STM JOURNALS

Journal of Artificial Intelligence Research & Advances
ISSN: 2395-6720(online)
Volume 2, Issue 3
www.stmjournals.com

Vision Based Gesture Recognition Techniques: A Review

Avinash S Pathare*, Supriya O Rajankar
Department of Electronic and Telecommunication, Sinhgad College of Engineering, Vadgaon (Bk), Pune, India

Abstract

With the ever-growing use of computers in the society, it is clearly seen that present popular mode of interactions with computers is mouse and keyboard. Mouse and keyboard will become an inconvenient way to convey information to the computers. Nowadays there are intelligent systems in which the flow of information takes place without any physical contact. The vision based (camera) gesture recognition has the potential to be a natural and powerful tool supporting efficient and spontaneous interaction between the human and the computer. Visual interpretation of hand gestures can help in achieving the ease and naturalness desired for Human-Computer Interaction (HCI). HCI has driven many researchers in computer vision-based analysis and understanding of hand gestures as a very active research area. This paper reviews some major developments and the recent evolution in the field of hand gesture recognition techniques. While providing a Non-Exhaustive catalog of the huge amount of past research in the field, the paper reviews in more detail vision-based approaches. We will discuss several traditional and new applications of gesture recognition in this paper. We surveyed the literature on the visual analysis of hand gestures in the context of its role in HCI and emphasized numerous important works of researchers. The tenacity of this review is to introduce the field of gesture recognition as a mechanism for interaction with computers. The aim of hand gesture recognition researchers is to develop such a system that will easily identify gestures and use them for a variety of application ranging from sign language to virtual reality.

Keywords: *Hand gestures, human-computer interaction, model-based approach, recognition, vision-based approach*

****Author for Correspondence*** E-mail: avinashpathare3@gmail.com

INTRODUCTION

We perform various gestures in our daily lives. It is the nature of human to use gestures to improve the communication. It is very difficult to understand if the message without a gesture. Between all kind of gestures such as facial, hand and sometimes body we can say that hand gestures play an important role to convey information.

In these days, computers have become an imperative part of our lives, so why not use a hand gesture to communicate with them. The usage of the hand as an input device is a very popular method for providing natural HCI. The use of hand gestures offers an attractive alternative to these bulky interface devices for HCI. The data glove used is an electronic device with motion capturing sensors, such as flex sensors, capturing the movements of each finger in analog form and converts them to digital using the analog-to-digital converter. Also, this method employs mechanical or optical sensors attached to a glove.

Fig. 1: *Data Glove.*

The transducers convert finger-bending Figure 1 angle into electrical signals for detection of the hand posture. It does the angular measurement of hand bones and provides data based it. However, there are several disadvantages that make this technology not so popular. Firstly, it makes interaction with the computer-controlled environment loses its naturalness and acceptance, the user is needed to carry a load of cables, which are connected to the computer and it also needs calibration and arrangement procedures.

Systems that are Vision Based: Computer vision based techniques have the potential to provide more natural and noncontact solutions. They are non-invasive and are based on the way human beings perceive information about their surroundings.

For the vision-based gesture recognition, the compound backgrounds, changing lighting conditions and a deformable human hand shape can cause wrong gesture recognition. The developed system also needs to satisfy other limitations, such as the size of the hand gesture in the image and variation to motion. Although, it is hard to design a vision-based interface for general usage, yet it is possible to design such an interface for a human controlled environment but has many challenges including accuracy and processing speed.

GESTURE RECOGNITION

In this paper, we consider vision-based gesture recognition technique as a main part of our discussion. It is a challenging research area, which involves Computer Vision (CV) as well as graphics, image processing, machine learning, bioinformatics and psychology.

Using visual input makes it possible to communicate remotely with electronic equipment, without the need for physical contact. To make a system work there are some requirements for the system to have:

Computational Efficiency

Generally, Vision-Based interaction frequently requires real-time systems. The vision and learning algorithms used in the vision-based interaction should be effective as well as cost efficient.

