CHAPTER -2

LITERATURE SURVEY

A Literature Review is a systematic and compressive analysis of books, scholarly articles and other sources relevant to a specific topic providing a base of knowledge on a topic. Literature reviews are designed to identify and critique the existing literature on a topic to justify your research by exposing gaps in current research. This investigation should provide a description, summary, and critical evaluation of works related to the research problem and should also add to the overall knowledge of the topic as well as demonstrating how your research will fit within a larger field of study. A literature review should offer critical analysis of the current research on a topic and that analysis should direct your research objective. A Literature Review can be a stand-alone element or part of a larger end product, know your assignment. Key to a good Literature Review is to document your process.

2.1 Hand Gesture Recognition using Web Cam - Ankita Saxena, Deepak Kumar Jain, Ananya Singhal, Central Electronics Engineering Research Institute, Pilani 333031, Rajasthan. IEEE conference paper on hand gesture recognition, 2014.

2.1.1 Architecture:

Gestures can originate from any bodily motion or state but commonly originate from the face or hand. Gesture recognition can be seen as way for computers to begin to understand human body language, thus building a richer bridge between machines and humans. It enables humans to communicate with the machine (HMI) and interact naturally without any mechanical devices. There has been always considered a challenge in the development of a natural interaction interface, where people interact with technology as they are used to interact with the real world. A hand free interface, based only on human gestures, where no devices are attached to the user, will naturally immerse the user from the real world to the virtual environment. The block diagram is shown in Figure 2.1 below.

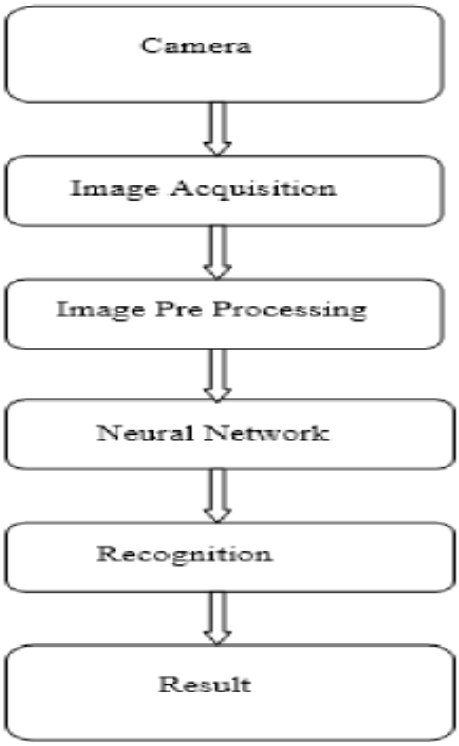

Figure 2.1: Block Diagram of Hand Gesture Recognition Model

2.1.2 Methodology:

- Image acquisition is the first step in any vision system, only after this process you can go forward with the image processing. In this application it is done by using IPWebCam android application. The application uses the camera present in the phone for continuous image capturing and a simultaneous display on the screen. The image captured by the application is streamed over its Wi-Fi connection (or WLAN without internet as used here) for remote viewing. The program access the image by logging to the devices IP, which is then showed in the GUI.

- In this paper the edge detection technique used is sobel edge detector. The image captured is then passed through sobel filter. Thinning is a morphological operation that is used to remove selected foreground pixels from binary images, somewhat like erosion or opening. In this mode it is commonly used to tidy up the output of edge detectors by reducing all lines to single pixel thickness.

- Edge Detection is the early processing stage in image processing and computer vision, aimed at detecting and characterizing discontinuities in the image domain. It aims at identifying points in a digital image at which the image brightness

changes sharply or, more formally, has discontinuities. The points at which image brightness changes sharply are typically organized into a set of curved line segments termed edges. Some of the different types of edge detection techniques are:

1.Sobel Edge Detector

2.Canny Edge Detector

3.Prewitt Edge Detector

The Sobel operator is used in image processing to detect edges of an image. The operator calculates the gradient of the image intensity at each point, giving the direction of the largest possible increase from light to dark and the rate of change in that direction. The result therefore shows how ”abruptly” or ”smoothly” the image changes at that point, and therefore how likely it is that, that part of the image represents an edge, as well as how that edge is likely to be oriented.

