International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2016): 79.57 | Impact Factor (2017): 7.296

# Dynamic Hand Gesture Recognition

**Riddhi Laxman Zunjarrao**

Thakur College of Engineering and Technology, Mumbai

**Abstract:** ***This Survey takes a brief overview of Dynamic Hand Gesture Recognition. We will take a look at various techniques and algorithms. We will compare all the algorithms and also see the pros and cons of each of them. Literature Survey is done on the extensively on Algorithm of Dynamic Gesture Recognition. Gesture recognition is a topic in computer science and language technology with the goal of interpreting human gestures. Gestures can originate from any bodily motion or state but commonly originate from the face or hand. It shows that a lot of work is being carried out in this field of gesture recognition and is used for various applications. Gesture recognition helps in building a richer bridge between machine and human.***

**Keywords:** Gesture recognition, Histogram of oriented gradients, Haar-like feature, Mean shift, Deep Belief Network, Support Vector Machine

## 1. Introduction

Hand gesture recognition aims to design and develop such systems than can identify explicit human gestures as input and process these gesture.

Human hand gestures provide the natural and effective mode of non-verbal communication with the computer interface. Hand gestures are the meaningful body motions that are movement of the hands, face, or other parts of the body, to convey information or interact with the environment. Hand gesture can be static gesture and dynamic gesture. The hand is directly use as the input to the machine, for the communication purpose of gesture identification there is no need of an intermediate medium.

Dynamic hand gesture recognition has various techniques and algorithms that are used to identify different type of gesture. This report briefly describes techniques and the various algorithms which are used for implementation. The report also has the pros, cons, applications along with various parameters for the comparative study for choosing the best known algorithm within the technique. We will take a brief look in the basics of the techniques.

## 2. Literature Survey

### A. Gesture Recognition

With the development of HCI (Human-Computer Interaction) in recent years, hand gesture has been considered as the most natural way to communicate between humans and computers. Gesture recognition enables humans to communicate with the machine and interact naturally without any mechanical devices. Gestures is a powerful communication mode for Human Computer Interaction. Gesture is a motion of body to express thoughts or to deliver a message. Gesture is nonverbal communication that includes communication through body postures, hand gestures and facial expressions makes up most of all communication among human. Gesture recognition is the process of identifying the gestures by the computer which is made by the user.

Hand gestures can be divided into two types from the aspect of time interval:

a) Static gesture
b) Dynamic gesture

***Static Gesture:***
A static gesture is observed at the spurt of time. It could only recognize the predefined gesture. In static gesture recognition, the hand shape, size of palm, length and width of fingers need to be kept in mind. The stop sign is an example of static gesture.

***Dynamic gesture:***
A dynamic gesture intended to change over a period of time. Gesture is recognized by its movement. Dynamic hand gestures need spatiotemporal information to track hand. A waving hand means goodbye is an example of dynamic gesture.

### B. Application of gesture

*Gaming interface:* Gestures used to play computer games. Tracks a player's hand or body position to control movement and orientation of interactive game objects such as cars. It can be used to control the movement and it helps to track hand movements for interactive games.

*Sign Language*: Sign language is an important case of communicative gestures. It can also be a good way to help the disabled to interact with computers. Indian Sign Language and American Sign Language is by deaf person.

*Virtual Reality:* Gestures for virtual and augmented reality applications have experienced one of the greatest levels of uptake in computing. Virtual reality interactions use gestures to enable realistic manipulations of virtual objects using one's hands, for 3D display interactions or 2D displays that simulate 3D interactions.

*Automated Homes:* To control electronic devices in your house with gestures. Gesture recognition system uses gestures such as waving your arms, punching, and kicking. It can help you turn out lights, control the television, music system etc.

*Switching channels without a remote*: Support motion control as you can switch channels, change volume, play games, pause videos, and even surf the web all using hand gestures.

*Image controlling and scaling:* Zooming in and out of image & document can be done by gesture.

### C. Process of Gesture Recognition

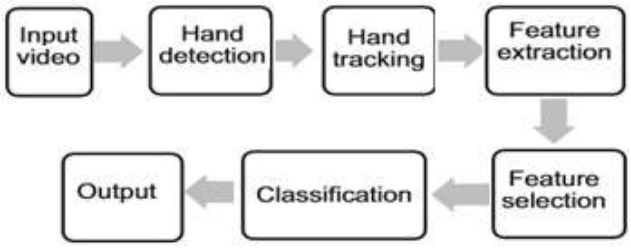


**Figure 1:** The process of gesture recognition

Input video: The Input which is the Raw data is collected basically from camera or by input devices worn by user i.e. sensors.
Hand detection: In this phase segmentation of hand from the background. Edge detection and removal of noise is done.