Robustness

The visual information could be very rich, noisy and incomplete, due to changing illumination, cluttered backgrounds and occlusion. The vision-based systems should be user independent and robust against all these factors.

Scalability

The Vision-based interaction system should be easily adapted to different scales of applications. The main part of the Vision-based communication should be the same for desktop situations, Sign Language Recognition (SLR), robot navigation and also for virtual equipment development.

User’s Tolerance

The failures or mistakes of Vision-based interaction should be tolerated. When we make a mistake, it should not incur much loss. Sometimes, the users can be asked to repeat some actions repeatedly, instead of that let the computer make more incorrect decisions.

Most of the systems reviewed depend on the mere idea of detecting and segmenting the gesticulating hand from the background using skin color or motion detection. According to Wacs *et al.* proper selection of features and their combination with sophisticated recognition algorithms can disturb the success or failure of any existing field of HCI using hand gestures [1].

The approaches to Vision-based hand gesture recognition can be divided into two categories: appearance based approaches and model-based approaches.

Appearance-Based Approach

Appearance-based models are derived directly from the data of the images and usually used for gesture detection. No explicit model of the hand is compulsory, which means that we have no particular model for internal degrees of freedom. When only the presentation of the hand in the video frames is recognized, differentiating between gestures is not as simple as with the model based approach. The gesture recognition will therefore, typically involve some statistical classifier built on a set of features that represent the hand. In many gesture applications, all that is mandatory is a mapping between gesture and input video. Therefore, many have argued that the full restoration of the hand is not crucial for gesture recognition. As a substitute, many approaches have used for the extraction of low-level image dimensions that are fairly robust to noise and we can extract rapidly. Low-level features that have been anticipated in the literature include: the hand centroid region [2]. Also, an elliptical bounding region of the hand with principal axes and the optical affine flow of the hand region in an image [2].

There exist another approach is to look for skin colored regions in the input image. This is a very widespread method but has some disadvantages [3–5]. First, skin color detection is very subtle to lighting conditions. While there exist some achievable and effective methods for skin color detection under known illumination. The problem of learning a supple skin model and upgrading it over time is a challenging task. Lindberg used scale-space color features which can be seen in an image at different scales. Therefore, the hand can be described as one biggest blob feature for the palm. It is having smaller blob features signifying the fingertips that are connected by certain rigid features. This method was proposed to perform the feature extraction directly in the defined color space. This method allows the combination of probabilistic skin colors in the feature extraction phase.

Another method is to use the Eigen space for gesture detection. At a start, a defined set of images and Eigen space approaches construct a small set of basis images. These images characterize the variation of images in the training set and can be used to estimate any of the training images. Now, to reconstruct an image in the training set, a linear grouping of the basis vectors is taken. In grouping the coefficients of the basis vectors is the outcome of projecting the image to be reconstructed on to the individual basis vectors. In, an approach for tracking hands by Eigen space improvements to the unique Eigen space approach formulation [6]. It is a large invariance to occlusions, some invariance to variations in the background from the input images.

AdaBoost learning system with Scale-Invariant Feature Transform (SIFT) features is used [7]. The SIFT introduced by Lowe consists of a histogram representing gradient and magnitude information within a small patch of the same image [8]. SIFT is a rotation and scale-invariant feature and is robust to some deviations of illuminations, viewpoints and noise. The sharing feature concept improves the accuracy of multiclass hand posture recognition. However, different features such as contrast context histogram need to be studied and applied to accomplish hand posture recognition in real time. To demonstrate the power of their algorithm, three hand posture classes, “fist”, “palm,” and “six,” are trained and recognized. For testing, about 275 images were collected using the onboard QuickCam Pro 5000 with a resolution of 640x480. By, using the feature sharing concept, the recognition rate enhances from 90.9% to 95.6%.

