Gesture Recognition and Control

Part 2 – Hand Gesture Recognition (HGR) System & Latest Upcoming Techniques

Prof Kamal K Vyas
Director (SIET) SGI Sikar
kamalkvyas@gmail.com

Amita Pareek
Tech-Advisor, eYuG
amushiva1@gmail.com

Dr S Tiwari
GE eYuG & Freelance Academician
Stiwari.eyug@rediffmail.com

Abstract: **This Exploratory paper's second part reveals the detail technological aspects of Hand Gesture Recognition (HGR) System. It further explored HGR basic building blocks, its application areas and challenges it faces. The paper also provides literature review on latest upcoming techniques like – Point Grab, 3D Mouse and Sixth-Sense etc. The paper concluded with focus on major Application fields.**

Key Words: **Hand Gesture Recognition (HGR), HGR basic building blocks, HGR Applications, HCI, MMI, Point Grab, 3D Mouse, Sixth-Sense.**

*

Introduction

The essential aim of building hand gesture recognition system is to create a natural interaction between human and computer where the recognized gestures can be used for controlling a robot or conveying meaningful information. How to form the resulted hand gestures to be understood and well interpreted by the computer considered as the problem of gesture interaction. Human computer interaction (HCI) also named Man-Machine Interaction (MMI) refers to the relation between the human and the computer or more precisely the machine, and since the machine is insignificant without suitable utilize by the human. There are two main characteristics should be deemed when designing a HCI system, functionality and usability. System functionality referred to the set of functions or services that the system equips to the users, while system usability referred to the level and scope that the system can operate and perform specific user purposes efficiently. The system that attains a suitable balance between these concepts considered as influential performance and powerful system. Gestures used for communicating between human and machines as well as between people using sign language. Gestures can be static (posture or certain pose) which require less computational complexity or dynamic (sequence of postures) which are more complex but suitable for real time environments. Different methods have been proposed for acquiring information necessary for recognition gestures system. Some methods used additional hardware devices such as data glove devices and color markers to easily extract comprehensive description of gesture features. Other methods based on the appearance of the hand using the skin color to segment the hand and extract necessary features, these methods considered easy, natural and less cost comparing with methods mentioned before. Some recent reviews explained gesture recognition system applications and its growing importance in our life especially for Human computer Interaction HCI, Robot control, games, and surveillance, using different tools and algorithms.

1. Hand Gesture Recognition (HGR)

It has been a dream of every nerd and geek to emulate the sweeping and swaying of hands to control your supercomputer of the future as shown in almost every sci-fi movie comprehensible. Tom Cruise in Minority Report [9] as well as Hritik Rosan in Krish did it and made us believe such sorcery would be possible in the near future. Techies all over the world have been developing downright crazy and innovative ways and means to bring this type of technology to our home.

1.1 Basic Building Blocks

Most of the researchers classified gesture recognition system into mainly three steps after acquiring the input image from camera(s), videos or even data glove instrumented device. These steps are:

a) Extraction Method,
b) Features Estimation and Extraction
c) Classification or Recognition

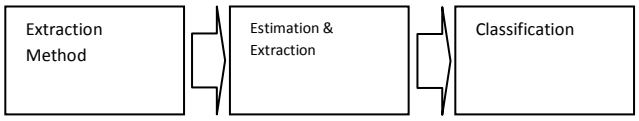

Figure 1[2]: Gesture recognition system steps

Segmentation process is the first process for recognizing hand gestures. It is the process of dividing the input image (in this case hand gesture image) into regions separated by boundaries. The segmentation process depends on the type of gesture, if it is dynamic gesture then the hand gesture need to be located and tracked, if it is static gesture (posture) the input image has to be segmented only. The hand should be located firstly, generally a bounding box is used to specify, depending on the skin color and secondly, the hand has to be tracked, for tracking the hand there are two main approaches; either the video is divided into frames and each frame have to be processed alone, in this case the hand frame is treated as a posture and segmented, or using some tracking information such as shape, skin color using some tools such as Kalman filter [1].

