IJST EXPLORE International Journal of Science and Technology
CENTRE OF PROFESSIONAL RESEARCH PUBLICATIONS
ISSN 2049-7318

International Journal of Science and Technology Volume 2 No. 6, June, 2013

Dynamic Hand Gesture Recognition Using Neural Networks

Parul Vashist, K.Hema
GNIOT ,Greater Noida

ABSTRACT

Gesture recognition is important for developing an attractive alternative to prevalent human–computer interaction modalities. Neural networks, with their remarkable ability to derive meaning from complicated or imprecise data, can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques. A trained neural network can be thought of as an "expert" in the category of information it has been given to analyze. This expert can then be used to provide projections given new situations of interest and answer "what if" questions. Neural networks take a different approach to problem solving than that of conventional computers. In this paper we focus on these problems: How to adapt the hand model to specific target? How to establish correspondences and combine (fuse) image data from multiple cameras in a 3-D framework? How good an algorithm handles occlusions and performs in highly cluttered environment? How to interpret the semantic meanings of a hand gesture dynamically?

1. INTRODUCTION

Artificial Neural Network (ANN) is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information. The key element of this paradigm is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements (neurons) working in unison to solve specific problems. ANNs, like people, learn by example. An ANN is configured for a specific application, such as pattern recognition or data classification, through a learning process. Learning in biological systems involves adjustments to the synaptic connections that exist between the neurones. This is true of ANNs as well.

1.1 Advantages of ANN

1. Adaptive learning: An ability to learn how to do tasks based on the data given for training or initial experience.

2. Self-Organisation: An ANN can create its own organisation or representation of the information it receives during learning time.

3. Real Time Operation: ANN computations may be carried out in parallel, and special hardware devices are being designed and manufactured which take advantage of this capability.

4. Fault Tolerance via Redundant Information Coding: Partial destruction of a network leads to the corresponding degradation of performance. However, some network capabilities may be retained even with major network damage.

2. GESTURE

A movement of a limb or the body as an expression of thought or feeling.

2.1 Human Computer Interface using Gesture

-Replace mouse and keyboard
-Pointing gestures
-Navigate in a virtual environment
-Pick up and manipulate virtual objects
-Interact with a 3D world
-No physical contact with computer
-Communicate at a distance

2.2 Hand Gestures

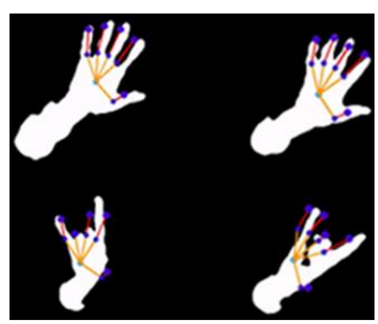


Fig1:-Hand gestures

Hand gesture recognition systems to successfully meet the demands of verification applications it is necessary to develop testing and scoring procedures that specifically address these applications this process would probably start with image processing techniques such as noise removal, followed by (low- level) feature extraction to locate lines, regions and possibly areas with certain textures.

The Purpose of human gestures are conversational, controlling, manipulative, and communicative. For Hand gesture recognition systems to successfully meet the demands of verification applications it is necessary to develop testing and scoring procedures that specifically address these applications this process would probably start with image processing techniques such as noise removal, followed by (low-level) feature extraction to locate lines, regions and possibly areas with certain textures.

2.3 The Challenges of Vision-based Hand Gesture recognition are

-Highly articulated, with many joints and high DOFs
-Highly constrained: static and dynamic constraints, hard to model
-Two representations: Appearance-based and 3-D model-based
-Two steps: Static posture recognition & Gesture understanding (semantics)
-Finger motion constraints are applied to define the ranges each finger may move within.

