

# Vision Based Interface in Human Computer Interaction

Aditya Verma, Ankita Verma

Delhi, India

## ABSTRACT

This paper provides information on Human Computer Interaction, it actually means that, its present use in the world of technology and its importance in the tech- savvy world of today. The paper goes through the various types of interaction that can take place between a human and a computer, the technicalities of the same, and the future scope of each interface. We always want a path of communication which is fast, is efficient and is user friendly which gives us maximum output and good performance consistently.

Vision based interface between human and computer is studied in detail. It can be stated as the most popular types of interaction between human beings and computers. Vision based interaction in HCI makes use of four main techniques; gesture recognition, eye movement recognition or tracking, head tracking and facial expressions recognition and judgement. Analysis for their usage in practical systems has been made. This is the most worked upon area is the vision based hand gesture recognition which has been discussed in this paper. The present studies and key findings in this area have been listed and its future scope and utility has also been discussed in this paper.

**Keywords :** HCI, Human Computer Interaction, Science, AI, Vision Based

## I. INTRODUCTION

Human Computer Interaction is basically the study of the interaction between the machine (computer) & human beings and how it takes place. It is the subject involving the study, design, construction, implementation and the use of computer technology for efficient communication between the computer and man. It studies the different methods, modes and types of communication or interaction between two independent systems (computer and human) across a boundary also known as an interface [1].

One such user interface is the vision-based interaction which is based on the intelligence of the computer to communicate visually with man through techniques like head tracking, eye movements, gesture recognition and interaction through facial expressions. It is one of the most important and popular interface

which enhances the level of interaction with the machine.

## II. OBJECTIVE

The clear objective of this paper is to study the various developments made in the field of human computer interaction, particularly vision based interaction between the two systems. To study the practical utilities of the interface keeping in mind its scope in the future and list the key areas of study, key findings, conclusions and discussions regarding the topic.

## III. RESULTS AND DISCUSSION

Human Computer Interaction works around the communication between the two independent systems across a boundary. It is the study of the way in which computer technology influences and helps human work and activities. The main focus of research in HCI

is to design and implement such an interface or mode of interaction which is easy and productive to use. It should satisfy our needs and should be effective & efficient. With the increase in the usage of computer systems, the computer systems would be embedded into our environment [3].

As defined by ACM, Human Computer Interaction is a branch co-related with the design, assessment and execution of interactive computing systems for the use of human and with the major phenomenon surrounding them [4]. As HCI is a concept revolving around humans and machines both, thus human knowledge as well as technical machine knowledge is required for its study as the subject is mixture of social sciences, communication skills, graphic and industrial designs, linguistics and operating systems, computer graphics and languages.

### 3.1 WHAT IS AN INTERFACE?

There are 3 types of interface in computer technology [5]:

- Software Interface-It consists of the coding that helps the applications to communicate with each other as well as the hardware.
- Hardware Interface-It consists of the parts that are used by hardware devices and helps in communicating with each other. Eg. Plugs, Sockets etc.
- User Interface-It enables the user to interact with the operating system. Eg. Keyboard, mouse etc.

#### 3.1.1 USER INTERFACE

The user interface is the most essential part of HCI (Human Computer Interaction). It is the component that handles the interaction between a human and a machine. It gives us the power to make the program do as we want. [6] The easiness with which we are able to get our purpose from the program is what signifies the efficiency and affectivity of the user interface. The interface may consist of commands or a menu. For a user to easily interact with the given program, the commands should be given in a command-driven

interface & command choices should be selected from various options in a menu for a menu driven interface.

### 3.2 USER INTERFACE & LOOP OF INTERACTION

3.2.1 User Interfaces are majorly classified into two categories:-

- Single User Interface is that in which only one user can access or use an operating system at the same time.
- Multi User Interface is that in which multiple users can use the operating system at the same point of time.

3.2.2 The flow of information or the communication between a human and the computer is known as 'Loop of Interaction' [7]. The various components or aspects related to the loop of interaction are:-

- Audio Based
- Video Based
- Task Environment: It includes the conditions set around or on the user.
- Machine Environment: It is the surrounding to which the machine (computer) is connected to.
- Areas of Interface: It includes the overlapping areas which obtain to the processes of the interaction.
- Input Flow: The information that flows the computer from the task environment.
- Output: The information that is provided from the machine environment.
- Feedback: Feedback includes loops that pass from the user through the interface to the computer and back. During this flow, the loop evaluates, moderates and confirms various processes.
- Fit: It is the race between the user, computer design and the task to optimize the human resources for the accomplishment of the task.