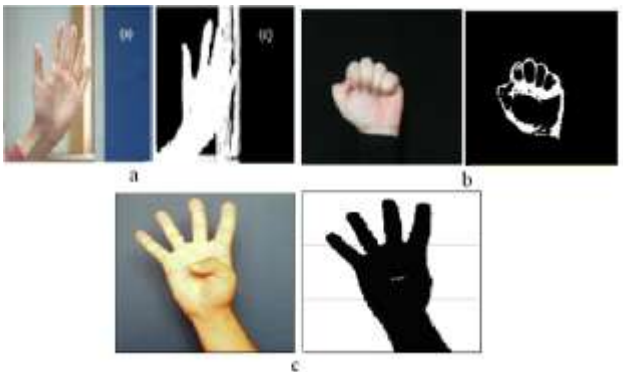

Fig 2 [2]

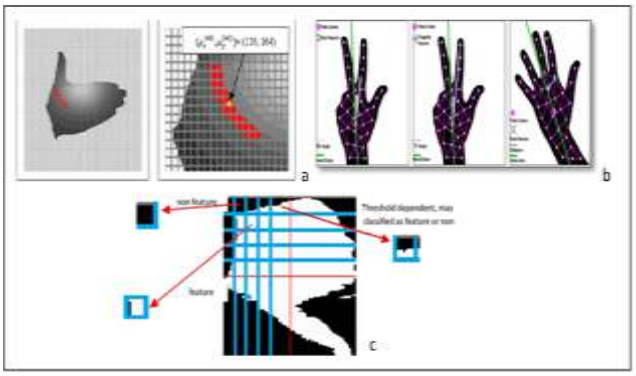

Fig 3 [3]

Good segmentation process leads to perfect features extraction process and the latter play an important role in a successful recognition process. Features vector of the segmented image can be extracted in different ways according to particular application. Various methods have been applied for representing the features and can be extracted. Some methods used the shape of the hand such as hand contour and silhouette while others utilized fingertips position, palm center, etc. created 13 parameters as a feature vector, the first parameters represents the ratio aspect of the bounding box of the hand and the rest 12 parameters are mean values of brightness pixels in the image used Self-Growing and Self-Organized Neural Gas (SGONG) neural algorithm to capture the shape of the hand, then three features are obtained; Palm region, Palm center, and Hand slope. Calculated the Center Of Gravity (COG) of the segmented hand and the distance from the COG to the farthest point in the fingers, and extracted one binary signal (1D) to estimate the number of fingers in the hand region. Divided the segmented image into different blocks size and each block represents the brightness measurements in the image. Many experiments were applied to decide the right block size that can achieve good recognition rate. Used Gaussian pdf to extract geometric central moment as local and global features. Figure 3 shows some applications of feature extraction methods. In figure 3 image :
a) The segmented image is partitioned into 11 terraces with 8 regions per terrace to extract local and global geometric central moment.
b) Three angles are extracted: RC angle, TC angle, and distance from the palm center. Segmented hand divided into blocks and the brightness factor for each block represents the feature vector (blocks with black area are discarded).

1.2 Architecture of Classification

After modeling and analysis of the input hand image, gesture classification method is used to recognize the gesture. Recognition process affected with the proper selection of features parameters and suitable classification algorithm. For example edge detection or contour operators cannot be used for gesture recognition since many hand postures are generated and could produce misclassification. Euclidean distance metric used to classify the gestures. Statistical tools used for gesture classification. HMM tool has shown its ability to recognize dynamic gestures besides, Finite State Machine (FSM), Learning Vector Quantization, and Principal Component Analysis (PCA). Neural network has been widely applied in the field of extracting the hand shape, and for hand gesture recognition. Other soft computing tools are effective in this field as well as Fuzzy C Means clustering (FCM), and Genetic Algorithms GAs. Figure 4 explain the architecture of classification system.