- An artificial neuron is a computational model inspired in the natural neurons. These networks consist of inputs (like synapses), which are multiplied by weights (strength of the respective signals), and then computed by a mathematical function which determines the activation of the neuron. Another function (which may be the identity) computes the output of the artificial neuron (sometimes in dependence of a certain threshold). Artificial Neural Networks (ANN) combine artificial neurons in order to process information. The higher a weight of an artificial neuron is, the stronger the input which is multiplied by it will be.

- The back propagation algorithm is used in layered feed-forward ANNs. This means that the artificial neurons are organized in layers, and send their signals forward, and then the errors are propagated backwards. The network receives inputs by neurons in the input layer, and the output of the network is given by the neurons on an output layer. There may be one or more intermediate hidden layers. The backpropagation algorithm uses supervised learning, which means that we provide the algorithm with examples of the inputs and outputs we want the network to compute, and then the error is calculated. The idea of the backpropagation algorithm is to reduce this error, until the ANN learns the training data.

- Recognition is the final step of the application. To fill the input neurons of the trained network, the previous calculated tokens discussed in section D are used. The number of output neurons is normally specified by the amount of different type of gestures, in this case it is fixed to 6.

Another main part of this work is the integration of a feed-forward backpropagation neural network. As described earlier the inputs for this neuronal network are the individual tokens of a hand image, and as a token normally consists of a cosinus and sinus angle, the amount of input layers for this network are the amount of tokens multiplied by two. The implemented network just has one input, hidden and output layer to simplify and speed-up the calculations on that java implementation. For training purpose the database of images located on the disk is used. It contains 6 different types of predefined gestures.

2.1.3 Limitations:

• Artificial neural networks require processors with parallel processing power, in accordance with their structure.

• This is the most important problem of ANN. When ANN produces a probing solution, it does not give a clue as to why and how. This reduces trust in the network.

• Gradient descent with backpropagation is not guaranteed to find the global minimum of the error function.

2.2 Hand Gesture Recognition Using Deep Learning - Soeb Hussain and Rupal Saxena Department of Chemistry Indian Institute of Technology, Guwahati Guwahati, India. IEEE conference paper on hand gesture recognition with deep learning, 2017.

2.2.1 Architecture:

The aim is to recognize six static and eight dynamic gestures while maintaining accuracy and speed of the system. For hand shape recognition, a CNN based classifier is trained through the process of transfer learning over a pretrained convolutional neural net which is initially trained on a large dataset. Our application belongs to the domain of hand gesture recognition which is generally divided into two categories i.e. contact-based and vision-based approaches. The second type is simpler and intuitive as it employs video image processing and pattern recognition. The aim is to recognize six static and eight dynamic gestures while maintaining accuracy and speed of the system. The recognized gestures are to command the computer.

In our work, VGG16 a CNN architecture is used as the pretrained model. It consists of 13 convolution layers followed by 3 fully connected layers. A convolutional neural network (CNN) is a type of feed-forward artificial neural network in which the connectivity pattern between its neurons is inspired by the organization of the animal visual cortex. We need to recognize eleven hand shapes, hence CNN is trained as a classifier using transfer learning method. To reach the desired output, network model needs to be altered. Therefore, two layers of the model were replaced with a set of layers that can classify 11 classes. All other layers remained unaltered. To avoid over fitting, the Regularization along with a more diverse dataset was introduced. Regularization involves modifying the performance function which is normally chosen to be the sum of the square of the network errors on the training set. The process flow diagram of our proposed method is shown in Figure 2.2.