Algorithms that can be used are:
- Histogram of Oriented Gradients(HOG) [6][9]
- Haar-like features [10]
- Scale-Invariant Feature Transform (SIFT)
- Speeded Up Robust Features(SURF)

Hand tracking: Movement is tracked. Movements can be very fast and their appearance can change vastly within a few frames.

Algorithms that can be used:
- Mean shift [7] [13]
- Particle Swarm Optimization(PSO)

Feature extraction and selection: Features are useful information that can be extracted from the segmented hand by which machine can understand meaning of posture. Selection phase provides useful information related to image.

Classification: It is gesture recognition phases in which given input gesture is identified.

Algorithms that can be used are:
- Deep Belief network (DBN) [2]
- Support vector machine (SVM) [1] [17]
- Dynamic time wrapping (DTW)
- K-nearest Neighbour (KNN)
- Hidden Markov Model (HMM)

## 3. Techniques Used for Hand Gesture Recognition

Glove-based and Vision based technique is used to collect hand configuration and hand movements.

**a) Gloved based approaches[4]:**
In this approach a device of glove type is worn by the user. This glove like device is having sensors which sense the movements of finger and hands. The information is then passed to the computer for processing. Glove approach uses the sensors for capturing the hand position and to track motion. Detection of hand is done by the sensors on hand and the correct coordinates of location of palm and fingers is to be find out using the sensors on the gloves. The main disadvantage of glove based devices is the health hazards, which are caused by its devices like mechanical sensor material which raises symptoms of allergy, magnetic devices which raises risk of cancer etc. Degree of freedom is less in glove based. Use of sensors makes glove based technique costly.

**b) Vision based approaches[3]:**
Recognition system only requires the camera to capture the image for natural interaction between human and computers. That is more useful in real time applications. This approach as compare to data glove based approach is the convenient, simple and natural. Although this technique is simple but there are lots of challenges while implementing such as complex background, lighting conditions and variations, and other skin color objects with the hand object. Vision based technology deals with the characteristics of image such as texture and color that are required for the identification of gesture.

**c) Comparison of Techniques Used [4][5]**

**Table 1:** Comparison table between Glove-based technique and vision based technique

| Criterion | Glove-based Technique | Vision based Technique |
|---|---|---|
| Cost | High | Less |
| User Comfort | Less | High |
| Portability | Less | High |
| User Co-operation | High | Less |
| Health issues | Yes/No | No |
| Hand anatomy | Restriction high | Less |

## 4. Phases in Hand Recognition

Three phases of gesture recognition

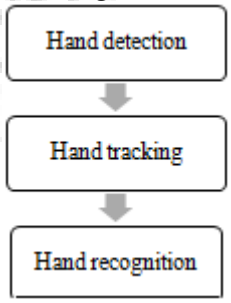


**Figure 2:** Phases in hand gesture recognition

**a) Hand detection [4]**
The primary step in hand gesture recognition systems is the detection of hands and the segmentation of the corresponding image regions. This segmentation is crucial because it isolates the task-relevant data from the image background, before passing them to the subsequent tracking and recognition stages. Skin color, motion, and depth information is used for hand detection. Skin color segmentation has been utilized by several approaches for hand detection. A majordecision towards providing a model of skin color is the selection of the color space to be employed. Several color spaces have been proposed including RGB, normalized RGB, HSV, YCrCb, YUV etc.

Algorithm used for hand detection are Histogram of oriented gradients(HOG) and Haar-based algorithm.

***Histogram of Oriented Gradient (HOG) [8]:***

Histogram of oriented gradients (HOG) is a feature descriptor used to detect objects in computer vision and image processing. The HOG descriptor technique counts occurrences of gradient orientation in localized portions of an image - detection window, or region of interest. First used for application of person detection [6].