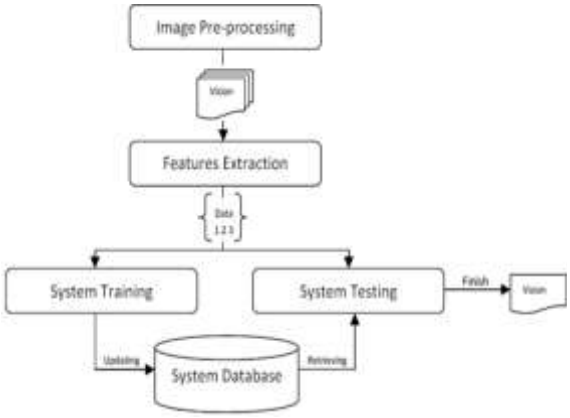

Figure 4 [4]: Architecture of classification system

1.3 Applications of Hand Gesture Recognition System

Hand gestures recognition system has been applied for different applications on different domains, as mentioned in including; sign language translation, virtual environments, smart surveillance, robot control, medical systems etc. overview of some hand gesture application areas are listed below :

- **Sign Language Recognition:** Since the sign language is used for interpreting and explanations of a certain subject during the conversation, it has received special attention. A lot of systems have been proposed to recognize gestures using different types of sign languages. For example recognized American Sign Language ASL using boundary histogram, MLP neural network and dynamic programming matching. recognized Japanese sign language JSL using Recurrent Neural Network, 42 alphabet and 10 words. Recognized Arabic Sign language ArSL using two different types of Neural Network, Partially and Fully Recurrent neural Network.

- **Robot Control:** Controlling the robot using gestures considered as one of the interesting applications in this field proposed a system that uses the numbering to count the five fingers for controlling a robot using hand pose signs. The orders are given to the robot to perform a particular task, where each sign has a specific meaning and represents different function for example, “one” means “move forward”, “five” means “stop”, and so on.

- **Graphic Editor Control:** Graphic editor control system requires the hand gesture to be tracked and located as a preprocessing operation used 12 dynamic gestures for drawing and editing graphic system. Shapes for drawing are; triangle, rectangular, circle, arc, horizontal and vertical line for drawing, and commands for editing graphic system are; copy, delete, move, swap, undo, and close.

- **Virtual Environments (VEs):** One of the popular applications in gesture recognition system is Virtual Environments VEs, especially for communication media systems provided 3D pointing gesture recognition for natural Human Computer Interaction HCI in a real-time from binocular views. The proposed system is accurate and independent of user characteristics and environmental changes.

- **Numbers Recognition:** Another recent application of hand gesture is recognizing numbers proposed an automatic system that could isolate and recognize a meaningful gesture from hand motion of Arabic numbers from 0 to 9 in a real time system using HMM.

- **Television Control:** Hand postures and gestures are used for controlling the Television device. In a set of hand gesture are used to control the TV activities, such as turning the TV on and off, increasing and decreasing the volume, muting the sound, and changing the channel using open and close hand.

- **3D Modeling:** To build 3D modeling, a determination of hand shapes are needed to create, built and view 3D shape of the hand [9]. Some systems built the 2D and 3D objects using hand silhouette. 3D hand modeling can be used for this purpose also which still a promising field of research.

1.4 Gesture Recognition through Non-geometric feature

Hasan applied multivariate Gaussian distribution to recognize hand gestures using non-geometric features. The input hand image is segmented using two different methods, skin color based segmentation by applying HSV color model and clustering based threshold techniques. Some operations are performed to capture the shape of the hand to extract hand feature; the modified Direction Analysis Algorithm are adopted to find a relationship between statistical parameters (variance and covariance) from the data, and used to compute object (hand) slope and trend by finding the direction of the hand gesture, as shown in Figure 5

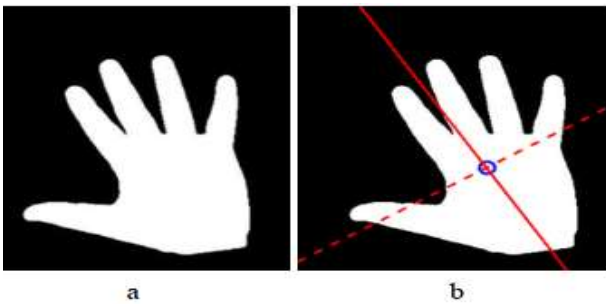

Figure 5[5]: Computing Hand Direction

Then Gaussian distinction is applied on the segmented image, and it takes the direction of the hand as shown in figure 6.