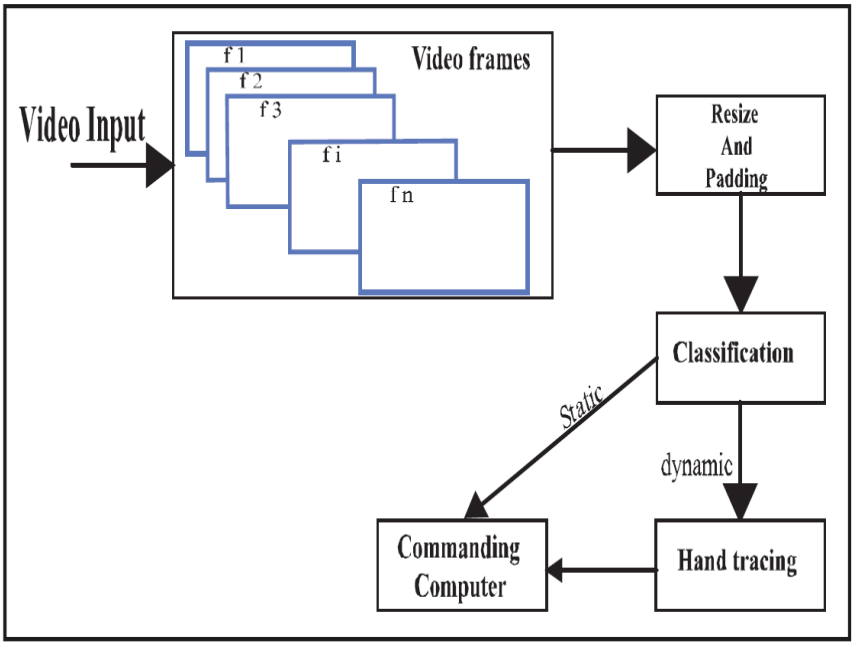

Figure 2.2: Process Flow Diagram Of Static Hand Gesture Recognition

2.2.2 Methodology:

- Hand segmentation is the first step in applications such as gesture recognition, fingers tracking,etc. Color is a strong tool for image segmentation. Color-based segmentation is better than edge-based and luminance histograming techniques because color is computational inexpensive, and it can give more information than a luminance-only image or an edge-segmented image.

  To avoid over fitting, the Regularization along with a more diverse dataset was introduced. Regularization involves modifying the performance function which is normally chosen to be the sum of the square of the network errors on the training set. Recognition of a static gesture requires only the hand shape. Once hand shape is classified as static gesture by the trained classifier, command is given to the computer. Unlike static gesture, dynamic gesture requires both the hand shape as well as the motion of hand.

- For tracing dynamic hand gestures, hand area is segmented out using HSV (Hue, Saturation, Value) skin color algorithm in a frame, followed by cropping blob area. Centroid of the blob is detected and traced. The main idea in this stage consists in retrieving the coordinates of the traced hand's center in each frame. These coordinates will be used in order to know which computer command corresponds to which motion. Coordinates will be used differently for each

gesture, depending on detected hand shape. Tracing involves extracting position of hand which is done by skin color detection, skin cropping, blob detection and centroid extraction. Hence tracing on the whole is a comparatively time consuming process.

The propose a vision based hand gesture recognition method using transfer learning. The method was made robust by avoiding skin color segmentation, blob detection, skin area cropping and centroid extraction for unidirectional dynamic gestures. Prototype was tested successfully on seven different volunteers at different backgrounds and light conditions with an accuracy of 93.09%.

2.2.3 Limitations:

• High computational cost.

• If you don't have a good GPU they are quite slow to train (for complex tasks)

• They use to need a lot of training data.

• SVM is defined by a convex optimization problem (no local minima) for which there are efficient methods.

2.3 A Basic Hand Gesture Control System for PC Applications- Charles J. Cohen, Glenn Beach, Gene Foulk, Cybemet Systems Corporation, Tunisia. IEEE journal paper on real time hand gesture, 2015.

2.3.1 Architecture:

In this paper they have chosen to integrate a free graphics library specializes in real time image processing named Open CV (Open Computer Vision) library to perform image processing. And then we have chosen to integrate the SVM classifier (Support Vector Machine) which can model real-world problems such as image classification, hand-writing recognition, text, bioinformatic and biosequence analysis. It has proposed a system based on SVM for recognizing various hand gestures. The system consists of four steps: hand segmentation, smoothing, feature extraction and classification. The idea here is to allow the smartphone to perform all necessary steps to recognize gestures without the need to connect to a computer in which a database is located to perform training process. With this system, all steps can be done by the smartphone. In this paper, for image acquisition, frontal camera of the smartphone is used. After that frames are gotten from the video, the color sampling is done which is followed by making binary representation of the hand, and then contours representing the hand were described with convex polygons to get information about fingertips and finally the input gesture was recognized using proper classifier. The system, unlike systems realized before, is designed to perform all necessary operations in order to recognize hand gestures, by the smartphone and without the need to use another device. Some applications that use gestures as an Human-Machine Interface were first imposed in game consoles (Microsoft Kinect and PlayStation Eye for Sony) and could now be used in other areas such as smart television and smartphone. If the classified hand is a static gesture then it immediately passes to commanding phase. Otherwise, it passes to hand tracing phase. The block diagram of our proposed method is shown in Figure 2.3 below.

