCI) technologies and the growing popularity of smart devices such as smart phones.

As the most flexible part of human body, hands play an important role in human's daily life. It is very natural for a person to use his/her hands when he/she wants to physically manipulate an object or communicate with other people. Traditionally, HCI is accomplished with devices such as mouse and keyboard, which are limited in terms of operational distance and convenience. By contrast, hand gesture recognition provides an alternative to these cumbersome devices, and enables people to communicate with computer more easily and naturally. In nowadays, hand gesture recognition technologies have already been successfully implemented in a wide range of applications, such as virtual reality [1], consumer electronics control [2], video games, sign language recognition [3] etc.

Note that there is a clear distinction between hand posture and hand gesture, although people often consider them to be identical. Hand posture refers to a single *static* pose of the hand (e.g., the victory sign) and can be represented by a single image. In contrast, hand gesture is *dynamic*, which is composed of a sequence of changing postures posed in a short period of time. For example, waving goodbye belongs to this category. In the following part of this paper, we will focus on dynamic hand gesture recognition, although both areas do share some common techniques.

To enable hand gesture recognition, numerous approaches have been proposed, which can be classified into various categories. A common taxonomy is based on whether extra devices are required for raw data collecting. In this way, they are categorized into data glove based hand gesture recognition [4], vision based hand gesture recognition [5], and color glove based hand gesture recognition [6] (see Fig. 1).

Figure 1. Data glove based, vision based and color glove based hand gesture recognition (from left to right).

Data glove based approaches require the user to wear a cumbersome glove-like device, which is equipped with sensors that can sense the movements of hand(s) and fingers, and pass the information to the computer. The advantages of these approaches are high accuracy and fast reaction speed. However, these techniques are not very natural and flexible, and the data gloves can be quite expensive.

Vision based approaches do not require the user to wear anything (naked hands). Instead, video camera(s) are used to capture the images of hands, which are then processed and analyzed using computer vision techniques. This type of hand gesture recognition is simple, natural and convenient for users and at present they are the most popular approaches to gesture recognition. However, there are still several challenges to be addressed, for instance, illumination change, background clutter, partial or full occlusion etc.

Color glove based approaches represent a compromise between data glove based approaches and vision based approaches. Intrinsically, they are similar to the latter, except that, with the help of colored gloves, the image preprocessing phase (e.g., segmentation, localization and detection of hands) can be greatly simplified. The disadvantages are similar to data glove based approaches: they are unnatural and not suitable for applications with multiple users due to hygiene issues.

Vision based hand gesture recognition can be further categorized into appearance based approaches and 3D model based approaches [7]. Appearance based approaches extract the low-level features from 2D images and compare them with predefined template gestures. They are relatively simple in computation but the loss of the depth information makes it susceptible to background disturbance. In contrast, 3D model based approaches can exploit the depth information and are much more computationally expensive but can identify hand gestures more effectively.

In this paper, we focus on vision based hand gesture recognition, more specifically the technologies and applications of appearance based approaches. The rest of the paper is organized as follows: Section II introduces the key algorithms used in hand gesture recognition. Applications of hand gesture recognition are presented in Section III. This paper is concluded in Section IV with some discussion and a list of possible directions for future work.

II. TECHNOLOGIES

Hand gesture recognition involves three important phases: detection and segmentation, tracking, and classification. The detection and segmentation part is to detect hands and segment the corresponding image regions from the background. Tracking exploits the spatial and temporal information of successive image frames to estimate the trajectories of hand motion. Classification takes hand trajectories as input to identify the specific types of gesture. The framework of a typical hand gesture recognition system is shown in Fig. 2.

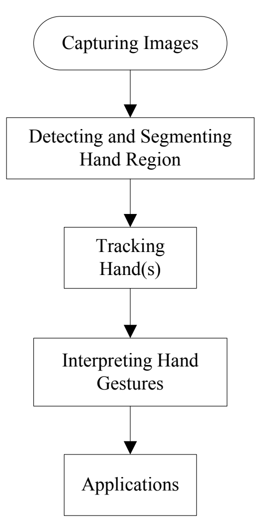

Figure 2. Framework of a typical hand gesture recognition system.

