IJCAT - International Journal of Computing and Technology
Volume 1, Issue 1, February 2014
www.IJCAT.org

Gesture Recognition of Human Behavior using Multimodal Approach

[1]Amit Pimpalkar, [2] Chaitali Nagalkar, [3] Sonal Waghmare, [4] Keshari Ingole

[1, 2, 3, 4] Gurunanak Institute of Engineering and Technology, RTM Nagpur University,
Dahegaon, Kalmeshwar Road, Nagpur, Maharashtra, 441501, India

Abstract - A gesture is a form of non verbal or non vocal communication in which visible bodily actions communicate particular messages. Gestures include movement of the face, hands or other parts of the body. Gestures allow persons to speak a variety of feelings and thoughts, from condescension and hostility to approval and affection, often together with body language in addition to words when they speak. An emotion plays a crucial role in person to person interaction. In recent years, there has been a growing interest in improving all aspects of interaction between humans and computers. Human expression is used to interact with the computers. This paper explores a ways of human-computer interaction that enable the computer to be more aware of the user's emotional expressions we present an approach for the emotion recognition from a facial expression, hand and body posture. Our model uses multimodal emotion recognition system in which we use two different models for facial expression recognition and for hand gesture recognition and then combining the result of both classifiers using a third classifier which give the resulting emotion.

Keywords **- Multimodal approach, Face Detection, Facial expression recognition system, Hand gesture recognition system, PCA, Cam Shift Algorithm.**

1. Introduction

The human computer interaction is done through graphical user interface. Now days, research is going in the field of human computer interaction by means of human gesture. The human computer interaction is done by observing the human expression and gesture. Here we present a multimodal approach in which we use two different models one for recognizing the emotion using facial expression and second for hand gesture. First the image is taken through the web cam then the emotion of that person is recognized.

Gesture recognition is a topic in computer science and language technology with the goal of interpreting human gestures via mathematical algorithms. Gestures can originate from any bodily motion or state but commonly originate from the face or hand. Current focuses in the field include Emotion recognition from the face and hand gesture recognition. Many approaches have been made using cameras and computer vision algorithms to interpret sign language. However, the identification and recognition of posture, gait and human behavior is also the subject of gesture recognition techniques.

Gesture recognition can be seen as a way for computers to begin to understand human body language, thus building a richer bridge between machines and humans than primitive text user interfaces or even GUIs (graphical user interfaces), which still limit the majority of input to keyboard and mouse. Gesture recognition enables humans to interface with the machine and interact naturally without any mechanical devices. Using the concept of gesture recognition, it is possible to point a finger at the computer screen so that the cursor will move accordingly. This could potentially make conventional input devices such as mouse, keyboards and even touch-screens redundant. Gesture recognition can be conducted with techniques from computer vision and image processing.

Recently, researchers have also turned to emotional body language, i.e. the expression of emotions through human body poses and/or body motion. An implicit assumption common to the work on emotional body language is that body language is only a different means of expressing the same set of basic emotions as facial expressions. Using a set of emotional body language stimuli, which was originally prepared for neuro scientific studies, we show that human observers, as expected, perform very well on this task, and construct a model of the underlying processing stream. The model is then tested on the same stimulus set. The data we use for our work is should based on the database which was originally created FABO[1] bimodal database consisting of combined face and body expressions recorded simultaneously. Here segmentation process is applied based on a background subtraction

method on image in order to obtain the silhouette of the upper body. We extract the face and the hands automatically from image, by exploiting skin color information. The hand position consists of the position of the centroid and in-plane rotation. We employ the Camshaft algorithm for tracking the hands and predicting their locations in image. Orientation feature helps to discriminate between different poses of the hand together with the edge density information. These body features we give to the classifier as input to get the emotion.

2. Related Work

2.1 Facial Expression Recognition System (FERS)

Here we consider emotions like: Anger, Despair, Interest, Pleasure, Sadness, Irritation, Joy, Pride, etc.

Figure 1. Facial Expressions

We applied Principle Component Analysis (PCA) algorithm is to find out the face from given image. It identifies all image regions which contain a face regardless of its three-dimensional position, orientation, and lighting conditions. Such a problem is challenging because faces are no rigid and have a high degree of variability in size, shape, color, and texture. In FERS, we use Principle Component Analysis Algorithm.