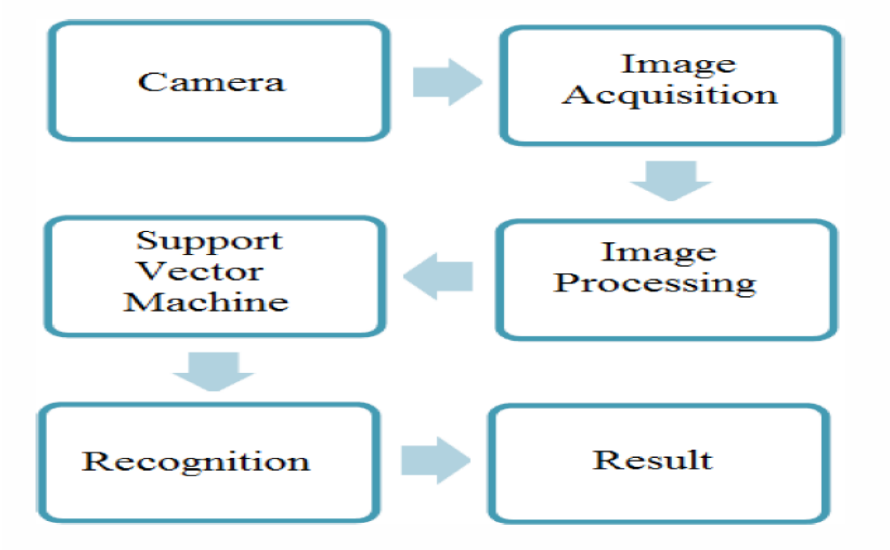

Figure 2.3: Architecture Of Hand Gesture Control System

2.3.2 Methodology:

The hand gestures are read by an input device like smartphone to detect frames. Hand gestures are then interpreted using algorithm either based on artificial intelligence techniques or statistical analysis. Gesture recognition research aims to convey information through a system which can recognize specific human hand gestures.

- Hand segmentation is the first step in applications such as gesture recognition, fingers tracking,etc. Color is a strong tool for image segmentation. Color-based segmentation is better than edge-based and luminance histograming techniques because color is computational inexpensive, and it can give more information than a luminance-only image or an edge-segmented image. One of the biggest problems in signal and image processing and computer vision, is to filter the noise while preserving the edges. For many years, many methods of adaptive filtering were developed following multiple approaches.

  The median filter is a nonlinear digital filter, often used for noise reduction. Noise reduction is a step in image pre processing to improve the results of future treatments. The median filter technique is widely used in digital image processing because it allows to reduce noise while preserving the contours in the image. The main idea of the median filter is to replace each entry with the median value of its neighborhood.

- After detecting hand we detect the features. When the binary image is generated, the segmented part, which represents the hand, is processed as follows:
    - get convex points in contour
    - get point furthest away from each convex vertex (convexity defect)
    - Filter out convexity defect not relevant

- Support vector machines (SVMs) are a set of supervised learning techniques aimed at solving classification and regression problems. SVMs are a generalization of linear classifiers. SVMs were developed in the 1990s from theoretical considerations of Vladimir Vapnik on the development of a statistical learning theory: the theory of Vapnik-Chervonenkis. SVMs were quickly adopted for their ability to work with large data, small number of hyper parameters, their theoretical guarantees, and their good results in practice. SVMs can be used to solve problems of classification, such as knowing to which class a sample belongs, or regression, such as predicting the numerical value of a variable. The SVM algorithm in its original form is like looking for a Linear boundary between two classes, but this model can greatly be enhanced by projecting in another space to increase the separability of the data. We can then apply the same algorithm in the new space, which results a nonlinear boundary in the initial space. So SVM can perform both linear and nonlinear classification using the kernel trick.The basic idea of the kernel trick is to preprocess initially the data by a non-linear mapping <l> and then to apply the same linear algorithm but in the image space.

- SVM is a binary supervised classification algorithm. It can deal with problems involving large numbers of descriptors, it provides a unique solution (no local minimum problems as with neural networks) and it has provided good results on real problems. 10 gestures have been collected to constitute the initial data base. In this system the gestures from the image of the whole palm.