2.4 Human Hand Modeling

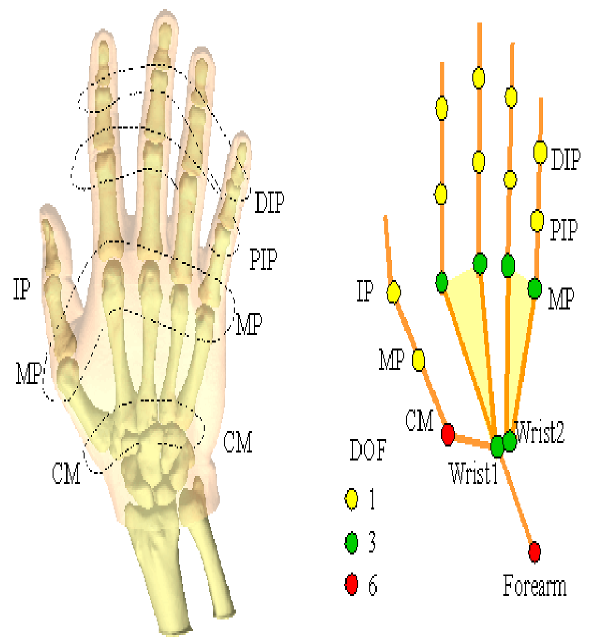
To model human hand motion, it is important to measure and reconstruct human hand postures accurately. Product engineering requires reconstructions to be accurate to within a few millimeters at the finger tip.

An optical motion capture system, in which many reflective markers are put on the skin surface, is used to model hands of various size with adequate degrees of freedom. Due to a hand's high degrees of freedom (about 30), only a limited number of markers can be used to avoid occlusion and subsequent failure of marker identification. In addition, skin movement over a human hand is relatively large for each joint's range of motion. Also, the previous technique of link identification developed for the entire body cannot be used effectively. Therefore, we propose a link identification method from calibrating motions and associating them with a catalogued set of calibrated motions and related anatomical knowledge.

2.5 Link Model of a Hand

When we deal with a hand represented as a link one of the characteristic points is how to deal with the palm. The palm is divided into two parts so that the palmer arch may be studied without directly considering the CM joints.

Hand joints has the following names which inspired from their location, Distal Interphalangeal (DIP) and Proximal Interphalangeal (PIP) used for forming a single finger except for thumb which has neither distal nor proximal since the count of his bones are less by one comparing with other fingers, so,it has just Interphalangeal (IP), the connection joint between the fingers and metacarpal bones is called Metacarpophalangeal (MCP)which are five one for each finger, finally, the connection joint between the metacarpal bones and carpal bones (wrist) are called Trapeziometacarpal (TM) or also called Carpometacarpal (CMC), these different forms of connections have a different DoF as follows: DIP's have 4 DoF, PIP's have 4 DoF, one for IP, the total is 9 hitherto, MCP's have 2 DoF each, this will bring up the sum into 19 DoF, two more DoF for TM joint for thumb finger only and 6 DoF for wrist, so, 27 DoF are provided by this mystery creature, out of question, and outstanding human hand



(a) (b)

(a)Human Hand Bones (b)Link Model

Fig.2 Link Model

2.6 Hand Contour

We have implemented a contour-based hand tracker, which combines two techniques called condensation and partitioned sampling. During tracking, we record the trajectory of the hand which will be used in the hand recognition stage. A fourteen-dimension state vector is used to describe the dynamics of the hand contour:

$$\chi = (tx, ty, \alpha, s, \theta_L, l_L, \theta_R, l_R, \theta_M, l_M, \theta_I, l_I, \theta_{\mathrm{Th1}}, \theta_{\mathrm{Th2}})$$

where the subvector (tx, ty, α, s) is a nonlinear representation of a Euclidean similarity transform applied to the whole hand contour template, (tx, ty) is the palm center. (θ_L, L_l)represents the nonrigid movement of the little finger, θL means the little finger's angle with respect to the palm, and l_L means the little finger's length relative to its original length .