There are a number of user interfaces. The most popular user interface is the Graphical User Interface (GUI) which accepts the input via keyboard, mouse etc. and provide a graphical output on the monitor [8]. Apart from GUI there are other interfaces such as Direct manipulation interface, Web based user

interface, Touch user interface, Conversation interface, Vision based interface, Intelligent user interfaces, Object oriented user interfaces, Voice user interface and many more.

These various types of interfaces are designed on the principles and knowledge of ergonomics, computer science, human factors and usability engineering.

### 3.3 VISION BASED INTERFACE

Day by day efforts are being made to develop such interfaces which ease the communication between the computer and man. Researches are trying to develop a perfect interface that is not only easy to operate but is also efficient. They are trying to increase the usability of interfaces. Today, a variety of interfaces are constructed by designing them to suit a particular purpose having the most desired property and quality for the same. What the researches seek to achieve varies according to the need and the purpose.

In today's tech savvy world, computers are an important and embedded part of our environment. With such increasing role of computers in our environment, there is a need for interfaces that are easy to use and natural, with which we can interact with the computer without the need of special parts and equipment like the keyboard, mouse, remote controls, joysticks, trackballs, data gloves etc. [9]. But what has grasped researchers' attention is the fact that in our daily life, we humans communicate with each other through our senses, through our vision, through the voice and hearing. So why not communicate with the computer through these mediums? For this, special features and interfaces have to be designed for easier and effective human computer interaction. Computers must thus be given the ability so that they understand visual and auditive input.

Thus researchers are now developing interfaces based on visual input with a try to replace traditional interfaces through keyboards, mouse etc. Vision based interaction has a lot many advantages over other types of interfaces. One can visually communicate with the

computer from a distance without having physical contact with the equipment. Moreover, in environments that are noisy or not very quiet, visual communication is better than communication through speech [9]. Studies are being made extensively in this field and new techniques are being developed under vision based interface for effective and efficient human computer interaction. The main techniques that are being explored by scientists and researchers are interaction through gesture recognition, eye tracking methods, head movement tracking and interaction through facial expressions. The area which has been studied the most is hand gesture recognition.

Keeping this in mind, researchers from Sweden are focusing on developing algorithms that help in gesture recognitions and they are working on the project by combining CVAP's expertise in computer vision with CID's experience in the designing and evaluation of new interfaces [9].

### 3.4 GESTURE RECOGNITION

As computers are becoming an integrated part of our lives today, the target is to bring human computer interaction to a point where the communication between man and machine is much simpler, easier and much more real. Thus researchers have thought of incorporating hand gestures as a way to interact between the two independent systems. This is because hand gestures have long been considered to be a part of human speech and communication and they fulfil our needs for a more natural, creative and intuitive method to communicate with computers [10].

Gestures can be basically defined as movements of the hand that convey or intent to convey some information. They form an important part of non-verbal communication. Gesture recognition involves tracking and detecting human hand gestures and then relating them to some meaningful commands through conversion.

As mentioned in FIGURE 1, the only method today that uses the hand as a direct input is "glove based sensing". But it has many disadvantages. Data gloves

worn on the hand help locate the position of the hand and the finger joint angles. They help importing all the information of the functionality of the hand through position measurements. Such devices that help in interacting through a physical contact of user are known as contact based devices. But with efforts to make the process realistic, vision based gesture recognition techniques are being developed. Two directions of research have emerged. First is hand gesture classification and the second is pose estimation systems. The first one aims to extract information related to motion patterns of the hand whereas the second one aims to capture 3 dimensional hand movements [11]. The user can also learn some personalized gestures through multimodal intuitive media browsing.