- The concept of tramming is important. Training by induction allows reaching conclusions by examining particular examples. It is divided into supervised and unsupervised training. The cases regarding SVM is supervised training. Particular examples are represented by a set of input / output pairs. The goal is to learn a function which corresponds to those seen examples and which predicts the outputs

for the entries that have not yet been seen. Entries can be object's descriptions, and outputs the class of given objects as input. In our case, Multi class SVM is used to build the training model and make predictions. For training purpose, the database of images is used. When we want to increase the number of a hand gesture, which is already exist in the training set, we can do it in our application by choosing at first image that represents the gesture. We can then add a gesture to the training set; we must first make a static hand gesture in front of the camera and after that, we can add this gesture to the training set. Gestures in the training set are first processed and then the extracted features are passed to the SVM model for training purpose.

This system can be used to control the smartphone without the need to touch it. With this system, we only have to make a gesture in front of the camera so that it performs the appropriate action. It could also serve to deaf people to communicate with others.

2.3.3 Limitations:

• It has a regularisation parameter, which makes the user think about avoiding over-fitting.

• SVM is defined by a convex optimisation problem (no local minima) for which there are efficient methods

• It is an approximation to a bound on the test error rate, and there is a substantial body of theory behind it which suggests it should be a good idea.

2.4 Recognition of Static Hand Gesture - Khadidja Sadeddine, Rachida Djeradi, Fatma Zohra Chelali and Amar Djeradi LCPTS Laboratory, Electronics and Computer Science Faculty University Houari-Boumediene of Sciences and Technology Algiers, Algeria, IEEE conference paper on static hand gesture, 2018.

2.4.1 Architecture:

In order to design static hand gesture recognition system, we propose in the present paper, an alphabet of hand postures for both American (ASL) and Arabic (ArSL) sign languages, and NUS database composed of static hand gestures. For all databases used, the postures are made by free hand and gestures are mono-manuals for many people. Some processing on images is performed before the classification task. The architecture is shown in Figure 2.4 below.

This work aims to develop recognition system for static gestures. Three databases are used. The First one is the Jochen Triesch dataset for American Sign Language (ASL) alphabet with ten hand posture images of ASL (A, B, C, D, G, H, I, L, V, Y) with simple background made by 24 persons. The second dataset is the Halawani dataset for Arabic Sign Language (ArSL) that contains 30 sign alphabets corresponding to 60 peoples.

2.4.2 Methodology:

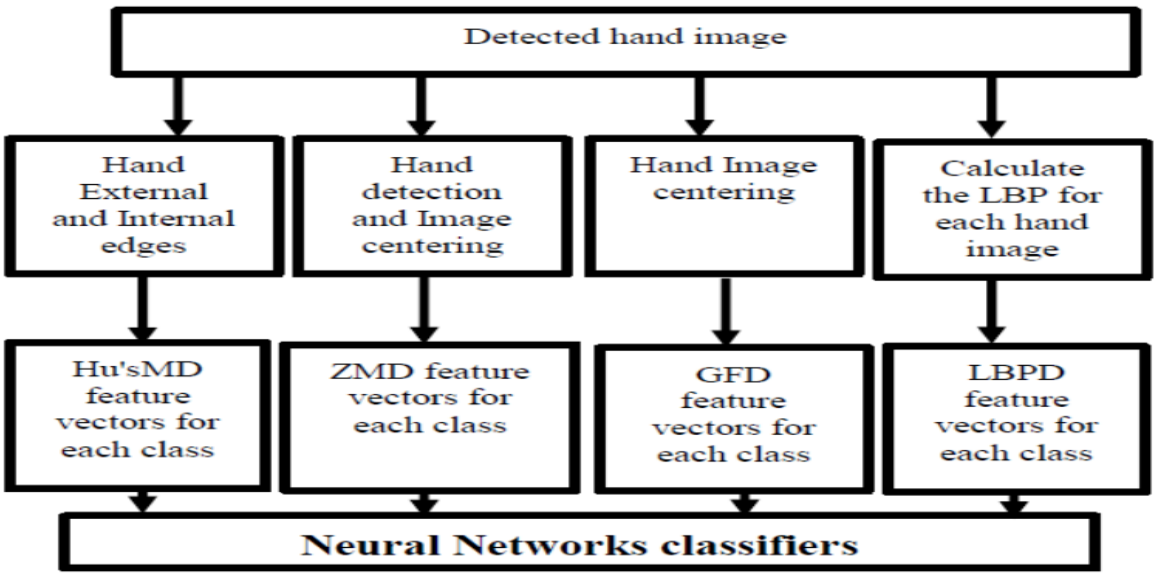

Figure 2.4: Architectural Diagram Of Static Hand Gesture System

- The system comprises three phases: hand detection, feature extraction phase and the classification phase using Neural Networks. The hand segmentation is made by Otsu threshold with some binary operations to detect the hand posture region.