Haar-like features are used for the task of hand detection [9, 10]. Haar-like features focus more on the information in a certain area of the image rather than a single pixel. To gain real-time performance, the AdaBoost learning algorithm is used, it also increases classification accuracy. The algorithm can adaptively choose the best features in each step and combine them into a robust classifier, can be used. The training algorithm constructed on AdaBoost learning algorithm takes a set of “positive” samples. The samples contain the object of interest and a set of “negative” samples, i.e., images that do not contain objects of interest.

Color likelihood maps and Maximally Stable Extremal Region (MSER) detector are used for hand tracking [11]. Such a grouping allows performing repeated ground segmentation in each frame in an efficient manner. The MSER detector is one of the best interest regions detectors in computer vision [12]. MSER detection is mostly applied to single grayscale images, but the method can be easily protracted for analysis of color images by defining an appropriate ordering relationship on the color pixels. The MSER detector discovers bright connected regions that have darker values along their boundaries subsequently. The set of MSERs is closed under constant geometric transformations and is not variant to affine intensity changes. Also, detect MSERs at all scales. So, due to these properties MSER detection is suited for segmentation purposes. The proposed method was implemented in C++ and runs at 25 frames per second on 320×240 video sequences.

Model-Based Approach

Model-based approaches create model assumptions and estimate them on the available visual interpretations. Essentially, this is achieved by articulating an optimization problem. Its aim function measures the inconsistency between the chromatic cues that are expected due to a model assumption and the real ones. The employed optimization technique must be able to evaluate the objective function at random points in the multidimensional model parameters space. By large, the methodology consists of searching for the kinematic parameters that brings the 2D projection of a 3D model of the hand into correspondence with an edge-based image of a hand.

In Figure 2 the CMC joints are supposed to be fixed, which quite idealistically models the hand palm as a rigid body part. Model, the hand fingers as serial kinematic chains devoted to the palm at anchor points placed at MCP joints. In recent years, the kinematic model was enhanced by adding extra twist gesticulation to MCP joints, by introducing one extension DOF to CMC joints [13]. They have used 27 DOF to create a 3D model for their DigitEyes system. They proposed a model based hand tracking system, which can be able to recover the 27 DOF states from ordinary grayscale images at speeds of up to 10 Hz.

Heap *et al.* proposed a deformable 3D hand model and modeled the entire surface of the hand by a surface smash constructed via PCA from training examples [14]. Real-time tracking is achieved by finding the closest possibly deformed model matching with the image. Such an illustration requires further processing to extract useful higher-level data, such as pointing direction; though, it was shown to be effective to locate reliably and track the hand in images. The technique, however, is not able to handle the blocking problem and is not scale and rotation invariant.