A. Detection and Segmentation

This part belongs to the preprocessing stage, which locates the task-relevant regions of the images and eliminates the background. It is both important and challenging for the reason that it generates the source data for the subsequent tracking and classification stages and the background might be very noisy in real-world situations. A large number of methods have been proposed based on various types of visual features or their combinations. For dynamic hand gesture recognition, the motion of hands instead of the pose of hands is the major concern and the computational complexity should be kept as low as possible. Consequently, low-level features and simple methods such as skin color, shape and background subtraction are often preferred.

1) Skin Color

Skin color detection is the most popular method for segmentation for its simplicity and convenience. Color space selection is a key factor for skin color detection. Since the RGB color space is sensitive to lighting conditions and has relatively high computational cost, several other orthogonal color spaces that can better separate the chromaticity from the luminance components have been proposed, such as HSV [8], YCbCr [9], YUV [10] and YIQ [11-12]. In these color spaces, the chromaticity information is often represented by one or two components, which are used in segmentation. In the HSV color space, the color information is represented by the H (hue) component and S (saturation) component, and is invariant to the change of V (value) for most of the time. As a result, H and S are often chosen to represent the color information. In the YCbCr color space, the luminance component of the image is represented by Y, while the chrominance component is represented by Cb and Cr.

With the chosen color space, the common methods for segmentation are: 1) using constant threshold or adaptive threshold selected empirically to identify skin areas, usually referred to as skin filters; 2) calculating the similarity between a predefined skin color model and the input image (i.e., the probability of a pixel or pixel block being part of the hand).

The advantages of skin color segmentation are: 1) it is computationally inexpensive and can be implemented in real time; 2) it is invariant against rotation, partial occlusion and pose change. However, this method is not capable of solving the task satisfactorily due to: 1) it is sensitive to illumination change; 2) it cannot handle complex background with skin-like colors or other skin object, such as human face and human arm; 3) many of these approaches are constrained to predefined models and are not robust enough.

2) Shape

The geometric features of hands can be also used for detection in different ways. The contour of hands can be obtained by applying edge detection operators, such as

Canny, Sobel, Prewitt, Roberts and Laplace operator. Compared with skin color detection, the contour feature has the superiority of being independent of illumination and skin color variation. However, its effectiveness can be hindered by occlusions and viewpoint change. Furthermore, edge detection may often result in a large number of edges that belong to irrelevant background objects. Therefore, the integration of shape, texture and color features may produce better performance [13].

There are several methods focusing on the morphology of hands such as fingertips [14]. One common clue to detect fingertips is the curvature [15]. Template matching is another technique that is often adopted in fingertip detection. Templates can be images of fingertips [16] or fingers [17]. However, template matching has some inherent drawbacks: 1) it is computationally expensive; 2) it cannot cope with scaling change or rotation of the target hand; 3) hand appearance can vary dramatically due to its 27 degrees of freedom, making it difficult to choose the right template.

3) Background Subtraction

Background subtraction is a method for segmenting the foreground. It compares the current image with the background image and extracts the foreground. It can only work in restricted situations, such as static background and unchanged illumination, which do not hold for most real scenarios. For slowly changing backgrounds, the model for background can be updated from frame to frame [18].

For motion detection, the former image can be regarded as the background image and the subtraction of successive frames can produce a rough estimation of the moving objects.

B. Tracking

The challenge of hand tracking mainly comes from the fact that the hand is a highly articulated object and cannot be treated as a rigid object, and the speed of gesture can be rather fast. Typical algorithms for hand tracking are: optical flow, Camshift, Kalman filtering, particle filtering, condensation algorithm etc. All these algorithms are based on certain assumptions and the integration of different algorithms can often generate better results.