The human hand consists of 27 bones .Eight of these are located within the wrist, and four make up the palm. We ignore both the bones within the wrist and those in the palm ,and represent the palm as a single segment in our model. Thus, we consider 3 × 5 = 15 phalanges for the five fingers, plus the palm which is defined as the remaining hand above the wrist. As the wrist itself is irrelevant in terms of object grasping we do not include it in the model. The degrees of freedom between hand segments are constrained it in the model. The degrees of freedom between hand segments are constrained by revolute joints: hinge joints between finger phalanges allow only for bending within a certain range (*flex*, 1 DOF); saddle joints connecting the fingers to the palm additionally allow for spreading the fingers (*flex* and *abduction*, 2DOF). Neither hinge nor saddle joints permit a *twist* around the bone axis.

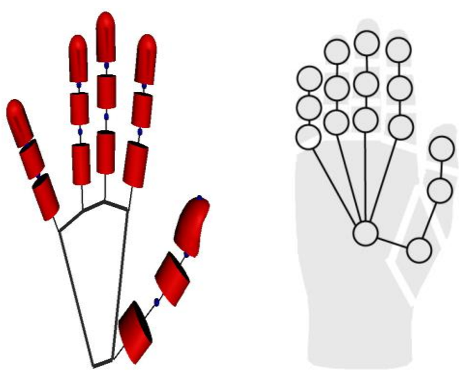
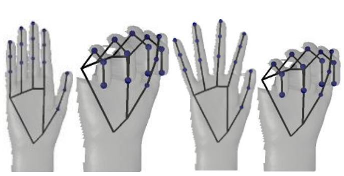


Fig3:-Hand Contour

3. OCCLUSION

Occlusion by an object. To verify the robustness of our method in the presence of an occluding object, we have introduced artificial occluders into a sequence of 45 frames. The hand first reads and then returns to its initial pose Fig. 4 demonstrates the seven tested degrees of occlusion, ranging from no occlusion to full occlusion. The error over all hand segments in the different occlusion scenarios is plotted . Up to occlusion level four there is almost no increase of the error. At higher levels fingers are fully occluded so their state has to be hallucinated, based only on the anatomic constraints. The system can hardly be blamed for the larger errors in such situations.



(a) Two fists

(b) Spreading

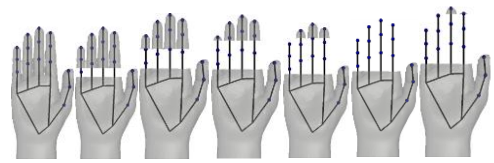


Fig4:-Occlusions

4. 3D MODEL BASED ALGORITHM

The 3D model approach can use volumetric or skeletal models, or even a combination of the two. Volumetric approaches have been heavily used in computer animation industry and for computer vision purposes. The models are generally created of complicated 3D surfaces, like NURBS or polygon meshes.

The drawback of this method is that is very computational intensive, and systems for live analysis are still to be developed. For the moment, a more interesting approach would be to map simple primitive objects to the person's most important body parts (for example cylinders for the arms and neck, sphere for the head) and analyse the way these interact with each other. Furthermore, some abstract structures like super-quadrics and generalised cylinders may be even more suitable for approximating the body parts. The very exciting about this approach is that the parameters for these objects are quite simple. In order to better model the relation between these, we make use of constraints and hierarchies between our objects.

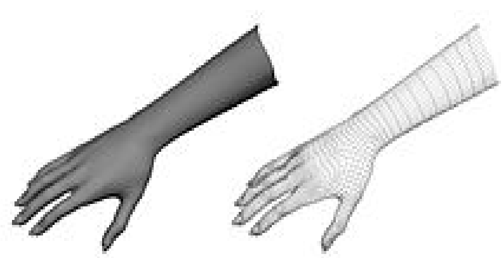


Fig5:-3 D Model

A real hand (left) is interpreted as a collection of vertices and lines in the 3D mesh version (right), and the software uses their relative position and interaction in order to infer the gesture.