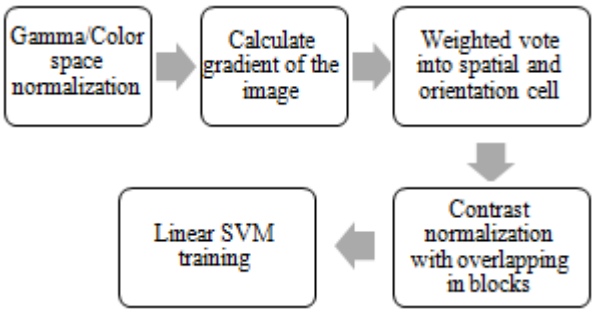


**Figure 3:** Process of Histogram of oriented gradients (HOG)

Gamma/Color space normalization: Gamma correction method to normalize input image color space. The aim is to adjust the contrast of the image, reduce the impact of local image shadows and illumination changes, at the same time, suppressing the noise interference.

Calculate gradient of the image: In this step the gradient of each pixel in the image (including the size and direction) to capture the contour information, and further weakening of the interference of illumination. Masks have been used to generate gradient field of images. Simple 1-D filter [ -1, 0, 1] performs best. Using simple filter increases accuracy and decreases complexity. Sobel Mask can also be used for calculation of gradient of image.

Weighted Vote: Vote is function of gradient magnitude, it strengthen the contrast between sharp and vague background.

Normalization: It can be done by using

$$\text{L1-norm: } f = \frac{v}{(\|v\|_1 + e)}$$

where v is non normalized vector, $\|v\|_k$ is k-norm, e is constant.

SVM classifier: The final step in gesture recognition using histogram of oriented Gradient descriptors is to feed the descriptors into some recognition system based on supervised learning.

Pros and Cons of Histogram of oriented gradient:

**Pros**

- Hog feature are not affected by varying light condition [11].
- False positive per window is reduce [11].

**Cons**

- It is computationally complex [12].

**Haar-like feature[10]:**

Haar-like feature uses simple luminance information or color information, to speed up the computation of feature values.Viola and Jones proposed a statistical approach to manage the variety of human faces. This famous face detector architecture called Haar like feature to describe hand. This is very simple algorithm because rather than using the intensity values of a pixel, they use the change in contrast values between adjacent rectangular groups of pixels. There are four different haar like feature such as Edge feature, Line feature, Center surround feature and Diagonal feature.

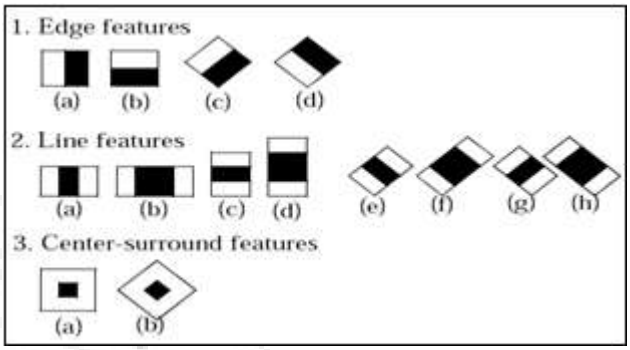


**Figure 4:** Different Haar-like features

The Haar-like features describe the ratio between the dark and bright areas within a kernel. One typicalexample is that the eye region on the human face is darker than the cheek region, and one Haar-like feature can efficiently catch that characteristic. The motivation is that a Haar-like feature-based system can operate much faster than a pixel based system. Haar-like feature-based system can operate much faster than a pixel based system. It describes the ratio between the dark and bright areas within a kernel The Haar-like features are also relatively robust to noise and various lighting condition because they compute the gray-level difference between the white and black rectangles.

It considers adjacent rectangular regions at a specific location in a detection window, sums up the pixel intensities in each region and calculates the difference between these sums. This difference is then used to categorize subsections of an image.

$$F(x) = \sum_{\text{Black}} (\text{pixel value}) - \sum_{\text{white}} (\text{pixel value})$$

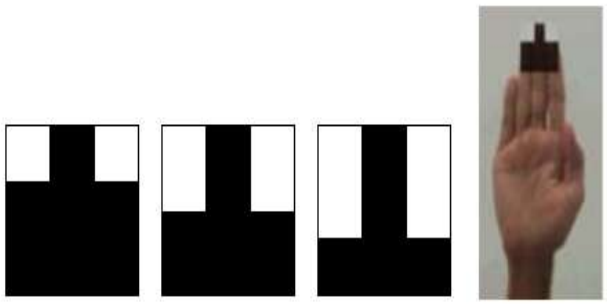

**Figure 5:** Applying Hand-like feature on hand

Pros and Cons of Haar-like feature

Pros

- It is efficient and detection time required is less than hog [11].