1) Optical Flow

Optical flow methods are pixel-based approaches to object motion estimation where a direct relationship is assumed between object motion and intensity changes within an image sequence [19]. Among the existing methods for optical flow estimation, gradient based techniques are characterized by the assumption of brightness constancy and spatial-temporal smoothness [20].

The major advantage of using optical flow in tracking is that it does not require any *priori* knowledge of the object appearance. However, this advantage may be compromised by its high computational complexity. For real-time hand gesture recognition applications, the Lucas–Kanade method based on block motion model is often used, which does not contain the iterative procedure [21].

Optical flow based methods perform well when objects move not very fast and there are only a few moving objects. For applications demanding high accuracy, such as robot navigation, optical flow itself may not be sufficient due to the factor of drift, which is caused by error accumulating. In this situation, exploiting the color information in optical flow tracking may result in better results [22].

2) Camshift

Camshift, short for “Continuously Adaptive Mean Shift”, is an algorithm based on the mean shift algorithm. It was originally used for face tracking [23], and achieved good results. It is robust against noise and distractions. However, this algorithm alone cannot obtain good performance on hand tracking. The reason is that Camshift specifies the target using a rectangle while the hand is an object with plenty of concavo-convex sections, which cannot be well represented by a rectangle. For hand tracking, adjustment of the original algorithm is necessary, such as the incorporation of Kalman filtering [24].

3) Particle Filtering

Particle filters can be used to track objects in densely cluttered background. It performs a random search to obtain an estimate of the posterior distribution describing the object’s configuration. Compared with Kalman filtering, this approach has apparent superiority in that it is not limited by the unimodal nature of Gaussian densities, which cannot simultaneously represent alternative hypotheses. However, particle filters have a common issue: the degeneracy phenomenon [25]. After a few iterations, all but one sample will have negligible weights and a great amount of computational effort is wasted.

To address the degeneracy problem, the unscented particle filter (UPF) uses the unscented Kalman filter to generate sophisticated proposal distributions that seamlessly integrate the current observation [26]. In the meantime, the condensation algorithm, which was used to learn to track curves against cluttered backgrounds, can achieve better performance than Kalman filters, and also operates in real-time [27]. Particle filters and the mean shift algorithm can be also combined to preserve the advantage of mean shift while solving the degeneracy issue of particle filters [28].

4) TLD

Most recently, a new tracking algorithm called TLD (Tracking-Learning-Detection) [29] has received immense interests from academia. It uses a tracker and a detector to track the object independently while the learner estimates the detector’s errors and updates it to avoid similar errors in the future. With these three modules combined together, competitive tracking results have been obtained on various test samples. However, when applied to hand tracking, it often cannot achieve the same good result as on other objects such as cars. The reason again lies in the specific features of human hands and the fast changing appearances, which cannot be adequately learnt in a timely manner.

C. *Classification*

For hand gesture recognition, the final step is classification. It uses the information extracted and processed by previous steps and outputs the recognition results. Note that predefinition is required to distinguish real gestures from meaningless hand moves.

There have been various approaches to gesture recognition, which can be roughly divided into two categories: techniques based on mathematical models, such as Hidden Markov Model (HMM) and Finite State Machine (FSM), and techniques based on soft computing, such as neutral networks. Among them, HMM is the most frequently used approach to hand gesture recognition.

HMM is a Markov model in which the system being modeled is assumed to be a Markov process with unobserved (hidden) states [30]. A set of hidden parameters is determined from a set of related, observable parameters. Although initially introduced and studied in the late 1960's and early 1970's, HMM became popular in temporal pattern recognition in the 1990's. In the early time, HMM was mostly used in speech recognition and achieved good results [31]. Gradually, it started to be applied to gesture recognition, especially sign language recognition [32]. The popularity of HMM in hand gesture recognition is largely due to the fact that each gesture, which is spatial-temporal, can be conveniently treated as a sequence of postures, and HMM has proven capability of efficiently modeling spatial-temporal information. Each gesture is modeled by a different HMM and the given unknown gesture is tested by each HMM and the gesture corresponding to the HMM with the best match is returned as the final recognition result.