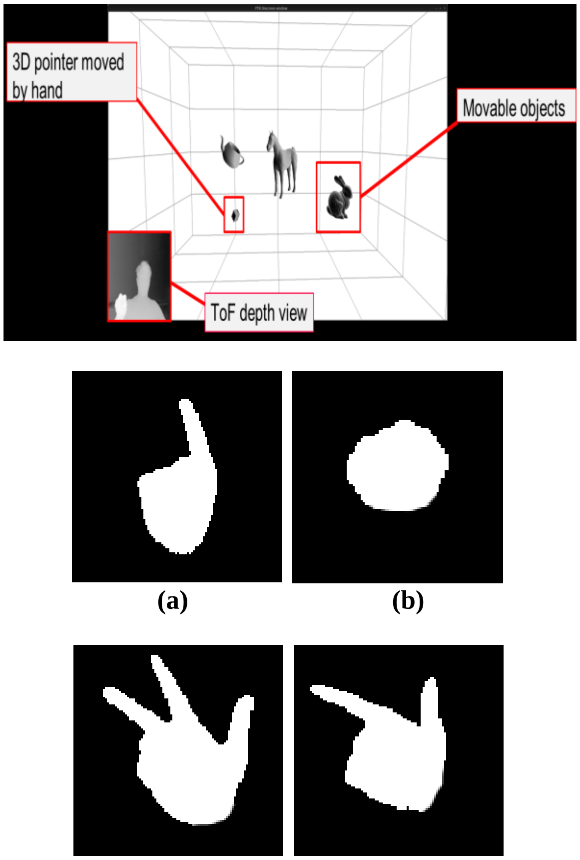


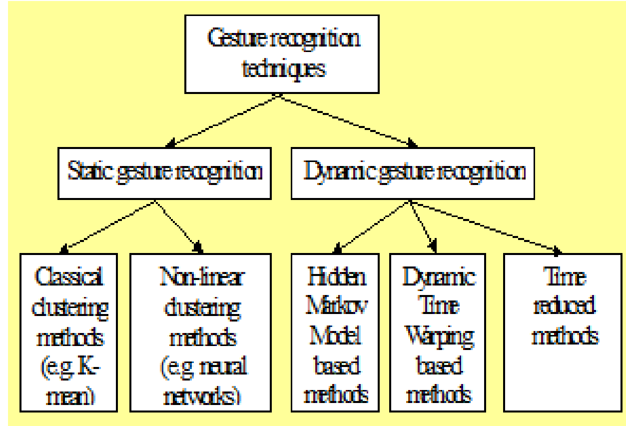
Fig.6:Screenshot of the virtual 3D desktop environment (a). The pointer is moved by the hand in 3D and its function (moving (b), grabbing (c), rotating (d), scaling (e)) is changed according to the hand gesture

5. MODEL-BASED HAND POSTURE RECOGNITION

5.1 Gesture Recognition Techniques

5.1.1 Dynamic Gestures

A dynamic hand gesture comprises a sequence of hand shapes with associated spatial transformation parameters (such as translation, rotation, scaling/depth variations etc.) that describe the hand trajectory. Gesture recognition schemes can be broadly classified into two groups. In the first approach, a gesture is modeled as a time sequence of states. Here, one uses Hidden Markov models (HMM), discrete finite state machines (DFA), and variants thereof for gesture recognition. In the second approach, one uses dynamic time warping to compensate for the speed variations (undulations in the temporal domain) that occur during gesticulation. Gesture recognition schemes can also be categorised on the basis of the parameters that are used to model the appearance of the hand e.g., hand silhouette- based model, graph-based model, use of Fourier descriptors, b-splines etc.



To recognize continuous dynamic hand gesture:

- Design of gesture command set and interaction model.
- Real-time segmentation of gesture streams.
- Modeling, analysis, and recognition of gestures.
- Real-time processing is mandatory for practically using hand gestures in HCI
- Dynamic gestures have been handled using tracking framework.

Continuous classical left to right Hidden Markov Models (HMMs) with their excellent dynamical time warping capabilities and recognition performance are utilized to handle dynamic gestures . With this paradigm the robust recognition of gestures is guaranteed no matter how fast or slow they are expressed.