- The second phase, implements many descriptors such as Hu's moments, Zernike moments, Local binary pattern and Generic Fourier descriptors. The third phase is the classification/recognition that employs probabilistic neural network PNN and Multilayer neural network MLP. The external edges of hand image were used as feature but unfortunately, the results obtained were unappreciated. This is why we apply the Canny contour detection filter on hand images to get external and internal edges. Neural networks NN are used to evaluate the efficiency of descriptors features for recognition of sign alphabets.

- The feed-forward MLP neural network is used to evaluate the efficacy of the descriptors vectors. A three-layer NN with weights adjusted using the scaled conjugate gradient (SCG) algorithm that is the appropriate algorithm for our case. The PNN is a multilayer neural network defined as an implementation of a statistical algorithm called Kernel discriminate analysis in which the operations are organized into multilayered feed-forward network. PNN possess four layers. Input layer, where number of neurons corresponding to number the input training samples. Pattern layer that contains one neuron for each training case available. Summation layer that computes the PDF estimation of x for the ith class with some smoothing parameters representing standard deviation (also called window or kernel width) around the mean of p random variables. Finally, the output layer with 10 neurons for ASL and 30 for ArSL. The classification decision is made by computing the posterior probabilities of all classes and selecting that class which yields the maximum value.

2.4.3 Limitations:

• This is designed for static hand gesture, not for dynamic hand gesture.

• This system will not work when there is complex background.

2.5 A static hand gesture recognition system for real time mobile device monitoring

- Hanene Elleuch, Ali Wali, Anis Samett and Adel M. Alimi, REGIM: Research Groups in Intelligent Machines, University of Sfax, IEEE 2015

2.5.1 Architecture:

The hand gesture recognition is considered as the most important alternative that can be deployed because it is natural, intuitive and easy to use. In this paper, we proposed our system of static hand gesture recognition to control mobile devices based only on a real time video streaming capture from the front-facing camera of the device. Our proposed system is based on the skin colour algorithm and face subtraction to detect hand area. The features derived from contour extraction, convex hull detection, convexity defects extraction and the palm center detection are used on SVM classifier for the recognition step. This paper proposes a system of static hand gesture recognition to control and command and tablet based on Android operation system. The flowchart is shown in Figure 2.5 below.