Cons
- It is sensitive to illumination change [12].

**Comparison of hand detection algorithms [11] [12]:**

**Table 2:** Comparison table between Histogram of Oriented gradient and Haar-like feature

| Parameters | Histogram of Oriented Gradient (HOG) | Haar-like feature |
|---|---|---|
| Detection speed | 134.209 ms/frame | 77.46 ms/frame |
| False rate | 0.0438 | 0.2017 |
| Detection rates | 0.820 | 0.933 |
| Accuracy | Low | High |
| Stable | Low | High |

**b) Hand Tracking[4]**

Hand tracking is second phase in hand gesture recognition. Tracking can be defined as the frame-to-frame correspondence of the segmented hand regions or features towards understanding the observed hand movements. It provides the inter-frame linking of hand/finger appearances, giving rise to trajectories of features in time. These trajectories convey essential information regarding the gesture.

**Mean Shift Algorithm[7][13]:**

Mean Shift is one of the most successfully used nonparametric deterministic optimization methods.It is also called as mode seeking algorithm. Mean shift is the most powerful clustering technique. It is quite similar to k-mean algorithm. In k-mean number of clusters are fixed but in mean shift it can vary as needed.Mean shift algorithm uses iterative method to find the maximum of probability density estimation in the neighborhood, i.e.the target position. Target region of the tracking algorithm is usually chosen by a rectangular or ellipsoidal region. It is based on the kernel density function estimation, which utilize an iterative method to find the maximum value of the estimated probability density. Mean shift is used for image segmentation, clustering, visual tracking, space analysis.

$$y_1 = \frac{\sum_{i=1}^{M} x_i w_i g\left(\left\|\frac{x_i - y_0}{h}\right\|^2\right)}{\sum_{j=1}^{M} w_j g\left(\left\|\frac{x_j - y_0}{h}\right\|^2\right)}$$

where h indicates the scale of target window, $x_i$ denotes the $i^{th}$ pixel in the target area having the centre point $y_0$, M is the number of pixels in the target area and $w_i$ is the weight of $x_i$. For tracking, $y_1$ is initially found and then this $y_1$ becomes $y_0$ for the next iteration. This process is repeated until $\|y_1 - y_0\|$ is less than a certain threshold.

Steps in mean shift tracking algorithm:
1) Place the window at random position in the target area.
2) Find centre of mass in the target area.
3) Repeat till it convergences.

Pros and Cons of mean shift algorithm

**Pros**
- It does not require prior knowledge of the number of clusters.

**Cons**
- It was unable to detect hand when background had similar color as object[3].
- Window size is a crucial parameter in the performance of the Mean-shift algorithm[7].

**c) Hand Recognition[2]**

Hand recognition is the last phase in gesture recognition system. In this phase, the recognition of a hand trajectory from image sequences obtained over a period of time. Dynamic gesture recognition can be split into 2 types: rule-based approaches and statistical model-based approaches. Rule-based approaches require the definition of the overall structure of every gesture carefully and are usually used for simple directive gestures, such as up, down, left, and right etc. Statistical model-based approaches apply different statistical learning theories, such as Hidden markov model (HMM), Dynamic time wraping (DTW) and neutral network.

**Deep Belief Network (DBN) [2][14]:**

Deep belief network is stack of Restricted boltmann machine (RBM). Deep belief network is deep learning algorithm. DBN works as a generative model consisting of three kinds of layers: label data, hidden layers, and the visible layer. Lines connecting these layers are composed of softmax regression, RBM (Restricted Boltzmann Machine), and directed belief nets. Softmax regression is used for obtaining the output from the hidden layer 'h3'. RBM is an effective mechanism for mining the potential relation between hidden layers by applying contrastive divergence learning. A belief net is a directed acyclic graph composed of stochastic variables.

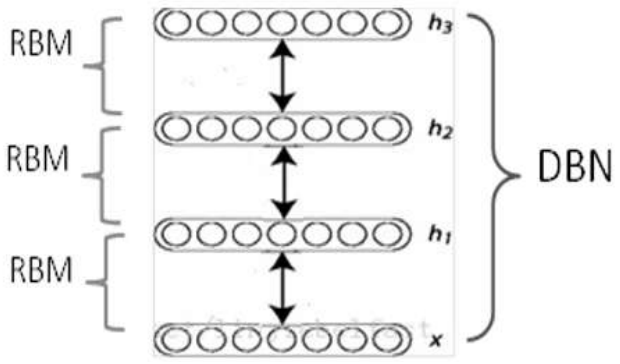


**Figure 6:** Structure of Deep Belief Network

**Restricted Boltzmann Machine**

RBM is a two-layer stochastic network. Its two layers are visible layer $\boldsymbol{v}$ and hidden layer $\boldsymbol{h}$, whose visible layer has 4 nodes and hidden layer has 3 nodes. RBMs learning process is unsupervised learning. So the Deep Belief Network can only work without supervising.