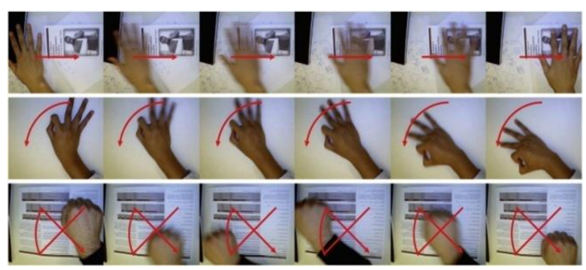


Fig7:-HMM

5.2 Dynamic Gesture Representation

A dynamic hand gesture is a spatial-temporal pattern and has four basic features: velocity, shape, location, and orientation. The motion of the hand can be described as a temporal sequence of points with respect to the hand centroid of the person performing the gesture. The hand shape is not considered and each dynamic hand gesture instance is represented by a time series of the hand's location:

$$P_t = (x_t, y_t), \qquad t=(1,2,\ldots T)$$

where *T* represents the length of gesture path and varies across different gesture instances. Consequently, a gesture containing an ordered set of points can be regarded as a mapping from time to location.

5.3 The Continuous Hand Gesture Recognition Scheme

In this paper, we consider online-continuous-handed dynamic gestures based on discrete HMM. The hand gesture recognition system consists of three major parts: palm detection, hand tracking, and trajectory recognition. Figure 3 shows the whole process. The hand tracking function is trigged when the system detects an opened hand before the camera; the hand gesture classification based on HMM is activated when the user finishing the gesture.

- The basic algorithmic framework for our recognition process is the following.
- Detect the palm from video and initialize the tracker with the template of hand shape.
- Track the hand motion using a contour-based tracker and record the trajectory of palm center.
- Extract the discrete vector feature from gesture path by the global and local feature quantization.
- Classify the gesture using HMM which gives maximum probability of occurrence of observation sequence.

5.4 Hand Detection and Tracking

We calculate the HOG features of a new observed image to detect the opened hand at different scales and location. When the hand is detected, we update the hand color model which will be used in hand tracking. The system requires user to keep his palm opened vertically and statically before the palm is captured by the detection algorithm. In this paper, we have considered single handed dynamic gestures. A gesture is composed of a sequence of epochs. Each epoch is characterized by the motion of distinct hand shapes.

Fig 8: Epoch

5.5 Dynamic Time Warping

DTW assumes that the endpoints of the two patterns have been accurately located and formulates the pattern matching problem as finding the optimal path from the start to the end on a finite grid. The optimal path can be found efficiently by dynamic programming.

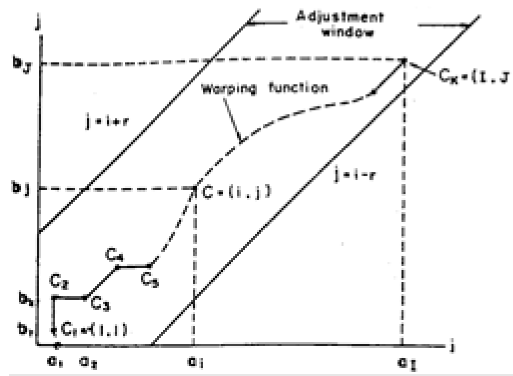


Fig9:-Time Wraping

A modified Dynamic Time Warping algorithm is suggested for eliminating time variation of spatio-temporal appearance patterns due to various gesturing rates.

- It is a kind of non-linear re-sampling technique.
- It can reserve necessary temporal information and spatial distribution of original patterns.

6. CONCLUSION

Hand gesture recognition is challenging, due to its complex articulate and constraints, high DOF, and heavy self-occlusion. 3-D model-based recognition is suitable in multi-camera vision-based systems. Global configuration of hand should be determined first to reduce the search space. Particle filtering and tree-based searching help improve tracking robustness and conquer the computation hurdles. The proposed algorithm can reach better recognition results than the traditional hand recognition method. However, the tracking algorithm is still very sensitive to light and the system can only report the detection until a gesture reaches its end.