2.2 Hand Gesture Recognition System (HGRS)

Hand gesture is the most easy and natural way of communication. Real-time vision-based hand gesture recognition is considered to be more and more feasible for Human-Computer Interaction. Hand tracking and segmentation are the primary steps for any hand gesture recognition system

Figure 2.Hand Movements

3. Motivation

The latest computer vision technologies and the advanced computer hardware capacity make real-time, accurate and robust hand tracking and gesture recognition promising. Many diverse approaches have been proposed such as appearance-based approaches and hand model-based approaches. The majorities of these approaches deals the hand gesture as a whole object and try to extract the corresponding mathematical description from a huge number of training samples. These approaches investigate hand gestures without breaking them into their component atomic elements that could simplify the complexity of hand gestures. As a result, many current approaches are still limited by the lack of speed, accuracy and robustness. They are either too fragile or demand too many prerequisites such as markers, clean backgrounds or complex camera calibration steps, and thus make the gesture interaction indirect and unnatural. At present there is no real-time vision-based hand tracking and gesture recognition system that can track and identify hand gestures in a fast, accurate, robust and easily accessible manner.

4. Proposed System

We propose a multimodal approach in which we use two different models one for recognizing facial expression and second for hand gesture.

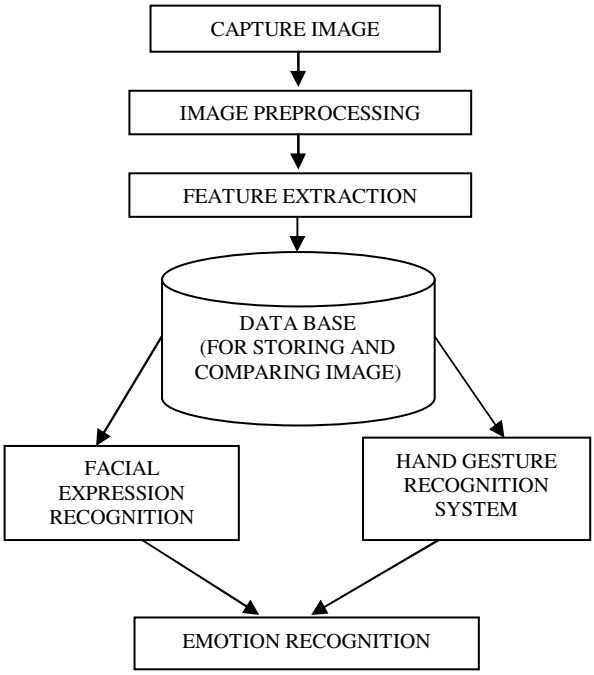


Figure 3.System Architecture

Multimodal system gives more accurate result than a signal or bimodal system. Our software is to help the disabled persons to express themselves, which is also very less in number. The new system need to be developed in efficient automotive system for proper functioning by which drawback of the previous system is removed. The target system provides a mechanism to understand sign language. For motion tracking and recognition of facial and hand gesture we use template matching method.

A. **Image Capturing:** In our system, the task of this phase is to capture an image using webcam, which is then processed in the next phases.
B. **Image preprocessing:** In the image preprocessing the skin color detection is done. The body part is separated from the image. The approach is used a previously acquired image of the background, subtracting it from the image with the gesture. Based on perimeter lengths, the hand region can then be extracted. Then from the database the image will go into the face recognition system particular emotion.
C. **Feature extraction:** The aim of this phase is to find and extract features that can be used to determine the meaning of a given gesture. Ideally a feature, or a set of features, should uniquely describe the gesture in order to achieve a reliable recognition. We extract the features by using Principle Component Analysis (PCA)[2] and Cam Shift Algorithm[5].

4.1 Principle Component Analysis

The Principal Component Analysis (PCA) is one of the most successful techniques that have been used in image recognition and compression. The jobs which PCA can do are prediction, redundancy removal, feature extraction, data compression, etc. Because PCA is a classical technique which can do something in the linear domain, applications having linear models are suitable, such as signal processing etc.

Face recognition is one of the most relevant applications of image analysis. It's a true challenge to build an automated system which equals human ability to recognize faces. One of the most used and cited statistical method is the Principal Component Analysis (PCA). It is a mathematical procedure that performs a dimensionality reduction by extracting the principal components of the multi-dimensional data. First principal component is the linear combination of the original dimensions that has the highest variability. However, high computational cost and dimensionality is a major problem of this technique. There is evidence that PCA can outperform over many other techniques when the size of the database is small.

