ISSN 2229-5518

Gesture Recognition Using Wrist Watch

Sonal Singh, Alpika Tripathi

Abstract-Sign Language is the only mode of communication between deaf/dumb and normal human beings. It is a common platform for both but it is not necessarily taught to common person in their school, so there occur a need for an interpreting system. In this paper, we present an approach to maintain the efficiency and accuracy of gesture recognition by identifying the hand gestures from images provided as input to the computer. Previously used methods were using multiple color gloves; sensor gloves, wrist band, but these all were the biggest limitations for its practical implementation. In this paper we will overcome these limitations by using most commonly worn wrist watch by deaf and dumb.

Index terms- Master gesture, Canonical frame, HMM, ANN, VBGMG, NVBG, Template

1.INTRODUCTION

Human gestures play a vital role in human and machine interface. This mode of interaction can lead to overcome various limitations such as mouse, television remote, joystick etc. A gesture of a sign language is a particular hand movement with specific shapes made out of them. Facial gestures also count toward the emotion recognition, as another mode of communication. A posture is a static shape of the hand to indicate a sign. A sign language usually provides signs for whole alphabets and numerals. It also provides signs to some words by using a series of gestures and postures that don't have a corresponding sign in that sign language. So, even sentences can be made using these signs for letters, numerals performing with signs of words is faster. The sign language used here is the American Sign Language.

A vision based approach is on boom, face detection, smile detection, blinked eyes has been implemented in various digital cameras and various social networking sites. Facial recognition is being used for authentication also. But when it comes to communication of deaf and dumb they are left with analyzing their sign language, but it is not necessary that a normal person must be knowing sign language. So here emerges a need of mediator to translate their gestures in textual or audio format which could be understood by normal human being. To figure out location of hand from any image many researchers have used gloves, sensor gloves, multi colored gloves, specific color bands. But once when it comes to implementation they have limitations as wearing this all the time is not possible and these all becomes as a sign to recognize deaf and dumb persons if worn publically so here emerges a need of any common accessory which is used most commonly by deaf and dumb persons**.**

Wrist watch is the solution as they coordinate with the world according to time, keeping a mobile phone to look time is worthless for them. So focusing on wrist watch for hand recognition can lead us in a very helpful way to implement gesture recognition. In vision-based approaches, the system architecture is generally divided mainly into two parts. The first part is the ***feature extraction*****.** This part extract important features from the image using computer vision and does image processing such as background noise subtraction, hand motion tracking, and feature characterization (shape, orientation, and location). The second part is ***recognition*** phase. From the extracted and characterized features, the recognizer learns the pattern from training dataset and recognizes testing data accurately. The recognizer uses and implements machine learning algorithms. Artificial Neural Network (ANN) and Hidden Markov Model (HMM) are the most common and simple algorithms to be used and implemented [1].

2. RELATED WORK

Attempts for automatic recognition of sign language became visible in the 90s. Researches on hand gestures can be classified into two categories: First category depends on electromechanical devices (sensor) that are used to observe different gesture parameters such as identification of hand's position, yaw, and the fingertips and gives numerical values. Systems that use such devices are called wearable computation-based systems. A major limitation with such systems is that they restrict the poser to wear inconvenient devices. Hence the way by which the user interacts with the system will be complicated and less natural. The second one uses visual techniques and image processing techniques (pattern recognition) to create hand gesture recognition systems. Visual gesture recognition systems are further divided into two categories: The first one depends on specially designed gloves with visual markers called "vision based glove-marker gestures **(VBGMG**)" that helps in determining hand position and postures. But using gloves and markers leads to the limitation and do not provide the natural feeling required in human computer interaction systems. Besides, colored gloves increases, the processing complexity.