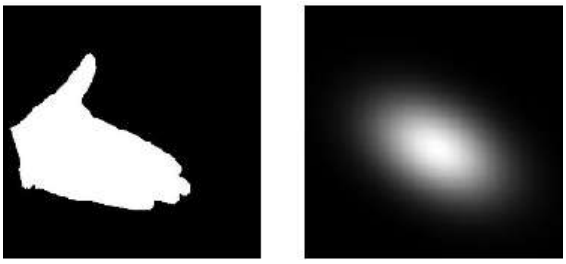

Figure 6 [5] : Gaussian distribution applied on the segmented image

Form the resultant Gaussian function the image has been

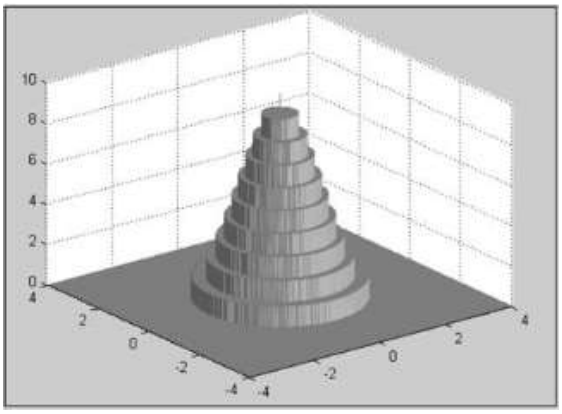

Figure 7 [5]: Terraces

divided into circular regions in other words that regions are formed in a terrace shape so that to eliminate the rotation affect. The shape is divided into 11 terraces with a 0.1 width for each terrace. 9 terraces are resultant from the 0.1 width division which are; (1-0.9, 0.9-0.8, 0.8-0.7, 0.7-0.6, 0.6, 0.5, 0.5-0.4, 0.4-0.3, 0.3-0.2, 0.2-0.1), and one terrace for the terrace that has value

smaller than 0.1 and the last one for the external area that extended out of the outer terrace. An explanation of this division is demonstrated in Figure 7.Each terrace is divided into 8 sectors which named as the feature areas, empirically discovered that number 8 is suitable for features divisions, To attain best capturing of the Gaussian to fit the segmented hand, re-estimation are performed on the shape to fit capturing the hand object, then the Gaussian shape are matched on the segmented hand to prepare the final hand shape for extracting the features, Figure 8 shown this process.

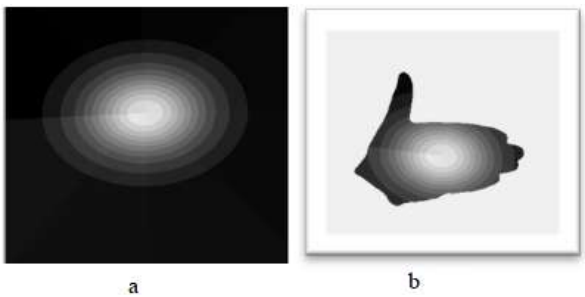

Figure 8 [5]: Features divisions. a) Terrace area in Gaussian. b) Terrace area in hand image