In [18], Chen and Fu applied HMM to the recognition of 20 different dynamic hand gestures. They combined the spatial and temporal features as the input feature vector to HMM. The training set was composed of 1200 image sequences, including 60 different gestures performed by 20 different people. Another 1200 image sequences were used in the test phase. With stationary background and normal lighting condition, the system's recognition rate was over 90% while running in real time. The recognition of Arabic numbers performed by hand can be also realized using HMM with recognition rate over 95% [33].

III. Applications

Since hand gestures have long been serving as an essential communication medium in human society, we believe that they can also play an equally important role in the computerized world. This section introduces a series of new applications of vision based hand gesture recognition in recent years to give an indication of its prospect in the future.

A. *Alternative to Touch-Based Devices*

Tablet PCs and smart phones have driven our everyday life into the era of touch-to-control, by getting rid of traditional input devices such as mouse and keyboard. In the meantime, vision based hand gesture recognition techniques are pushing the user experience one step further by providing a touch-free solution (the touch sensors are no longer required). Many major IT companies including Microsoft, Samsung, Sony and Qualcomm have integrated hand gesture recognition techniques into their products such as smart TVs, video game consoles and mobile processors. In the near future, it is very likely that people can use fingers and hand movement to conveniently control various devices [34, 35]. For example, in the project of Kinect Magic Cursor, a left mouse button click can be simulated by raising the left hand above the shoulder.

B. *3D Virtual Environment*

In the 3D virtual environment, the Kinect depth sensor can be used for producing 3D animations [36]. Users can upload 3D models, such as toys, into the system and perform as puppeteers. The system can then use the 3D models as puppets for an animation show (Fig. 3). In a low cost interface device for interacting with objects in the virtual environment using hand gestures [37], the core architecture consists of a central computing module, which can locate the position of fingers in specific moment and then remodel the hand postures. With a number of key patents [38 – 40], GestureTek has grown as a world-leader in the field of 3D video hand gesture control. For example, its WallFX system can create dynamic wall displays by tracking and responding to subtle human body movements, allowing for real-time interaction between the user and the display, and turning any vertical surface into an interactive display system. Its GestureTrack3D Hand Tracker is able to follow two hands to form a multi-touch mode, and can track up to ten hands simultaneously.

Figure 3. Example of 3D animations in front of a Kinect depth sensor.

C. *Surgical System*

In a surgical environment, hand gesture recognition systems can help doctors manipulate digital images during medical procedures using hand gestures instead of touch screens or computer keyboards (Fig. 4). For example, the hand gesture control system allows doctors to remain in place during an operation, without the need to operate traditional computer-based interfaces [41, 42].

Figure 4. Example of hand gesture recognition in surgical system.

D. Sign Language

A sign language uses body language to convey meaning instead of acoustically conveyed sound patterns. Various vision based gesture recognition methods have been embedded into sign language interpreters [43]. Usually, a capture device is used to find and track hands and record the shapes and trajectories of hands, which are represented by feature vectors. After being matched to corresponding signs, the feature vectors are compared against a grammar library to determine whether the signs make sense in a grammar context. A meaningful grammar context requires reasonable previous and successive signs implication. For static hand gestures, interpreters can reach over 99% accuracy [44, 45]. However, for dynamic hand gestures, it is still difficult to achieve comparable performance and the incorporation of user demographic information and contextual databases may elevate the performance to some extend [46].

E. Robot Control

Venetsky and Tieman [47] proposed a robot gesture recognition system, which allows users to control the robot via hand gestures. The system consists of a robot unit, a video or infrared camera affixed to the robot unit for capturing hand images, a gesture recognition unit and a gesture database. It is also possible to use train robots to learn new gestures in an online or interactive manner [48].