2.5.2 Methodology:

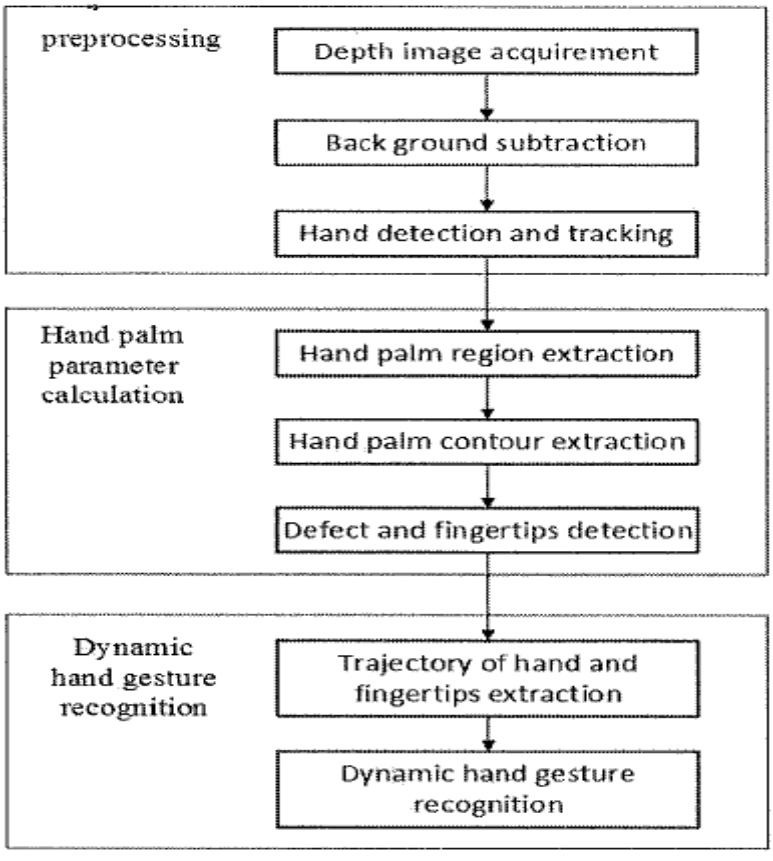

Figure 2.5: Process Flow Diagram Of Real Time Gesture Recognition System

- Video capturing is the first step in any vision system, only after this process you can go forward with the image processing. In this application it is done by using IPWebCam of laptop. The application uses the camera present in the laptop for continuous image capturing and a simultaneous display on the screen.

- The second step is the processing of video frames. Here each frames from the video is processed separately. There are many ways as shown in Figure 2.6 below:
    - Skin Segmentation: The first step in implementing our particular gesture-recognition system is being able to effectively segment skin pixels from non-skin pixels. By using only a simple RGB-based webcam we are limited in methods for locating and distinguishing static hand gestures. For this reason, we have chosen to focus on segmentation using various color spaces. Skin segmentation methods are generally computationally inexpensive, and moreover, they can function robustly across many different models of simple webcams - an important feature given the notable difference in quality, color, etc. that can exist between webcams.
    - Canny Edge detection:Edge detection is one of the fundamental operations when we perform image processing. It helps us reduce the amount of data (pixels) to process and maintains the structural aspect of the image. We're going to look into two commonly used edge detection schemes - the gradient (Sobel - first order derivatives) based edge detector and the Laplacian (2nd order derivative, so it is extremely sensitive to noise) based edge detector. Both of them work with convolutions and achieve the same end goal - Edge Detection. Canny Edge detection was invented by John Canny in 1983 at MIT. It treats edge detection as a signal processing problem. The key idea is that if you observe the change in intensity on each pixel in an image, it’s very high on the edges.
    - Noise removal: Since this method depends on sudden changes in intensity and if the image has a lot of random noise, then it would detect that as an edge. So, it’s a very good idea to smoothen your image using a Gaussian filter of 5×5.
    - Gradient Calculation: In the next step, we calculate the gradient of intensity(rate of change in intensity) on each pixel in the image. We also calculate the direction of the gradient. Gradient direction is perpendicular to the edges. It’s mapped to one of the four directions(horizontal, vertical,

and two diagonal directions).

- Non-Maximal Suppression: Now, we want to remove the pixels(set their values to 0) which are not edges. You would say that we can simply pick the pixels with the highest gradient values and those are our edges. However, in real-world images, gradient doesn't simply peak at one pixel, rather it's very high on the pixels near the edge as well. So, we pick the local maxima in a neighborhood of 3×3 in the direction of gradients.