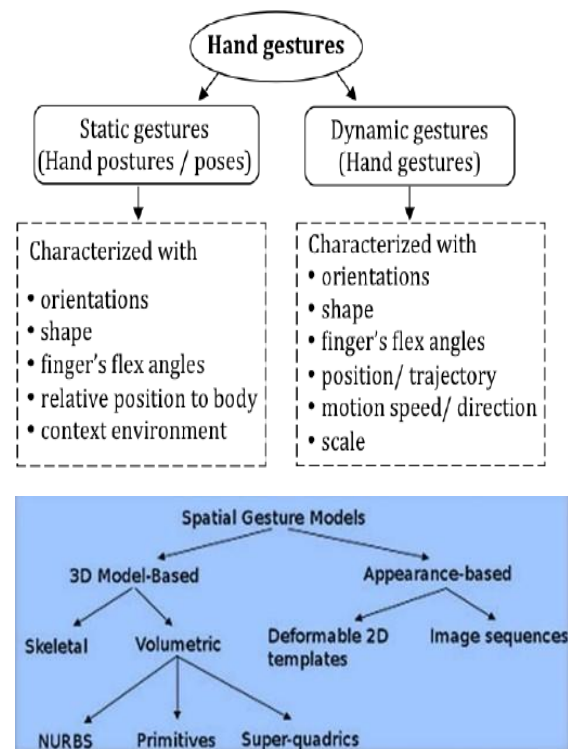


**Fig. 1** Hand Gesture Recognition, Retrieved from [12] The devices that work on the principles of vision based gesture recognition are based on captured video sequence by many cameras for the analysis and study of motion. Hand markers have also been invented for this purpose. These help in the detection and recognition of hand gestures by either shine or flash techniques [10]. For effective communication through this technique, a good knowledge and judgement regarding the consideration of various degrees of freedom, camera positioning, the two dimensional constraint needs to be taken care of. Also the computer system should be enabled with the power to analyse and recognize the image at the rate of the input video so that it is able to provide instant feedback to the user. Moreover the system needs to be made in a way such that it recognizes the gestures of a number of people or in other words, it should be user independent. It

should not be designed to recognize and respond to the hand gestures of only one user.

**3.4.1 GESTURE TAXONOMIES**

Because the meaning of hand gestures is interpreted differently in different cultures, thus for ease hand gestures have been divided into two types, ‘static and dynamic gestures’. Static gestures are those in which the orientation of the hand remains the same during a particular amount or interval of time while dynamic gestures are those in which the position of the hand involves movement in the time interval. Gestures which are significant and typical of emotions are called affect displays. While emblematic gestures are a substitute for short verbal communication. As represented in FIGURE 2, hand gesture representation is also divided into two categories; 3D model based and appearance based representation.



**Fig. 2** Difference in Static and Dynamic gestures Retrieved from [13]

The mechanism that enables vision based hand gesture interaction comprises of three main phases (FIGURE 3) - Detection, tracking and Recognition.

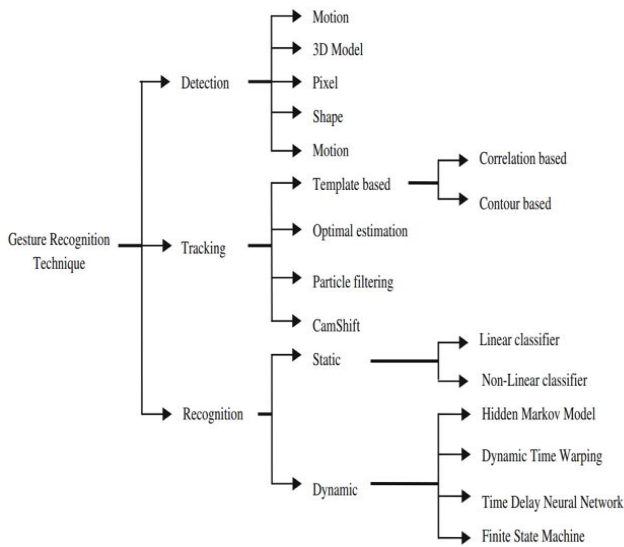


Fig. 3 Gesture Recognition Technique, Retrieved from [10]

### 3.4.2 DETECTION

The basic step of any interaction is detection. For the computer to understand what the gesture means, it first needs to detect it. In this stage, task relevant data is separated from other image backgrounds and then passed to the tracking and recognition stages.