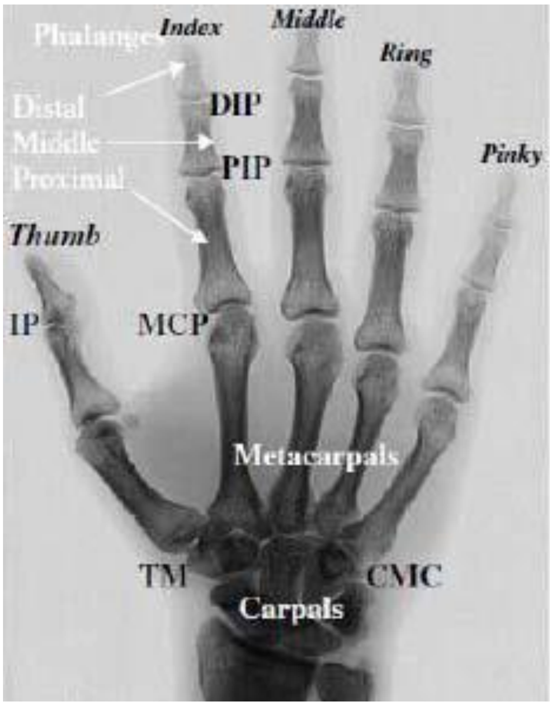

Fig. 2: *Skeletal Hand Model for Kinematic Model [13].*

Stenger *et al.* have used quadrics model as shape primitives in [15]. The usage of quadrics to build the 3D model produces an elegant and practical method for making the contours of the model, which are then equated with the image data. The posture of the hand model is predicted with an Unscented Kalman filter (UKF), which decreases the geometric error between edges extracted and the profiles from the images. The use of the UKF permits higher frame rates than more sophisticated estimation methods such as particle filtering while providing higher precision than the extended Kalman filter. In, they have formulated the problem within a Bayesian framework [16]. Bayesian algorithms allow for the pooling of multiple sources of information to attain at both an optimal evaluation of the parameters and a probability distribution of the parameter space, to guide the future search for parameters. Against the Kalman filter approaches, Bayesian approaches allow nonlinear system formulations and non-Gaussian (multi-modal) uncertainty.

APPLICATIONS

Robotics and Telepresence

At the start when robots are introduced into our daily lives, they have to face many challenges such as collaborating with humans in complex and ambiguous environments or upholding long-term human-robot relationships. Both telepresence and telerobotic applications are considered in the domain of space exploration and military-based projects. The gestures used to control and interact with robots are similar to fully-immersed virtual reality interactions with machines. Though the worlds are often real, giving the operator with the video feed from cameras installed on the robot [17]. Here, gestures can control a robot's hand, camera and arm movements to reach and operate on actual objects, as well its movement over the world.

Sign Language

It is considered as a significant case of communicative gestures. We know that sign languages are highly structural; they are very suitable as testbeds for vision based systems [17].

Also, it can also be a better way to assistance the disabled person to interact with machines (Computers). Sign language for the deaf is a good example in U.S. that has received substantial attention in the gesture literature [18, 19].

Virtual Reality

The hand gestures for virtual and enlarged reality applications have experienced one of the extreme levels of acceptable interactions or 2D displays that mimic 3D interactions.

Vehicle Interfaces

Some hand gesture recognition techniques for a human vehicle interface have been proposed in [20]. The motivation hand gestures in vehicles are from in-vehicle secondary controls. It is widely based on the principle that taking the eyes off the way to operate conventional secondary controls can be abridged by using hand gestures.

Desktop and PC Applications

In PC/desktop computing applications, gestures can provide substitute interaction methods to the mouse and keyboard. Many gestures for desktop computing tasks involve working graphics or interpreting and editing documents using pen-based gestures [21]. Recently perception introduced gesture recognition Technology for Windows-based Portable Computers and Android Tablets.

Games

If, we look at gestures for computer games. Freeman *et al.* tracked a player's body or hand position to control orientation and movement of interactive game objects such as cars, bat and ball [22]. Konrad *et al.* used all body gestures to control the avatar's movements in a virtual reality [16]. Also, the Play Station 2 has presented the Eye Toy, a camera that tracks complete hand movements for interactive video games.

Hospitals and Healthcare Centers

Wachs *et al.* developed a hand-gesture recognition system that facilitates doctors to use digital images virtually [23]. These digital images captured by medical equipment during medical procedures using hand gestures instead of computer keyboards or touch screens. A spotlessly clean human-machine interface is of more importance because it is the means by which the doctor/surgeon controls medical information about that patient, avoiding patient infection, the operating room and the other surgeons. The gesture based system could substitute touch screens now used in many hospital operating rooms. It will prevent increase or spreading of contaminants and wants even surfaces that must be thoroughly cleaned after each process-even that sometimes they are not. With infection rates at hospitals now at unacceptably high rates, the hand gesture recognition system offers a possible alternative.

TYPES OF GESTURES

The gesture used for HCI applications classifies hand gestures into three groups: communicative gestures, controlling gestures, manipulative gestures:

Communicative Gestures

These are projected to express an idea or a model. These gestures are either used together with words or are a substitute for vocal communications. Communicative gestures require a high structured set of gestures such as those well-defined in sign languages.