IV. CONCLUSION AND DISCUSSION

Vision based dynamic hand gesture recognition plays an important role in the development of HCI. This paper reviewed some key algorithms and the applications of hand gesture recognition in the real world. In the segmentation module, skin color model, shape feature and background subtraction are the three popular methods to extract the hand region. In the tracking module, algorithms such as Camshift, optical flow, particle filtering, Kalman filtering, TLD have achieved promising results. At last, in the classification module, approaches based on HMM are the dominant techniques. As to applications, hand gesture recognition has been generating significant impact on our everyday life in many aspects.

With the rapid development of sophisticated techniques, a number of successful applications have emerged [49, 50]. However, most of these applications rely on highly restricted situations, such as simple background, normal illumination and static camera. For gesture recognition in the general circumstance, a number of issues are to be addressed, such as complex background, partial-occlusion, background disturbance, object reappearance, illumination change and running in real time. Moreover, the hand is the most flexible part of human body and has high DOFs. It poses additional challenges as it has a variety of appearances and can change quickly. The speed of hand motion itself can be also very fast.

To cope with these challenges, many algorithms can be strategically combined together to collaborate with each other, and encouraging results have been achieved in various cases. For complex background, the fusion of the skin-color model and the shape model can perform better in the segmentation task. A number of new ideas have also been proposed, such as background modeling, self-learning, adaptive object model and historical object template. Note that, similar to the fusion of multiple algorithms, the exploitation of the background information and the object information in previous frames can improve the accuracy of tracking but will inevitably result in increasing computational complexity.

REFERENCES

[1] D. Y. Xu, "A Neural Network Approach for Hand Gesture Recognition in Virtual Reality Driving Training System of SPG," In Proceedings of the 18th International Conference on Pattern Recognition, vol. 3, pp. 519-522, 2006.

[2] W. T. Freeman and C. D. Weissman, "Television Control by Hand Gestures," In Proceedings of International Workshop on Automatic Face and Gesture Recognition, pp. 179-183, 1995.

[3] T. Starner, J. Weaver and A. Pentland, "Real-Time American Sign Language Recognition Using Desk and Wearable Computer Based Video," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20(12), pp. 1371-1375, 1998.

[4] J. Weissmann and R. Salomon, "Gesture Recognition for Virtual Reality Applications Using Data Gloves and Neural Networks," In Proceedings of the International Joint Conference on Neural Networks, vol.3, pp. 2043-2046, 1999.

[5] Y. Wu and T. S. Huang, "Vision-Based Gesture Recognition: A Review," In Proceedings of the International Gesture Workshop on Gesture-Based Communication in Human-Computer Interaction, pp. 103-115, 1999.

[6] R. Y. Wang and J. Popović, "Real-Time Hand-Tracking with a Color Glove", ACM Transactions on Graphics, vol. 28(3), Jul 2009.

[7] V. I. Pavlovic, R. Sharma and T. S. Huang, "Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19(7), pp. 677-695, 1997.

[8] A. R. J. Francois and G. G. Medioni, "Adaptive Color Background Modeling for Real-Time Segmentation of Video Streams," In Proceedings of the International Conference on Imaging Science, Systems, and Technology, vol. 1(121), pp. 227-232, 1999.

[9] D. Chai and K. N. Ngan. "Locating Facial Region of a Head-and-Shoulders Color Image," In Proceedings of the 3rd IEEE International Conference on Automatic Face and Gesture Recognition, pp. 124-129, 1998.

[10] A. A. Argyros and M. I. A. Lourakis. "Real-Time Tracking of Multiple Skin-Colored Objects with a Possibly Moving Camera," In Proceedings of the 8th European Conference on Computer Vision, Lecture Notes in Computer Science, vol. 3023, pp. 368-379, 2004.