A variety of segmentation techniques are required involving the skin colour, shape, motion etc. For the skin colour, the luminance component is removed, eliminating shadows and other effects. By background subtraction and illumination this problem is solved. The typical or characteristic shape of hands enables us to identify them in an image and extract information through contour extraction of the other objects. Occlusions pose a problem to hand identification or in general identification of 2D objects. Fingertip identification [14] techniques and shape context descriptor [15] techniques are being used for easier detection. Hand appearance also poses a serious problem for detection. But studies have proved that through a combination of weak classifiers and images,

we can derive a strong classifier through the process of boosting, hence this problem has been solved through algorithms [10]. Eg. AdaBoost algorithm (Given by Freud & Schapire) & FloatBoost algorithm. 3 Dimensional models and motion detection serve as cues for the efficient detection of hand gestures.

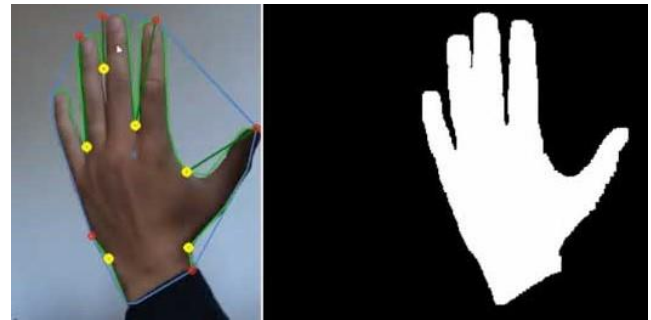


Fig. 4 Hand Gestures, Retrieved from [16]

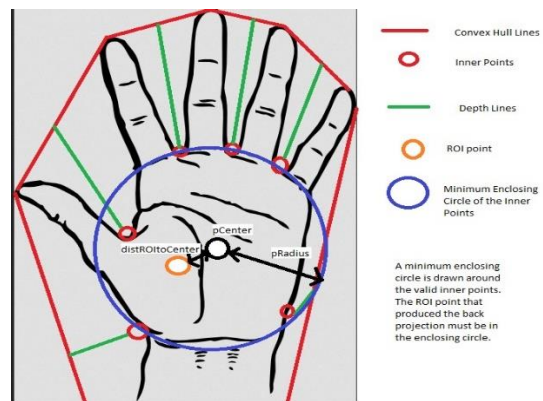


Fig. 5 Hand Gesture, Retrieved from [18]

### 3.4.3 TRACKING

The detection should be really fast so that tracking takes place properly. If the detection takes place at the image acquired frame rate, then tracking can be done efficiently. But the gesture can change very fast within a frame or a few frames. Fast or robust tracking enables us to interlink different appearances in frames so that the trajectories drawn help us get useful information regarding the gesture. In model based detections, robust tracking helps estimate the various model parameters which may not be observed before. Tracking is thus defined as the understanding of hand movements and gestures through frame to frame

correspondence of the segmented and identified hand regions [10]. If images are captured and detected frequently enough, it becomes easier to track the gesture. This forms the basis of template based tracking. Correlation based feature tracking and contour based tracking are the most popular techniques being studied. In the first, the hand is at first detected in a frame. Further, the regions in which the hand appears (image regions) are captured. Using these regions as a prototype and assuming that hands will be next identified in the same spatial neighbourhood, the hand is easily detected in the next frame. On the other hand, in contour tracking, contours or snakes are used to track the regions where the image of the hand is identified. The boundaries to these regions are then obtained by the colour gradient. The main advantage of using the contour method is that they exhibit better performance by handling more than one target at a time and allow real time tracking not only for some basic gestures but also for complex postures. But this method is mostly used when there is a considerable contrast between the image (object) and the background because contours or snakes do not perform well in dark backgrounds.

Optimal estimation technique is a major breakthrough in the tracking area of vision based hand gesture interaction. It enables our machines with real time tracking, performance and gives the system the power to deal with uncertainty and making logical predictions. Algorithms like CONDENSATION, CamShift have opened up many opportunities in the tracking technique of hand gestures.

### 3.4.4 RECOGNITION

Recognition is basically understanding or interpreting what a hand gesture, position or posture means or conveys. It is the most important part of the interaction. As already discussed, hand gestures are categorized into 2 main types; static and dynamic gestures. For the detection of static gestures, classifiers are used. But for dynamic gestures, Hidden Markov Models (HMM) are

used because simple classifiers can't detect hand gestures that are dynamic because of we take another dimension into consideration in dynamic gestures.