Controlling Gestures
These types of gestures are used to control a point or system and locate an object. Finger Mouse is an example application which detects 2D finger controls mouse movements on the computer/tablet. Analyzing hand gestures is completely application dependent and involves analyzing the hand motion, showing hand and arm, drawing the motion features to the model and reading the gesture in a predefined time interval.

Manipulative Gestures
These are used for interaction with things in an environment. These gestures are typically used for communication in virtual environments. The physical objects can be operated through gesture controlled robots in a real world.

A list in which hand gesture association with speech declines, language properties increase, impulsiveness decreases and social instruction increases can be seen:

Gesticulation
It is a spontaneous movement of hands and arms, associated speech, these movements constitutes about 90% of human gestures. People gesticulate when they are talking to someone on the telephone and even blind people regularly gesture when speaking with each other.

Language-like Gestures
The gesticulation integrated into a spoken utterance is looks like a language, by replacing a particular spoken word.

Pantomimes
The gestures are depicting objects or actions, with or without accompanying the language.

Emblems
It represents familiar signs such as “V for victory,” or OK, or Thumbs Up.

CONCLUSION
In this paper, we have presented a review of various vision-based hand gesture recognition methods. In these years, there is remarkable growth in the field of vision based hand gesture recognition methods, from software as well as hardware point of view, has been done. For advancement and to continue this growth, it is clear that further research in the areas of feature extraction and classification methods is required. Also, gesture representation is needed to study, to realize the ultimate goal of humans interfacing with machines on their natural terms. A newly developed cameras gives us depth information about the object which is now included in algorithms. The local invariant features permit us to use methods which may enclose Gestalt laws of visual perception. The Gestalt laws are a set of visual rules that direct the building process of groupings, creating causal relationships between grouping elements, and tends to emulate well. It is surely seen that in future hand gesture recognition will play an important/major role in our lives. It is clearly seen from the documentation on gesture recognition that there are some problems of gesture recognition. A main problem impeding most of the approaches is that they rely on several fundamental assumptions that may be appropriate in a controlled lab setting but do not generalize to arbitrary contexts. Some common assumptions are such as: assuming ambient lighting conditions and high contrast stationary backgrounds. Also, the results are dependent on every researcher's data collection and methodology. It is impossible for us to compare the parameters as everyone is assuming the parameters in different ways.

REFERENCES
1. Konstantinos G. Derpanis. A review of vision-based hand gestures. 2004; 1–18p
2. New JR, Hasanbelliu E, Aguilar M. Facilitating user interaction with complex systems via hand gesture recognition. *In Proc of ACM Conf*. 2003.
3. Mo Z, Lewis JP, Neumann U. Smartcanvas: a gesture-driven intelligent drawing desk system. *International Conference on Intelligent User Interfaces*. 2005; 239–243p.
4. Martin J, Devin V, Crowley JL. Active hand tracking. *3rd. International Conference on Face & Gesture Recognition in IEEE*. 1998; 575p.
5. Kjeldsen R, Kender J. Toward the use of gesture in traditional user interfaces. *International Conference on Automatic Face and Gesture Recognition in IEEE.* 1996; 151–156p.