7. FUTURE WORK

We currently assume that the moving skin color region in the scene is the gesturing hand, which could be invalid when there appears a moving human face. Exploiting simple geometrical model of human body can alleviate this problem, in that case multiple cameras can be necessary. To practically use hand gestures in HCI, more gestural commands will be needed. Some kind of commands would be more reasonably input by static hand gestures (hand postures). Cooperating hand gesture recognition into multi-modal interface (MMI) is our next work.

REFERENCES

[1] Ying Wu and Thomas S. Huang, Hand modeling, analysis and recognition For Vision-Based Human Computer Interaction. IEEE Signal Processing Mag, May 2001, p. 51-60

[2] A. Erol, et al, Vision-based hand pose estimation: A review. Computer Vision and Image Understanding 108 (2007) 52–73

[3] M. Potamias and V. Athitsos, Nearest Neighbor Search Methods for Handshape Recognition. *PETRA'08* July 1519, 2008, Athens, Greece

[4] D. P. Huttenlocher, et. al., Comparing Images Using the Hausdorff Distance. IEEE Trans, PAMI 15 (9) (Sept 1993) 850–863

[5] H.G. Barrow, et. al., Parametric Correspondence and Chamfer Matching: Two New Techniques for Image Matching, NASA Technical Report, Vision-7, p.659-670.

[6] S. Ekvall and D. Kragic. Grasp recognition for programming by demonstration tasks. In *ICRA*, 2005.

[7] A. Erol, G. Bebis, M. Nicolescu, R. D. Boyle, and X. Twombly. Vision-based hand pose estimation: A review.*CVIU*, 108(1-2):52–73, 2007.

[8] K.-H. Jo, Y. Kuno, and Y. Shirai. Manipulative hand gesture recognition using task knowledge for human computer interaction. In *FG*, 1998.

[9] H. Kjellstr¨om, J. Romero, D. M. Mercado, and D. Kragic.Simultaneous visual recognition of manipulation actions and manipulated objects. In *ECCV*, 2008.

[10] B. Leibe, A. Leonardis, and B. Schiele. Combined object categorization and segmentation with an implicit shape model. In *ECCV Workshop*, 2004.

[11] D. G. Lowe. Object recognition from local scale-invariant features. In *ICCV*, 1999.

[12] J. Maccormick and M. Isard.Partitioned sampling, articulated objects, and interface-quality hand tracking. In *ECCV*, 2000.

[13] R. Mann, A. Jepson, and J. M. Siskind. The computational perception of scene dynamics. 65(2):113–128, 1997.

[14] T. Oggier, B. B¨uttgen, and F. Lustenberger. Swissrangersr3000 and first experiences based on miniaturized 3d-t of cameras. Technical report, CSEM, 2005.

[15] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. 1988.

[16] B. Stenger, A. Thayananthan, P. Torr, and R. Cipolla. Modelbased hand tracking using a hierarchical bayesian filter. *PAMI*, 28(9):1372–1384, 2006.

[17] E. B.Sudderth, M. I. Mandel,W. T. Freeman, and A.S.Willsky.Visual hand tracking using nonparametric belief propagation. In *CVPR*, 2004.

[18] E. B. Sudderth, Michael, W. T. Freeman, and A. S. Willsky. Distributed occlusion reasoning for tracking with nonparametric belief propagation. In *NIPS*, 2004

[19] T.Weise, B. Leibe, and L. Van Gool. Fast 3d scanning with automatic motion compensation. In *CVPR*, 2007.

[20] Y.Wu and T. S. Huang. Hand modeling, analysis and recognition. *IEEE Signal Proc Mag*, (3):51–60, 2001.

[21] Y. Wu, J. Y. Lin, and T. S. Huang. Capturing natural hand articulation. In *ICCV*, 2001.

[22] J. S. Yedidia, W. T. Freeman, and Y. Weiss. Understanding belief propagation and its generalizations. In *Exploring artificial intelligence in the new millennium*, 2003.