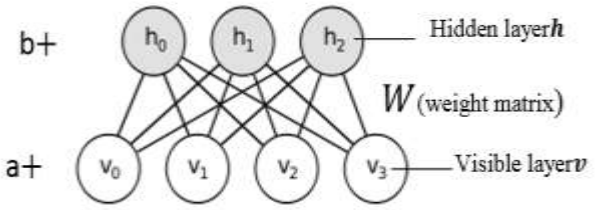


**Figure 7:** Structure of Restricted Boltzmann Machine

Energy Function:

$$E(v,h) = -\sum_{i \in v} a_i v_i - \sum_{j \in h} b_j h_j - \sum_{i,j} v_i h_j w_{ij}$$

Probability Distribution:

$$p(v,h) = \frac{1}{z} e^{-E(v,h)}$$

The principle of greedy layer-wise unsupervised training can be applied to DBNs with RBMs as the building blocks for each layer. The process is as follows:

1) Train the first layer as an RBM that models the raw input $x = h^{(0)}$ as its visible layer.
2) Use that first layer to obtain a representation of the input that will be used as data for the second layer. Two common solutions exist. This representation can be chosen as being the mean activations $p(h^{(1)}=1 | h^{(0)})$ or samples of $p(h^{(1)} | h^{(0)})$ .
3) Train the second layer as an RBM, taking the transformed data (samples or mean activations) as training examples (for the visible layer of that RBM).
4) Iterate (2 and 3) for the desired number of layers, each time propagating upward either samples or mean values.
5) Fine-tune all the parameters of this deep architecture with respect to a proxy for the DBN log likelihood, or with respect to a supervised training criterion (after adding extra learning machinery to convert the learned representation into supervised predictions, e.g. a linear classifier).

Pros and Cons of Deep Belief Network

**Pros**

- Robust to noise[15].

**Cons**

- Time consuming and computationally expensive as it is made of RBM[16].

**Support Vector Machine (SVM)[1][17]:**

Support Vector Machine (SVM) is a classification and regression prediction tool that uses machine learning theory to maximize predictive accuracy while automatically avoiding over-fit to the data.The goal is design a hyperplane that classifies all training vectors in classes.It is being used for many applications, such as hand writing analysis, face analysis and so forth, especially for pattern classification and regression based applications.The operation of the SVM algorithm is based on finding the hyperplane that gives the largest minimum distance to the training examples. The optimal separating hyperplane maximizes the margin of the training data.

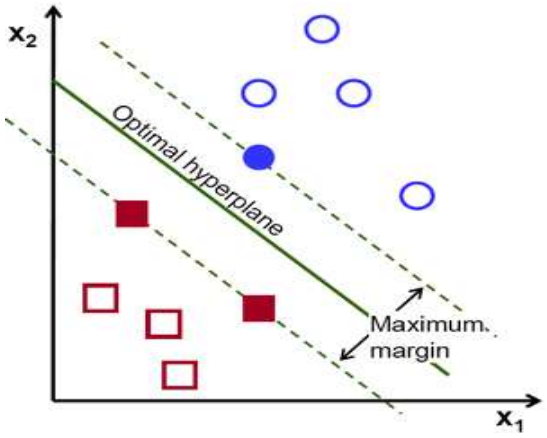


**Figure 8:** Representation of Support Vector

Hyperplane equation :

Distance between two planes:

$$\vec{w} \cdot \vec{x} + b = 1 \qquad \frac{2}{||\vec{w}||}$$

and

$$\vec{w} \cdot \vec{x} + b = -1$$

Pros and Cons of Support Vector Machine

a) Pros

The major strengths of SVM are the training is relatively easy. No local optimal, unlike in neural networks [17].

b) Cons

- Challenges is that of choosing an appropriate kernel for the given application [17].
- Long training time.