6. Bolt RA. Voice and gesture at the graphics interface. *Proc. SIGGRAPH80*. 1980; 14(3): 262–270p.
7. Wang CC, Wang KC. Hand posture recognition using adaboost with SIFT for human-robot interaction. *Springer*. 2008; 370: 317–329p.
8. Lowe DG. Object recognition from local scale-invariant features (SIFT). *Proc. of the International Conference on Computer Vision*. 1999; 2: 1150–1157p.
9. Barczak ALC, Dadgostar F. Real-time hand tracking using a set of co-operative classifiers based on haar-like features. *Res. Lett Inf. Math and Sci*. 2005; **7**: 29–42p.
10. Chen Q, Georganas ND, Petriu EM. Real-time vision-based hand gesture recognition using haar-like features. *IEEE Trans. on Instrumentation and Measurement.* 2007; 1–6p.
11. Donoser M, Bischof H. Real-time appearance-based hand tracking pattern. *In a Conference of ICPR.* 2008; 1–4p.
12. Mikolajczyk K, Schmid C, Zisserman A et al. A comparison of affine region detectors. *In International Journal of Computer Vision*. 2005; 65(1/2): 43–72p.
13. Bay M, Koller-Meier, Gool LV. Smart particle filtering for 3D hand tracking. 6th *IEEE Conference on Automatic Face and Gesture Recognition*. 2004; 675p.
14. Heap AJ, Hogg DC. Towards 3-D hand tracking using a deformable model. *In 2nd International Conference on Face and Gesture Recognition.* 1996; 140–145p.
15. Stenger B, Mendon PRS, Cipolla R. Model-based 3D tracking of an articulated hand. *Proc. In British Machine Vision Conference.* 2001; 2: 63–72p.
16. Stenger B, Thayananthan A, Torr PHS et al. Model-based hand tracking using a hierarchical bayesian filter. *In IEEE Trans. on Pattern Analysis and Machine Intelligence.* 2006; 28(9): 1372–1384p.
17. Rehg J, Kanade T. Visual tracking of high DoF articulated structures: An application to human hand tracking. *European Conference on Computer Vision and Image Understanding.* 1994; 801: 35–46p.
18. La Gorce M de, Paragios N, Fleet DJ. Model-based hand tracking with texture, shading, and self-occlusions. *In IEEE Conference on Computer Vision and Pattern Recognition*. 2008; 1–8p.
19. La Gorce M de, Fleet DJ, Paragios N. Model-based 3D hand pose estimation from monocular video. *In IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2011; 33(9): 1793–1805p.
20. Dong Guo Yonghua. Vision-based hand gesture recognition for human-vehicle interaction. *International Conference on Control, Automation and Computer Vision*. 1998.
21. Smith GM, Schraefel MC. Scrolling support for stylus/touch-based document navigation. *In Proc. seventeenth ACM Symposium on User Interface Software and Technology*. 2004; 53–56p.
22. Jinshi Cui, Zengqi Sun. Model-based visual hand posture tracking for guiding a dexterous robotic hand. *Optics Communications.* 2004; 235: 311–318p.
23. Wachs PJ, Stern HI, Edan Y et al. A gesture-based tool for sterile browsing of radiology images. *A Journal of the American Medical Informatics Association*. 2008; 15(3): 321–323p.

AUTHOR BIBLIOGRAPHY

Mr. Avinash S. Pathare is a student of Sinhgad College of Engineering, Pune for his post-graduation in Digital Systems. He is graduated in electronics and telecommunication from Vidya Pratishthan's College of Engineering, Baramati Pune. His main area of research is image processing and embedded systems. He is recently working on research of object detection and gesture detection.

Mrs. Supriya O. Rajankar is Associate Professor in Department of Electronics and Telecommunications Engineering at Sinhgad College of Engineering, Pune. She earned her master's degree from Government College of Engineering, Pune with the specialization in Digital Systems. She is a life member of Indian Society for Technical Education (ISTE) and a member of Institute of Electronics and Telecommunication Engineers (IETE).She has experience of more than 20 years in academics. She has a credit of being an author of two books published by John Wiley Publication. She has taught various subjects like Advanced Digital System Design, Network Theory, Electromagnetics and Digital Signal Processing. Her main area of interest is in the field of Signal Processing and Image Processing. She is currently working on a research project in the field of Signal Compression.

Cite this Article
Avinash S. Pathare, Supriya O. Rajankar. Vision Based Gesture Recognition Techniques: A Review. *Journal of Artificial Intelligence Research & Advances*. 2015; 2(3): 1–8p.