Learning algorithms are all different from one another and designed in such a way that they can be used to recognise a particular class or type of gesture. Its choice depends on the hand gesture representation. Automata based methods are being used along with various learning algorithms. But the main disadvantage of automata method is that whenever a new gesture is introduced and needs to be recognized, the whole system needs to be modified in accordance with the new criteria and gestures. The complexity also rises with increase in the number of gestures to be identified and introduced in the learning algorithm.

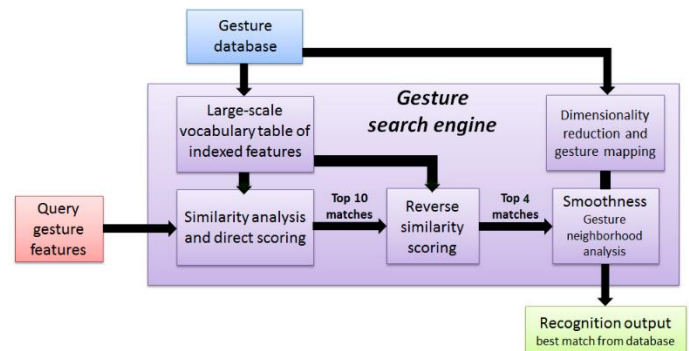


Fig. 6 Gesture Search Engine, Retrieved from [19]

### 3.4.5 DISADVANTAGES OF VISION BASED INTERFACE HAND GESTURE RECOGNITION SYSTEMS

No doubt, vision based hand gesture recognition is very real, much more natural than contact based devices, but there are some limitations towards this technique's development. Some of them are listed below [10]:-

- There are many possible gestures. But vision based interface does not yet cover the large variety of gestures we use in our day to day life. This is major challenge.
- More the gestures, more the degrees of freedom (DOF's), more the speed and resolution needed, more spatial scales. The handling of all these is a

difficult task at present. That is the reason we are able to recognise only a set of gestures today and not all of them.

- Also the according to the need, the system and the application, the technique needs to be constructed so as to meet the performance rate, real time tracking and recognition, accuracy and usability, robustness.
- Robustness is extremely important because the computer must give the user instant feedback and for that, the gesture needs to be detected, tracked and the image needs to be processed at the same frame rate as that of the input video.
- The system should be well equipped to handle in plane and out of plane rotations and identify the gesture successfully.
- The main advantage of vision based gesture recognition devices over contact based devices is that they are extremely user friendly and do not cause any health hazards. But at the same time, they are very complex, it takes time to get used to the device and also one needs to understand the procedure properly so as to obtain correct results.
- The problem of occlusion is also prevalent in vision based devices.
- Vision based hand gesture recognition devices are difficult to configure. They are not flexible.

Criterion	Contact-devices	Vision-devices
User cooperation	Yes	No
User intrusive	Yes	No
Precise	Yes/No	No/Yes
Flexible to configure	Yes	No
Flexible to use	No	Yes
Occlusion problem	No (Yes)	Yes
Health issues	Yes (No)	No

**Table. 1** Different Criterion to distinguish models of vision based interface hand gesture recognition systems, retrieved from [10]

### 3.4.6 APPLICATIONS

Vision based hand gesture recognition systems have been put up to a large use all over the world in various applications across a number of fields.

Hand gestures have replaced the need of keyboards and computers in many application and tasks particularly in the desktop application. For browsing the web and exploring, mouse gestures have come up. For editing documents and scrolling down pages, pen gestures have been developed. In response to some questions asked or options and dialogue boxes, nodding is another gesture.

Vision based gesture recognition has had the greatest impact on the development and advancement of Virtual Reality Interaction applications. It has enabled realistic imitation of virtual objects and two dimensional applications that behave and simulate three dimensional interactions [10]. It enables us to interact with data, computer applications and desktop in a more natural way. It helps us control and program robots, use signals to transfer data, automatic translation of the sign language. It provides an air of realism to man-made environments.

In more advanced terms, it has revolutionized the engineering as drawings of projects and construction techniques can be visually analysed. It enables no contact interaction in the medical field. This system makes the interaction much more comfortable and user friendly. It has enhanced mobile, tablet and PC navigation and usage through better performance and user satisfaction. In the gaming world, hand gesture recognition techniques have opened up a totally new unexplored area, where we can now interact with the players hand and body through gestures and no contact at all. We can sense ourselves as being a part of the game, existing into the game.