**Comparison of hand recognition algorithms[18]:**

**Table 3:** Comparison table between Deep Belief Network and Support Vector Machine

| Parameters | Deep Belief Network (DBN) | Support Vector Machine (SVM) |
|---|---|---|
| Accuracy | 98.12% | 87.58% |
| Recognition Time | 0.000899 | 0.015259 |
| Application | Image reconstruction and classification [16], Text analysis[15], Handwriting recognition | Pedestrian Detection |

## 5. Conclusion

Dynamic hand gesture recognition algorithms are compared with different parameters. Survey of Pros and Cons of Algorithm is mentioned. It is seen that different algorithms can be used together in various application.

## References

[1] Akanksha Singh I, Saloni Arora, Pushkar Shukla, Ankush MittaIndian, "Sign *Language Gesture Classification as Single or Double Handed Gesture*", Third International Conference on Image Information Processing,2015.

[2] Liu Ni, Muhammad Ali Abdul Aziz "*A robust deep belief network-based approach for recognizing dynamic hand gestures*" , Proceedings of 2016 13th International Bhurban Conference on Applied Sciences & Technology,Islamabad,Pakistan,12th-16th January,2016
[3] Joyeeta Singha, Amarjit Roy1 ,Rabul Hussain Laskar ,"*Dynamic hand gesture recognition using vision-based approach for human–computer interaction*", The Natural Computing Applications Forum 2016
[4] Siddharth S. Rautaray, Anupam Agrawal, "*Vision based hand gesture recognition for human computer interaction: a survey*", Springer Science+Business Media Dordrecht 2012
[5] R.Pradipa, Ms S.Kavitha,"*Hand Gesture Recognition – Analysis of Various Techniques, Methods and Their Algorithms*", ICIET'14
[6] Navneet Dalal and Bill Triggs ,"*Histograms of Oriented Gradients for Human Detection"*, Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)
[7] Farhad Dadgostar, Abdolhossein Sarrafzadeh, Scott P. Overmyer, Liyanage De Silva "*Is the Hand really quicker than the Eye? Variances of the Mean-Shift algorithm for real-time hand and face tracking*", International Conference on Computational Intelligence for Modelling Control and Automation,and International Conference on Intelligent Agents,Web Technologies and Internet Commerce (CIMCA-IAWTIC'06)
[8] Cun Hang, Fei Hu, Aboul Ella Hassanient, and Kai Xiao ,"*Texture-based Rotation-Invariant Histograms of Oriented Gradients*", 2015 IEEE
[9] Kai-ping Feng , Fang Yuan, "*Static Hand Gesture Recognition Based on HOG Characters and Support Vector Machines*", 2nd International Symposium on Instrumentation and Measurement, Sensor Network and Automation (IMSNA),2013
[10] Ruchi. M. Gurav, Premanand K. Kadbe ,"*Vision Based Hand Gesture Recognition with Haar Classifier and AdaBoost Algorithm*", International Journal of Latest Trends in Engineering and Technology 2015
[11] Kuizhi Mein, Lu Xu, Boliang Li, Bin Lin, Fang Wang,"*A real-time hand detection system based on multi-feature*" , Neurocomputing 158 (2015)
[12] Chong Xia, Shui-Fa Sun, Peng Chen,Heng Luo, and Fang-Min Dong,"*Haar-Like and HOG Fusion Based Object Tracking*", Springer International Publishing Switzerland 2014
[13] Youness Aliyari Ghassabeh, Tamás Linder, Glen Takahara "*On The Convergence and Application of Mean Shift Type Algorithm*", 25th IEEE Canadian Conference on Electrical and Computer Engineering 2012
[14] Yuming Hua, Junhai Guo, Hua Zhao , "*Deep Belief Networks and Deep Learning" International Conference on Intelligent Computing and Internet of Things 2015*
[15] Anupama Ray, Sai Rajeshwar,Santanu Chaudhury, "*Scene Text Analysis using Deep Belief Networks*",ACM 2014
[16] Amin Emamzadeh Eshmaeili Nejad,"*An Application Of Deep Belief Networks for 3-Dimensional Image Reconstruction"* Indian J.Sci.Res. 7 (1): 618-625, 2014
[17] Vikramaditya Jakkula, "Support Vector Machine"
[18] Ao Tang, Ke Lu, Yufei Wang Jie Huang, and Houqiang Li, "A *Real-Time Hand Posture Recognition System Using Deep Neural Networks*", ACM Transactions on Intelligent Systems and Technology, Vol. 6, No. 2, Article 21, Publication date: March 2015