The next one that is an alternative to the second kind of visual based gesture recognition systems can be called "natural visual-based gesture (**NVBG**)" means visual based gesture without gloves & markers. And this type tries to provide the ultimate convenience and naturalness by using images of bare hands for gesture recognition. Also many researchers have been trying very hard to introduce hand gestures to Human-Machine Interaction field. Year 1992: Charayaphan and Marble developed a way to understand American Sign Language using image processing. Year 1993: Fels and Hinton implemented a system using VPL Data glove Mark with a Polhemus tracker as input device in their neural network system; method was implemented to classify hand gestures. Year 1993: Another system was developed by Banarse using neural networks. That was a vision-based system which recognized hand postures using a neocognitron network which is a neural network based recognition system of the visual cortex of the brain. Year 1995: Heap and Samaria invented active shape models, or "Smart Snakes" technique to identify gestures using

ISSN 2229-5518

computer vision. Year 1995: Starner and Pentland used a view-based approach using a camera to extract two-dimensional features of poser using images as input to HMMs. Year 1996: Kadous developed a system based on power gloves used to recognize a set of 95 isolated Aulsan signs with an accuracy of 80%, with strongly focusing on computationally inexpensive methods. Year1996: Grobel and Assan used HMMs to recognize isolated signs. Year 1997: Vogler and Metaxas used computer vision methods and **HMM**s. Year 1998: Yoshinori, Kang-Hyun, Nobutaka, and Yoshiaki used colored gloves and have proven that using colored gloves faster & easier hand features extraction can be done than simply wearing no gloves at all.

Year 1998: Liang and Ouhyoung developed a continuous recognition of Taiwan sign language using HMMs with a vocabulary between 71 and 250 signs using data glove as an input device. However their system required that gestures performed by the signer to be slow to so that word boundary can be detected. Year 1999: Yang and Ahuja developed dynamic gestures recognition as they used skin color detection and transforms of skin regions which are in motion to figure out the motion path of ASL signs. Using a neural network, they recognized 40 ASL gestures with a success rate of around 96%. But their technique has very high computational cost whenever wrong skin regions are detected. Year 2000: A feature detection technique is employed to extract shapes of hand in sign language recognition by Imagawa, xMatsuo. Taniguchi, Arita, and Igi. They used Eigen method to detect hand shapes which was appearance based. Using a clustering technique, they generate clusters of hands on an eigen space. They have achieved accuracy of around 93%. Year 2000: Symeoinidis used orientation histograms to detect static hand gestures, specifically, a subset of American Sign Language. . Year 2002: Bowden and Sarhadi developed a non-linear model of motion and shapes for tracking American Sign Language [2].

Later a project in Oxford University using a red color wrist band was used to recognize hand from image and fast template matching was used for pattern recognition with an accuracy of 99.1%. [3]

3. PROPOSED METHOD

Our proposed method uses 2-d image captured from normal digital camera, webcam as input. Then watch is detected in the image and image is preprocessed. If watch is not detected then a different methods are used. If watch is detected then our proposed method can be applied on the preprocessed image. Then image is matched using pattern recognition. Then quantization of image is done. Tree technique is used to reduce time complexity and unnecessary matching's done. Accordingly winning neuron will be chosen. The output can be audio or textual.

3.1 WATCH DETECTION IN AN IMAGE

Set A {all pixels of image}
Set B {all skin color pixels}
Set C {wrist watch color pixels}
Preprocessed image = A ∩ (B U C)

Wrists watch pixels can be detected as between the line passing through the adjacent pixels of skin color and a different color (watch) will be nearly parallel to each other.

Fig. 1

3.2 PATTERN RECOGNITION

Now fast template matching using **canonical frame** is done for pattern recognition. Keeping wrist watch always on +ve x-axis will improve the chances of quick recognition as in training set the images will be set in the same format.

3.3 TREE TECHNIQUE

Many symbols are similar and have very minute difference and some are totally different to from others. In this technique we group visually similar symbols in groups and the input image is matched with the group to which it is most likely to be matched. This increases the possibility and chances of match to be increased.