After capturing the hand shape, two types of features are extracted to form the feature vector; local feature, and global features. Local features using geometric central moments which provide two different moments μ00 , μ11 as shown by equation (1):

$$\mu_{pp} = \sum_{x}\sum_{y}(x-\mu_x)^p(y-\mu_y)^p f(x,y) \quad (1)$$

$$\mu_{pp}^{(k)} = \sum_{y}\sum_{x}\left(x^{(k)}-\mu_x^{(k)}\right)^p\left(y^{(k)}-\mu_y^{(k)}\right)^p f(x^{(k)},y^{(k)}) \quad (2)$$

$$\forall k \in \{1,2,3,\ldots,88\} \;\&\; \forall p \in \{0,1\}$$

Where μx and μy is the mean value for the input feature area, x and y are the coordinated, and for this, the input image is represented by 88*2 features, as explained in detail in equation (2). While the global features are two features the first and second moments that are the computed for the whole hand features area. These feature areas are computed by multiplying feature area intensity plus feature area's map location. In this case, any input image is represented with 178 features. The system carried out using 20 different gestures, 10 samples for each gesture, 5 samples for training and 5 for testing, with 100% recognition percentage and it decreased when the number of gestures are more than 14 gestures. In 6 gestures are recognized with 10 samples for each gesture. Euclidian distance used for the classification of the feature. Kulkarni [5] recognize static posture of American Sign Language using neural networks algorithm. The input image are converted into HSV color model, resized into 80x64 and some image preprocessing operations are applied to segment the hand from a uniform background, features are extracted using histogram technique and Hough algorithm. Feed forward Neural Networks with three layers are used for gesture classification. 8 samples are used for each 26 characters in sign language, for each gesture, 5 samples are used for training and 3samples for testing, the system achieved 92.78% recognition rate using MATLAB language. Applied scaled normalization for gesture recognition based on brightness factor matching. The input image is segmented using threshold technique where the background is black. Any segmented image is normalized (trimmed), and the center mass of the image are determined, so that the coordinates are shifted to match the centroid of the hand object at the origin of the X and Y axis. Since this method depends on the center mass of the object, the generated images have different sizes see figure 9, for this reason a scaled normalization operation are applied to overcome this problem which maintain image dimensions and the time as well, where each block of the four blocks are scaling with a factor that is different from other block's factors. Two methods are used for extraction the features; firstly by using the edge mages, and secondly by using normalized features where only the brightness values of pixels are calculated and other black pixels are neglected to reduce the length of the feature vector. The database consists of 6 different gestures, 10 samples per gesture are used, 5 samples for training and 5 samples for testing. The recognition rate for the normalized feature problem achieved better performance than the normal feature method, 95% recognition rate for the former method and 84% for the latter one. Wysoski et al. [6] presented rotation invariant postures using boundary histogram. Camera used for acquire the input image, filter for skin color detection has been used followed by clustering process to find the boundary for each group in the clustered image using ordinary contourtracking algorithm. The image was divided into grids and the boundaries have been normalized. The boundary was represented as chord's size chain which has been used as histograms, by dividing the image into number of regions N in a radial form, according to specific angle. For classification process Neural Networks MLP and Dynamic Programming DP matching were used. Many experiments have implemented on different features format in addition to use different chord's size histogram, chord's size FFT. 26 static postures from American Sign Language used in the experiments. Homogeneous background was applied in the work. Stergiopoulou suggested a new Self-Growing and Self-Organized Neural Gas (SGONG) network for hand gesture recognition. For hand region detection a color segmentation technique based on skin color filter in the YCbCr color space was used, an approximation of hand shape morphology has been detected using (SGONG) network. Three features were extracted using finger identification process which determines the number of the raised fingers and characteristics of hand shape, and Gaussian distribution model used for recognition.

1.5 Challenges for Image based Gesture Recognition System

There are many challenges associated with the accuracy and usefulness of gesture recognition software. For image-based gesture recognition there are limitations on the equipment used and image noise. Images or video may not be under consistent lighting, or in the same location. Items in the background or distinct features of the users may make recognition more difficult. The variety of implementations for image-based gesture recognition may also cause issue for viability of the technology to general usage. For example, an algorithm calibrated for one camera may not work for a different camera. The amount of background noise also causes tracking and recognition difficulties, especially when occlusions (partial and full) occur. Furthermore, the distance from the camera, and the camera's resolution and quality, also cause variations in recognition accuracy. In order to capture human gestures by visual sensors, robust computer vision methods are also required, for example for hand tracking and hand posture recognition for capturing movements of the head, facial expressions or gaze direction.