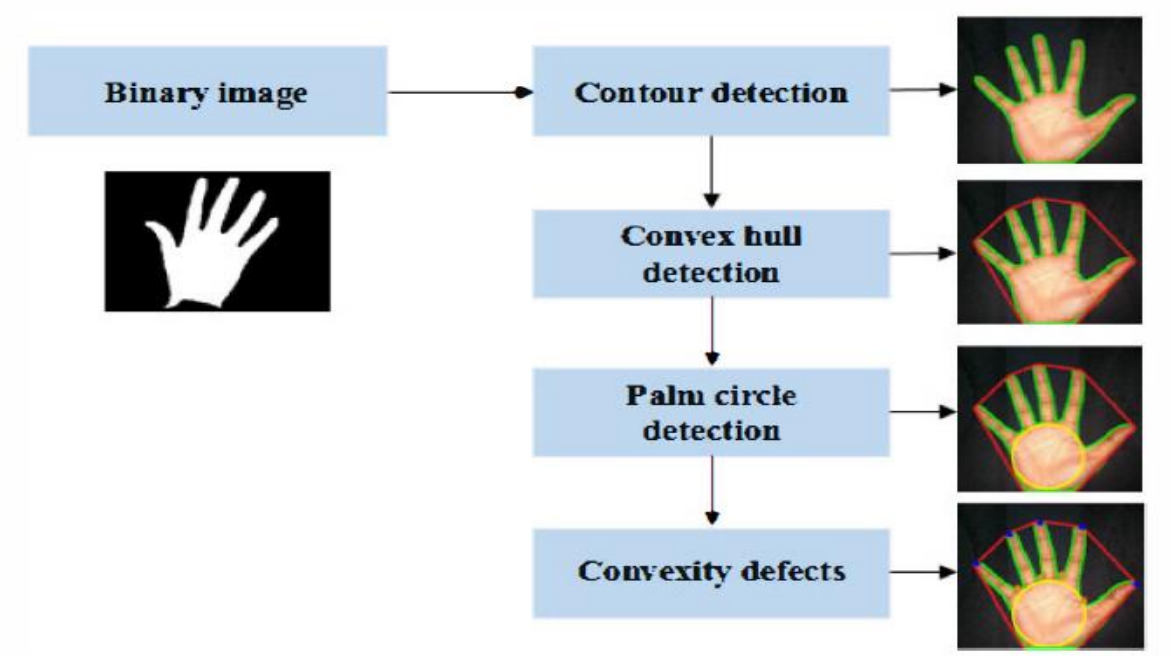

Figure 2.6: Hand Gesture Feature Extraction

- Background substraction: Background subtraction is a major preprocessing steps in many vision based applications. For example, consider the cases like visitor counter where a static camera takes the number of visitors entering or leaving the room, or a traffic camera extracting information about the vehicles etc. In all these cases, first you need to extract the person or vehicles alone. Technically, you need to extract the moving foreground from static background. If you have an image of background alone, like image of the room without visitors, image of the road without vehicles etc, it is an easy job. Just subtract the new image from the background. You get the foreground objects alone. But in most of the cases, you may not have such an image, so we need to extract the background from whatever images we have. It become more complicated when there is shadow of the vehicles. Since shadow is also moving, simple subtraction will mark that also as foreground.

- Thinning: Thinning is the transformation of a digital image into a simplified, but topologically equivalent image. It is a type of topological skeleton, but computed using mathematical morphology operators. Thinning is a morphological operation that is used to remove selected foreground pixels from binary images, somewhat like erosion or opening. It can be used for several applications, but is particularly useful for skeletonization. In this mode it is commonly used to tidy up the output of edge detectors by reducing all lines to single pixel thickness. Thinning is normally only applied to binary images, and produces another binary image as output. The thinning operation is related to the hit-and-miss transform, and so it is helpful to have an understanding of that operator before reading on. Thinning operation is calculated by translating the origin of the structuring element to each possible pixel position in the image, and at each such position comparing it with the underlying image pixels.

- Image thresholding: Thresholding is the simplest method of image segmentation. From a grayscale image, thresholding can be used to create binary images. The simplest thresholding methods replace each pixel in an image with a black pixel if the image intensity is less than some fixed constant or a white pixel if the image intensity is greater than that constant. If pixel value is greater than a threshold value, it is assigned one value (may be white), else it is assigned another value (may be black). The function used is cv2.threshold. First argument is the source image, which should be a grayscale image. Second argument is the threshold value which is used to classify the pixel values. Third argument is the maxVal which represents the value to be given if pixel value is more than (sometimes less than) the threshold value.

We conducted a study to validate the robustness of our system. To attain this purpose, we deployed our system entirely on an Android-based tablet with NVIDIA Tegra 3 Quad-Core and a 32 Go of RAM running an android version of 4.2.1. The front-facing camera has a 2 Mega pixels resolution. lO participants aged between 24 and 29 are asked to test our system. As our system is dedicated to be used in different situations, we ensure the variance of the test environment: different places and lighting conditions. In this system, we introduce the static hand gesture recognition to command mobile devices. The first step of our system is the detection of the hand. This detection is based on the Skin color

algorithm. As this method confused the user's face and the hand we add an module of face subtraction that detect the face by using the viola and Jones algorithm and then subtracted it.In the next step, we extracted hand features from the contour of the hand, the convex hull, the convexity defects and the palm center and we used for the SVM classifier to recognize the static hand gesture. We deployed our system on an Android-based tablet and 10 participants are asked to tested it. The experiment result showed that the successful rate is about 96.8%. An implementation based on parallel designis necessary to increase the processing time that attains 23 fps.