[11] H. S. Yoon, J. Soh, Y. J. Bae and H. S. Yang, "Hand Gesture Recognition Using Combined Features of Location, Angle and Velocity," Pattern Recognition, vol. 34(7), pp. 1491-1501, 2001.

[12] H. K. Lee and J. H. Kim, "An HMM-Based Threshold Model Approach for Gesture Recognition", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 21(10), pp. 961-973, 1999.

[13] J. Q. Wang and Y. Yagi, "Integrating Color and Shape-Texture Features for Adaptive Real-Time Object Tracking", IEEE Transactions on Image Processing, vol. 17(2), pp. 235-240, 2008.

[14] K. Oka, Y. Sato and H. Koike, "Real-Time Fingertip Tracking and Gesture Recognition", IEEE Computer Graphics and Applications, vol. 22(6), pp. 64-71, 2002.

[15] A. A. Argyros and M. I. A. Lourakis. "Vision-Based Interpretation of Hand Gestures for Remote Control of a Computer Mouse", In Proceedings of the International Conference on Computer Vision in Human-Computer Interaction, pp. 40-51, 2006.

[16] J. L. Crowley, F. Berard and J. Coutaz. "Finger Tracking as an Input Device for Augmented Reality," In Proceedings of the International Workshop on Gesture and Face Recognition, pp. 195-200, 1995.

[17] J. M. Rehg and T. Kanade. "Model-Based Tracking of Self-Occluding Articulated Objects", In Proceedings of the 5th International Conference on Computer Vision, pp. 612-617, 1995.

[18] F. S. Chen, C. M. Fu and C. L. Huang, "Hand Gesture Recognition using a Real-Time Tracking Method and Hidden Markov Model", Image and Vision Computing, vol. 21(8), pp. 745–758, 2003.

[19] S. Mitra and T. Acharya, "Gesture Recognition: A Survey," IEEE Transactions on Systems, Man and Cybernetics, Part C: Applications and Reviews, vol. 37(3), pp. 311-324, 2007.

[20] D. Fleet and Y. Weiss, "Optical Flow Estimation," Handbook of Mathematical Models in Computer Vision, pp. 237-257, 2006.

[21] H, Zhou, "An OPS Hand Tracking Algorithm Based on Optical Flow," In Proceedings of the WRI World Congress on Software Engineering, vol. 1, pp. 190-194, May, 2009.

[22] K. R. T. Aires, A. M. Santana and A. A. D. Medeiros, "Optical Flow Using Color Information: Preliminary Results", In Proceedings of the 2008 ACM Symposium on Applied Computing, pp. 1607-1611, 2008.

[23] G. R. Bradski, "Computer Vision Face Tracking For Use in a Perceptual User Interface," In Proceedings of the IEEE Workshop on Applications of Computer Vision, pp. 214–219, 1998.

[24] J. C. Peng, L. Z. Gu and J. B. Su, "The Hand Tracking for Humanoid Robot using Camshift Algorithm and Kalman Filter", Journal of Shanghai Jiaotong University, 2006.

[25] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A Tutorial on Particle Filters for On-Line Non-Linear/Non-Gaussian Bayesian Tracking," IEEE Transactions on Signal Processing, vol. 50(2), pp. 174-189, 2002.

[26] Y. Rui and Y. Q. Chen, "Better Proposal Distributions: Object Tracking Using Unscented Particle Filter", In Proceedings of the 2001 IEEE Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. II-786-793, 2001.

[27] M. Isard and A. Blake, "Condensation–Conditional Density Propagation for Visual Tracking", International Journal of Computer Vision, vol. 29(1), pp. 5–28, 1998.

[28] C. F. Shan, T. N. Tan and Y. C. Wei, "Real-Time Hand Tracking using a Mean Shift Embedded Particle Filter", Pattern Recognition, vol. 40(7), pp. 1958-1970, 2007.

[29] Z. Kalal, K. Mikolajczyk and J. Matas, "Tracking-Learning-Detection", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34(7), pp. 1409-1422, 2012.