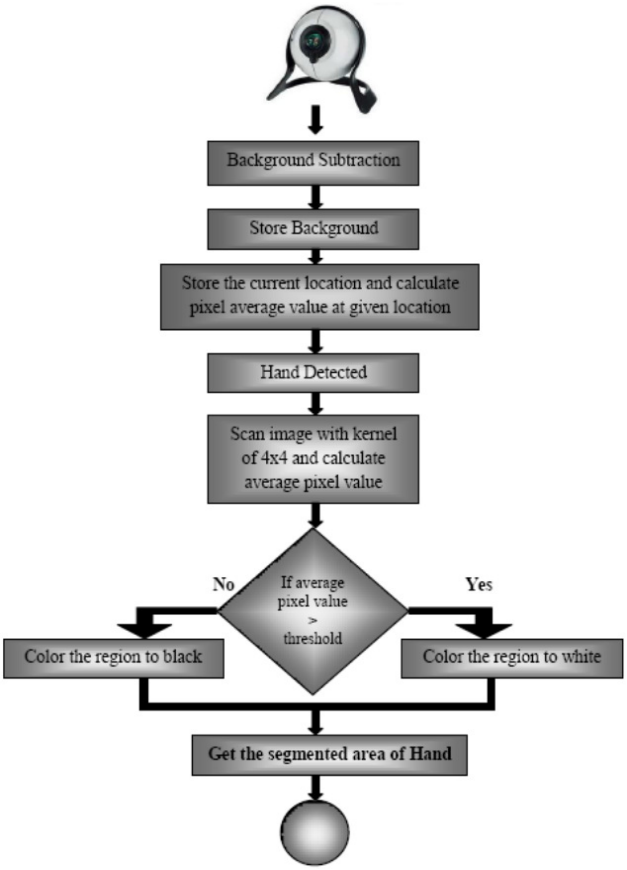


Figure 4 Architecture Design for Detection

4.2 Cam Shift Algorithm

The CAMSHIFT algorithm is based on the MEAN the search SHIFT algorithm. The MEAN SHIFT algorithm works well on static probability distributions but not on dynamic ones as in a movie. CAMSHIFT is based principles of the MEAN SHIFT but also a facet to account for these dynamically changing distributions. CAMSHIFT's is able to handle dynamic distributions by readjusting window size for the next frame based on the zeroth moment of the current frames distribution. This allows the algorithm to anticipate object movement to quickly track the object in the next scene. Even during quick movements of an object, CAMSHIFT is still able to correctly track.

In the CAMSHIFT Algorithm, a probability distribution image of the desired color in the video sequence is created. It first creates a model of the desired image that corresponds to projecting standard RGB color space along its principal diagonal from white to black. Color distributions derived from video image sequences change over time, so the mean shift algorithm has to be modified

to adapt dynamically to the probability distribution it is tracking.

The CAMSHIFT algorithm can be summarize with these steps:

1. Choose the initial region of interest, which contains the object we want to track.
2. Make a color histogram of that region as the object model.
3. Make a probability distribution of the frame using the color histogram. As a remark, in the implementation, they use the histogram back projection method.
4. Based on the probability distribution image, find the center mass of the search window using mean-shift method.
5. Center the search window to the point taken from step 4 and iterate step 4 until convergence.
6. Process the next frame with the search window position from the step 5.

D. Facial Expression Recognition System: Facial expression analysis involves the recognition and interpretation of movements in the faces of humans. Facial expression recognition is concerned with the recognition of certain facial movements without attempts to determine or presume about the underlying emotional state of the agent. The justification for this comes from the relationship between facial expression and underlying emotional states. Literature includes ongoing work in the computer vision field on capturing gestures or more general human pose and movements by cameras connected to a computer. Gestures are expressive, meaningful body motions involving physical movements of the fingers, hands, arms, head, face, or body with the intent of:

1. Conveying meaningful information
2. Interacting with the environment.

Initially Principle Component Analysis algorithm is to detect the face from given image. Face detection is to identify all image regions which contain a face regardless of its three-dimensional position, orientation, and lighting conditions. Such a problem is challenging because faces are no rigid and have a high degree of variability in size, shape, color, and texture.

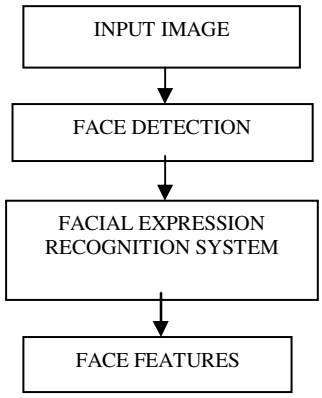


Figure 5. Facial Expression Recognition System

E. Hand Gesture Recognition System: Hand gesture is the most easy and natural way of communication. Real-time vision-based hand gesture recognition is considered to be more and more feasible for Human-Computer Interaction. Hand tracking and segmentation are the primary steps for any hand gesture recognition system. Hand tracking was done using mean shift algorithm Cam Shift function (a variation on mean shift algorithm) within the OpenCV library is used for tracking and detection. In Hand Gesture Recognition System, initially a cam shift algorithm is applied to detect the hand from given image. Then by applying cam shift algorithm our system is detect hand outline, the angle of hand, position of the hand and finally our system recognize hand gesture.