2 Upcoming New Technologies

Let's have a look at what the technology has brewed for the world of the future –

2.1 PointGrab [9] managed to refine that to perfection and presented us with a simple software that turns your camera into a motion-sensing device. Install the software while your webcam is plugged in and you can transform your computer,

laptop, smartphone, tablet, or a smart-television into a gesture-controlled device as shown in fig 9.

Fig 9 [9]

Working on simple principles, you can assign different commands to various gestures, right from zooming into a map, pause a video etc.

2.2 3D Mouse : This [7,9] allows you wear a small dongle-like thing on your finger and control the interface wirelessly. In other words, you are supposed to point your finger at the screen and control everything with the single effortless swoop of your finger. The advantages of this single implementation are endless. Gesturing with your finger means you won't have to use the might of your whole hand to do absurd and animated gestures to achieve a result, but by a simple flick of a finger. The multi-tasking capabilities are embedded in the dongle (as shown in fig 10) with three buttons snugly fitted in its sides that act as left-click, right-click and centre-scroller equivalent to that of a normal mouse. Throw away that old heckling mouse of yours and embrace the technology of the future.

Fig 10 [9]

2.3 Sixth Sense : It is a wearable gestural interface device developed by Pranav Mistry, a PhD student in the Fluid Interfaces Group at the MIT Media Lab. It is similar to Tele-pointer, a neck-worn projector/camera system developed by Media Lab student Steve Mann (which Mann originally referred to as "Synthetic Synesthesia of the Sixth Sense"). The Sixth-Sense prototype is comprised of a pocket projector, a mirror and a camera. The hardware components are coupled in a pendant like mobile wearable device. Both the projector and the camera are connected to the mobile computing device in the user's pocket. The projector projects visual information enabling surfaces, walls and physical objects around us to be used as interfaces; while the camera recognizes and tracks user's hand gestures and physical objects using computer-vision based techniques. The software program processes the video stream data captured by the camera and tracks the locations of the colored markers (visual tracking fiducials) at the tip of the user's fingers using simple computer-vision techniques. The movements and arrangements of these fiducials are interpreted into gestures that act as interaction instructions for the projected application interfaces. The maximum number of tracked fingers is only constrained by the number of unique fiducials, thus Sixth-Sense also supports multi-touch and multi-user interaction.

The Sixth-Sense prototype implements several applications that demonstrate the usefulness, viability and flexibility of the system. The map application lets the user navigate a map displayed on a nearby surface using hand gestures, similar to gestures supported by Multi-Touch based systems, letting the user zoom in, zoom out or pan using intuitive hand movements. The drawing application lets the user draw on any surface by tracking the fingertip movements of the user's index finger. Sixth-Sense also recognizes user's freehand gestures (postures). For example, the Sixth-Sense system implements a gestural camera that takes photos of the scene the user is looking at by detecting the 'framing' gesture. The user can stop by any surface or wall and flick through the photos he/she has taken. Sixth-Sense also lets the user draw icons or symbols in the air using the movement of the index finger and recognizes those symbols as interaction instructions. For example, drawing a magnifying glass symbol takes the user to the map application or drawing an '@' symbol lets the user check his mail. The Sixth-Sense system also augments physical objects the user is interacting with by projecting more information about these objects projected on them. For example, a newspaper can show live video news or dynamic information can be provided on a regular piece of paper. The gesture of drawing a circle on the user's wrist projects an analog watch.

2.3.1 Basic Technique

The Sixth-Sense prototype comprises a pocket projector, a mirror and a camera contained in a pendant like, wearable device as shown in fig 11. Both the projector and the camera are connected to a mobile computing device in the user's pocket. The projector projects visual information enabling surfaces, walls and physical objects around us to be used as interfaces; while the camera recognizes and tracks user's hand gestures and physical objects using computer-vision based techniques. The software program processes the video stream data captured by the camera and tracks the locations of the colored markers (visual tracking fiducials) at the tips of the user's fingers. The movements and arrangements of these fiducials are interpreted into gestures that act as interaction instructions for the projected application interfaces. Sixth-Sense supports multi-touch and multi-user interaction.