**Fig. 7** Applications of Gesture Recognition, retrieved from [20]

#### IV. ANALYSIS

In this paper, we have analysed the importance of human computer interaction in our lives today. Each moment we interact with the various machines and devices we depend upon. The interaction between the interface takes place through different mediums. We have studied one of the most popular and important interface type, visual based. Through the technique of hand gesture recognition, various arrays have been studied and real life applications have come up. Thus vision based interface using hand gesture recognition techniques has a very wide scope in the present as well as in the future. Apart from this technique, eye tracking, head movement tracking and facial expressions recognition provide much opportunity of research as they are yet to be explored. Thus vision based interface in human computer interaction aims at easing our lives and making communication with machines a real time experience.

#### V. CONCLUSION

Researchers have been working on vision based interface since long. It is surely a very attractive field. The aim is to design an interface which is efficient, effective and is user friendly. It should have no health hazards and should be very close to reality. Vision based interface is one such interface which uses one of the very basic sense that we have, vision. It makes use of this and enables users to control a system, a computer through this. Day by day scientists are working on vision based interface. They have made giant leaps in one of its area, gesture recognition. It has not only been studied but has also been put to use whose examples have been given in the 'Application' section. Now, the attention has been focussed to eliminate the disadvantages of vision based gesture recognition so that we develop systems that can detect, track and identify a wide variety of gestures. We need to develop systems that are easy to configure. Not only this, there is a lot of scope of other techniques under vision based interface such as eye gazing, identifying and interacting through facial expressions and head movement tracking that have not yet been explored much.

Human Computer Interaction is a field that has a lot of scope in the future because our lives today are interconnected with technology. The easier and user friendly it is to interact with machines, the simpler and exciting life gets.

#### REFERENCES

- [1] Siddharth S. Rautaray, Anupam Agrawal. "Vision based hand gesture recognition for human computer interaction: a survey", *Artificial Intelligence Review*, 2012
- [2] Rehg, J. M., & Kanade, T. (1994, November). Digiteyes: Vision-based hand tracking for human-computer interaction. In *Proceedings of*



- 1994 IEEE Workshop on Motion of Non-rigid and Articulated Objects (pp. 16-22). IEEE.
- [3] Murthy, G. R. S., & Jadon, R. S. (2009). A review of vision based hand gestures recognition. *International Journal of Information Technology and Knowledge Management*, 2(2), 405-410
- [4] Rautaray, S. S., & Agrawal, A. (2015). Vision based hand gesture recognition for human computer interaction: a survey. *Artificial intelligence review*, 43(1), 1- 54.
- [5] Wu, Ying, and Thomas S. Huang. "Nonstationary color tracking for vision-based human-computer interaction." *IEEE transactions on neural networks* 13, no. 4 (2002): 948-960.
- [6] Kim, K.N. and Ramakrishna, R.S., 1999, October. Vision-based eye-gaze tracking for human computer interface. In *IEEE SMC'99 Conference Proceedings. 1999 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No. 99CH37028) (Vol. 2, pp. 324-329)*. IEEE.
- [7] Murthy, G.R.S. and Jadon, R.S., 2009. A review of vision based hand gestures recognition. *International Journal of Information Technology and Knowledge Management*, 2(2), pp.405-410.
- [8] [https://en.wikipedia.org/wiki/Human%E2%80%93computer\\_interaction#Human.E2.80.93computer\\_interface](https://en.wikipedia.org/wiki/Human%E2%80%93computer_interaction#Human.E2.80.93computer_interface)
- [9] [https://en.wikipedia.org/wiki/User\\_interface#Types](https://en.wikipedia.org/wiki/User_interface#Types)
- [10] <http://www.nada.kth.se/cvap/gvmdi>
- [11] ELSEVIER<<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.76.6351&rep=rep1&type=pdf>>
- [12] Image: [www.researchgate.net](http://www.researchgate.net)
- [13] Song & takatsuka- Vision based hand gesture recognition
- [14] Belongie et al.- Vision based hand gesture recognition
- [15] Image: [www.andol.me](http://www.andol.me)
- [16] [www.hamlynkinect.wikispaces.com](http://www.hamlynkinect.wikispaces.com)
- [17] [www.site.uottawa.ca](http://www.site.uottawa.ca)
- [18] <https://www.google.co.in/search?q=vision+based+gesture+recognition>
- [19] Core applications of hand gesture recognition: [pawprintcooper.wordpress.com](http://pawprintcooper.wordpress.com)