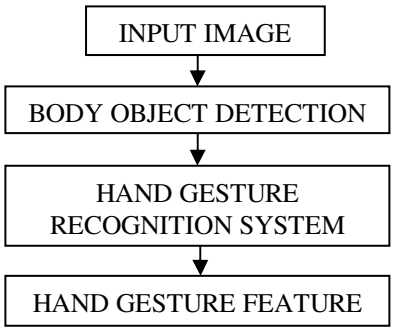


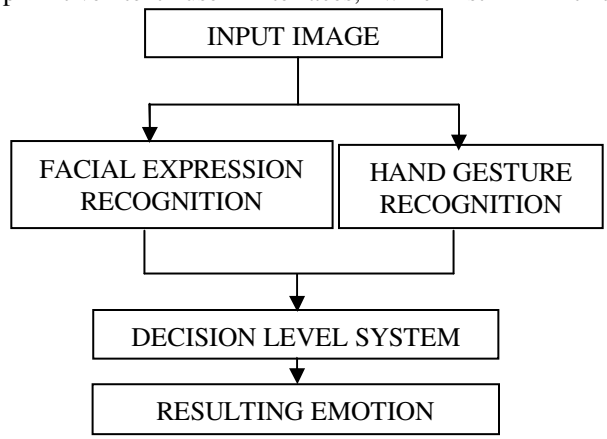
Figure 6.Hand Gesture Recognition System

F. Emotion Recognition: Our system is combine the result of facial expression recognition system and hand gesture recognition system and gives the final resulting emotion.

Figure 7.Emotion Recognition System

5. Conclusion

Gesture recognition can be seen as a way for computers to begin to understand human body language, thus building a richer bridge between machines and humans than primitive text user interfaces, which still limit the



majority of input to keyboard and mouse. We present gesture recognition system by using webcam. We proposed a multimodal approach in which we use two different models one is Facial Expression Recognition system (FERS) for recognizing facial expression and second is Hand Gesture Recognition System (HGRS) for recognize hand gesture as context. In our model we used the database which was originally created "FABO bimodal database [1]"consisting of combined face and body expressions recorded simultaneously. We used Principle Component Analysis algorithm (PCA) for recognize face and Cam Shift algorithm for tracking the hands and predicting their locations in image.

References

[1] Priya Metri, Jayshree Ghorpade and Ayesha Butalia "Facial Emotion Recognition Using Context Based Multimodal Approach", Int. J. Emerg. Sci., 2(1), 171-182, March 2012.

[2] Bajwa, I.S., Naveed, M.S., Asif, M.N., Hyder, I.S. (2008), "Feature Based Image Classification by using Principal Component Analysis", in Journal of Graphics, Vision and Image Processing, March 2008, 9(2) pp.11-17 .

[3] R. Highfield, R. Wiseman and R.Jenkins, "How your looks betray your personality", New Scientist, Feb 2012.

[4] J. Montagu, "The Expression of the Passions: The Origin and Influence of Charles Le Brun's 'Conférence sur expression générale et particulière,"Yale University Press, 2010.

[5] C. Darwin, edited by F. Darwin, "The Expression of the Emotions in Man and Animals,"2ndedition, J. Murray, London, 2009.

[6] C. Bell, "The Anatomy and Philosophy of Expression as Connected with the Fine Arts,"5thedition, Henry G. Bohn, London, 2009.

[7] A. Samal and P.A. Iyengar, "Automatic Recognition and Analysis of Human Faces and Facial Expressions: A Survey, Pattern Recognition".

[8] Paul Ekman David Mastumoto and Wallance V. Friesen University of California, "Facial Expression In Affective Disorders", New York Oxford University Press 2008.

[9] Paul Ekman and Wallance V. Friesen University of California, "Facial Signs Of Emotional Experience Journal Of Personality And Social Psychology". 2007 Vol 19, No 6, 1123 1134.

First Author has obtained his bachelor degree in Computer Engineering from Nagpur University in 2005 with first class and master degree with Honors in Computer Science & Engineering with specialization in Software Systems form RGPV, Jabalpur in 2013. He is currently working as Assistant Professor in Gurunanak Institute of Engineering & Technology, Nagpur. He published one international Journal paper and presented paper in two international conferences and five National conferences. His current research interests Data mining, Computer Graphics, Data structure, and Database.