Fig 11 [9]

2.3.2 Application Areas for Sixth-Sense

- The Sixth-Sense prototype contains a number of demonstration applications.
- The map application lets the user navigate a map displayed on a nearby surface using hand gestures to zoom and pan.
- The drawing application lets the user draw on any surface by tracking the fingertip movements of the user's index finger.
- Sixth-Sense also implements Augmented reality; projecting information onto objects the user interacts.
- The system recognizes a user's freehand gestures as well as icons/symbols drawn in the air with the index finger, for example - A 'framing' gesture takes a picture of the scene. The user can stop by any surface or wall and flick through the photos he/she has taken. Drawing a magnifying glass symbol takes the user to the map application while an '@' symbol lets the user check his mail.
- The gesture of drawing a circle on the user's wrist projects an analog watch.

3. Major Application of Gesture Recognition

Gesture recognition has wide-ranging applications such as the following:

- Developing aids for the hearing impaired;
- Enabling very young children to interact with computers;
- Designing techniques for forensic identification;
- Recognizing sign language;
- Medically monitoring patients' emotional states or stress levels;
- Lie detection;
- Navigating and/or manipulating in virtual environments;
- Communicating in video conferencing;
- Distance learning/tele-teaching assistance;
- Monitoring automobile drivers' alertness/drowsiness levels, etc.
- Public Display Screens: Information display screens in Supermarkets, Post Offices, Banks that allows control without having to touch the device.
- Robots: Controlling robots without any physical contact between human and computer.
- Graphic Editor Control: Controlling a graphic editor by recognizing hand gestures using HMM

References

1. Matthias Rehm, Nikolaus Bee, Elisabeth André, Wave Like an Egyptian - Accelerometer Based Gesture Recognition for Culture Specific Interactions, British Computer Society, 2007
2. Pavlovic, V., Sharma, R. & Huang, T. (1997), "Visual interpretation of hand gestures for human-computer interaction: A review", IEEE Trans. Pattern Analysis and Machine Intelligence., July, 1997. Vol. 19(7), pp. 677 - 695.
3. R. Cipolla and A. Pentland, Computer Vision for Human-Machine Interaction, Cambridge University Press, 1998, ISBN 978-0521622530
4. Ying Wu and Thomas S. Huang, "Vision-Based Gesture Recognition: A Review", In: Gesture-Based Communication in Human-Computer Interaction, Volume 1739 of Springer Lecture Notes in Computer Science, pages 103-115, 1999, ISBN 978-3-540-66935-7, doi 10.1007/3-540-46616-9
5. Alejandro Jaimesa and Nicu Sebe, Multimodal human–computer interaction: A survey, Computer Vision and Image Understanding Volume 108, Issues 1-2, October–November 2007, Pages 116-134 Special Issue on Vision for Human-Computer Interaction, doi:10.1016/j.cviu.2006.10.019
6. Thad Starner, Alex Pentland, Visual Recognition of American Sign Language Using Hidden Markov Models, Massachusetts Institute of Technology
7. Kai Nickel, Rainer Stiefelhagen, Visual recognition of pointing gestures for human-robot interaction, Image and Vision Computing, vol 25, Issue 12, December 2007, pp**9^** Lars Bretzner and Tony Lindeberg "Use Your Hand as a 3-D Mouse ...", Proc. 5th European Conference on Computer Vision H. Burkhardt and B.
8. Moniruzzaman Bhuiyan, Rich Picking Journal of Software Engineering and Applications, 2011, 4, 513-521 doi:10.4236/jsea.2011.49059 Published Online September 2011
9. www.mensxp.com/technology/portable-media/9202-gesture-recognition-the-controls-of-the-future.html.