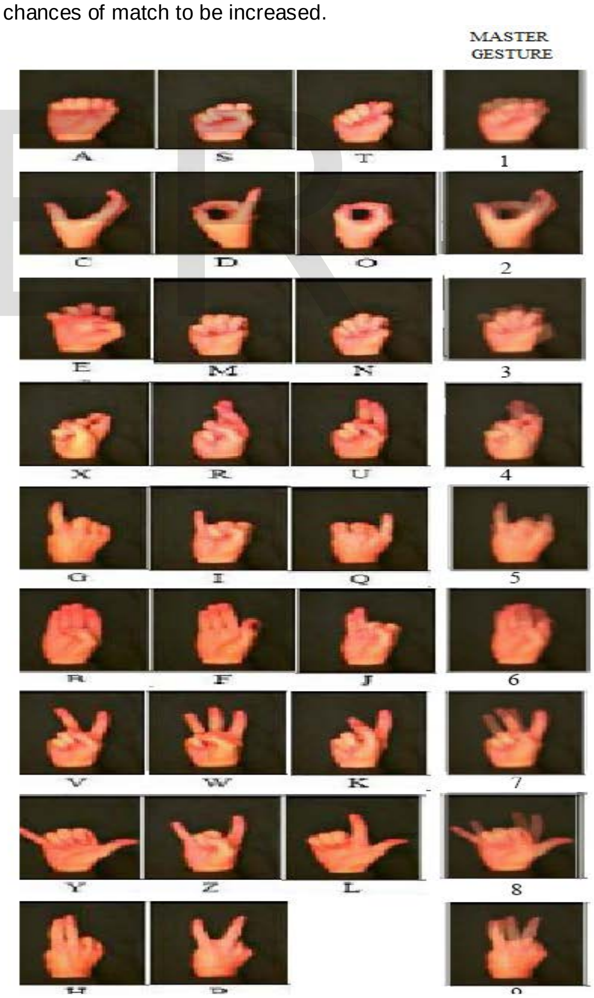

Fig. 2 [4]

Here we have divided 26 alphabets into 9 sets with 3 and less than three gestures per set. All the 3 gestures within a set are superimposed over each other to gives a **master gesture** representing that set. Then the input is matched with 9 master gestures of all sets. The closest match with a master gesture provides a security that the gesture lies in that particular set. Best case will have 2 matches for output and in worst case there will be 11 to 12 comparisons for 26 alphabets, average case will be 6-7 comparisons.

3.4 WINNING NEURON

After the selection of group to which the input lies comes the choice of winning neuron. In winning neuron technique there come only one output which is most similar in nature using kohonen algorithm. The output can be in audio and textual manner whatever desired.

3.5 ARCHITECTURAL DESIGN

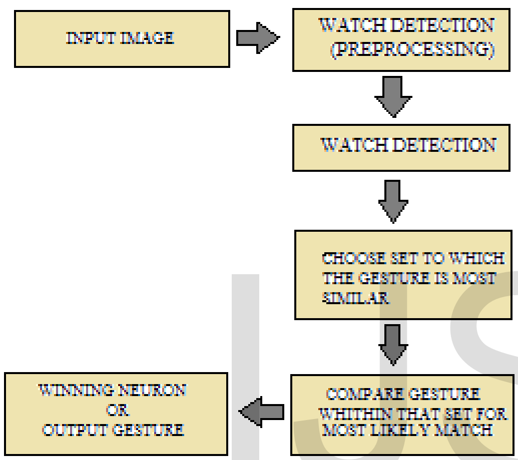

Fig. 3

4. CONCLUSION

This paper is an attempt of minimizing computational complexity and to overcome various limitations of previously used techniques like gloves, multicolored gloves, and sensor gloves. Methods summarized in this paper can be very helpful in near future in human-machine interface. Using Tree technique will reduce time complexity as it reduces number of comparisons easily.

5. FUTURE SCOPE

Methodology proposed in this paper will be implemented in our next research paper. Vision based control and access can lead us to overcome several basic limitations of our computers like mouse & keyboard. It will minimize the use of remote control in televisions, air-conditioners. Even in medical fields it can bring very drastic changes. Doctors can treat their patients using robotic arm imitating them without being physically present. It can be helpful in driving complicated vehicles just by signs or gestures. Using general things and commonsense leads to growth in practical implementations.

REFERENCES

[1] Iwan Njoto Sandjaja & Nelson Marcos (2009) "Sign Language Number Recognition**",** 2009 Fifth International Joint Conference on INC, IMS and IDC.

[2] B.Lekhashri & A.Arun Pratap(2011), "Use of Motion-Print in Sign Language Recognition", Proceedings of the National Conference on Innovations in Emerging Technology-2011,Kongu Engineering College, Perundurai, Erode, Tamilnadu, India.17 & 18 February, 2011.pp.99-102.

[3] Ray Lockton & Balliol College Oxford University "Hand Gesture Recognition Using Computer Vision",

[4] Syed Atif Mehdi & Yasir Niaz Khan, "SIGN LANGUAGE RECOGNITION USING SENSOR GLOVES", Proceedings of the 9th International Conference on Neural Information Processing (ICONIP'02), Vol. *5*