[30] Hidden Markov Model. [Online]. Available: http://en.wikipedia.org/wiki/Hidden_Markov_model

[31] L. R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," In Proceedings of the IEEE, vol. 77(2), pp. 257-286, 1989.

[32] T. Starner and A. Pentland, "Real-time American Sign Language Recognition from Video using Hidden Markov Models," In Proceedings of the International Symposium on Computer Vision, pp. 265-270, Nov 1995.

[33] M. Elmezain, A. Al-Hamadi, J. Appenrodt and B. Michaelis, "A Hidden Markov Model-Based Continuous Gesture Recognition System for Hand Motion Trajectory", In Proceedings of the 19th International Conference on Pattern Recognition, pp. 1-4, Dec 2008.

[34] D. Minnen, "Fast Fingertip Detection for Initializing a Vision-Based Hand Tracker," U.S. Patent, Application NO. 13/430,509, Oct. 25, 2012.

[35] P. T. Keaton, S. Dominguez and A. H. Sayed, "Vision-Based Pointer Tracking and Object Classification Method and Apparatus," U.S. Patent 7148913, Dec. 12, 2006.

[36] R. T. Held, A. Gupta, B. Curless and M. Agrawala, "3D Puppetry: A Kinect-Based Interface for 3D Animation," In Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology, pp. 423-434, 2012.

[37] S. S. Rautaray and A. Agrawal, "Real Time Hand Gesture Recognition System For Dynamic Applications," International Journal of UbiComp, vol. 3(1), pp. 21-31, Jan 2012.

[38] E. Hildreth and F. MacDougall, "Video Image Based Control System," U.S. Patent 5534917, Jul. 9, 1996.

[39] E. Hildreth, "Optical Flow Based Tilt Sensor," U.S. Patent 7379566, May. 27, 2008.

[40] A. Shamaie, "Tracking Bimanual Movements," U.S. Patent 7379563, May. 27, 2008.

[41] B. Itkowitz, S. P. DiMaio and T. Zhao, "Method and Apparatus for Hand Gesture Control in a Minimally Invasive Surgical System," U.S. Patent Application NO. 12/886,993, Sep. 21, 2010.

[42] N. Frielinghaus, C. Hamilton and W. Steinle, "Gesture Support for Controlling and/or Operating a Medical Device," U.S. Patent Publication NO. 2524279 A1, Nov. 21, 2012.

[43] H. Birk and T. B. Moeslund, "Recognizing Gestures From the Hand Alphabet Using Principal Component Analysis," Master Thesis, Laboratory of Image Analysis, Aalborg University, Denmark, 1996.

[44] S. Meena, "A Study on Hand Gesture Recognition Technique," Master Thesis, Department of Electronics and Communication Engineering, National Institute of Technology, Rourkela, Orissa 769 008, India,2011

[45] R. Lockton, "Hand Gesture Recognition using Computer Vision," [Online]. Available:http://research.microsoft.com/en-us/um/people/awf/bmvc02/project.pdf

[46] J. Tardif, "Machine Based Sign Language Interpreter," U.S. Patent, Application NO. 12/794,455, Jun. 4, 2010.

[47] L. Venetsky and J. W. Tieman, "Robotic Gesture Recognition System," U.S. Patent 7606411, Oct. 20, 2009.

[48] C. Lee and Y. S. Xu, "Online, Interactive Learning of Gestures for Human/Robot Interfaces," In Proceedings of the IEEE International Conference on Robotics and Automation, vol. 4, pp. 2982-2987, 1996.

[49] J. P. Wachs, M. Kölsch, H. Stern and Y. Edan, "Vision-Based Hand-Gesture Applications," Communications of the ACM, vol. 54(2), pp. 60-71, 2011.

[50] S. S. Rautaray and A. Agrawal, "Vision Based Hand Gesture Recognition for Human Computer Interaction: A Survey," Artificial Intelligence Review, Published Online: 06 November, 2012.