ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

VIDEO BASED EMOTION RECOGNITION

Vijaya Lakshmi R*, Dr. S. Palanivel, Professor

* M.E II Year Computer Science & Engineering, Annamalai University, Annamalai Nagar-608 002
Computer Science & Engineering, Annamalai University, Annamalai Nagar-608 002

DOI: 10.5281/zenodo.268690

ABSTRACT

The objective of Video based Emotion Recognition System is to recognize the emotions such as normal, happy and anger from the facial expression images in the video. The face and mouth regions are detected using Viola Jones algorithm for each frame in the video. Gray Level Co-occurrence Matrix (GLCM) and Local Binary Pattern (LBP) features are extracted from the mouth region and it is used to train the Support Vector Machine. The Support Vector Machine is used to classify the emotion in the test video. Experimental results show that GLCM and LBP based emotion recognition system archives an accuracy of % and %, respectively.

KEYWORDS: Emotion Recognition, Gray Level Co-occurrence Matrix, Local Binary Pattern, Viola Jones algorithm, Support Vector Machine.

INTRODUCTION

Emotion, in everyday speech, is any relatively brief conscious experience characterized by intense mental activity and a high degree of pleasure or displeasure. Scientific discourse has drifted to other meanings and there is no consensus on a definition. Emotion is often intertwined with mood, temperament, personality, disposition, and motivation. Those acting primarily on the emotions they are feeling may seem as if they are not thinking, but mental processes are still essential, particularly in the interpretation of events. For example, the realization of our believing that we are in a dangerous situation and the subsequent arousal of our body's nervous system (rapid heartbeat and breathing, sweating, muscle tension) is integral to the experience of our feeling afraid.

Emotion is often the driving force behind motivation, positive or negative. According to other theories, emotions are not causal forces but simply syndromes of components, which might include motivation, feeling, behavior, and physiological changes, but no one of these components is the emotion.

Emotion recognition is the process of identifying human emotion, most typically from facial expressions. This is both something that humans do automatically but computational methodologies have also been developed.

Humans show universal consistency in recognizing emotions but also show a great deal of variability between individuals in their abilities. This has been a major topic of study in psychology.

It is widely accepted from psychological theory that human emotions can be classified into six archetypal emotions such as surprise, fear, disgust, anger, happiness, and sadness.

The major issues in facial expression recognition are:

- The problem that arose from this type of facial recognition system was the fact that the person to be identified must be facing the camera at no more than 35 degrees for accurate identification to be possible.
- Light differences and facial expressions contributed to low accuracy in recognition of such systems.
- The new facial recognition systems make use of three-dimensional images and thus more accurate than their predecessors.
- These systems make use of distinct features in an human face and use them as node to create a face print of a person.

- Unlike 2-D face recognition systems, however, they have the ability to recognize a face even when it is turned 90 degrees away from the camera.
- Moreover, they are not affected by the differences in lighting and facial expressions of the subjects.
- Subject does not look directly into the camera.

APPLICATIONS

A system that could enable fast and robust facial expression recognition would have many uses in both research and application areas as diverse as behavioral science, education, entertainment, medicine, and security. Following is a list of applications that can benefit from automatic recognition of facial expressions.

Banks - Suspicious Person Detection Perimeter Intrusion for Critical Infrastructures.

Stadiums - Suspicious Person Detection, Abandoned object Detection.

Airports - Suspicious Person Detection, Abandoned object Detection.

Railways/metro stations - Suspicious Person Detection, Abandoned object Detection, Parking Management, Vehicle Monitoring on roads.

Avatars with expressions-Virtual environments and characters have become tremendously popular in the 21st century. Gaming industry would benefit tremendously if the avatars were able to mimic their user's facial expressions recorded by a webcam and analyzed by a facial expression recognition system as the level of immersion and reality in the virtual world would increase. This immersion into virtual world could have many implications i.e. the game could adapt its difficulty level based on information from the facial expressions of the user.

EmotiChat -Another interesting application has been demonstrated by Anderson and McOwen, called the "EmotiChat". It consists of a chat-room application where users can log in and start chatting. The face expression recognition system is connected to this chat application and it automatically inserts emoticons based on the user's facial expressions.

Smart homes- As mentioned earlier, computing environment is moving towards human-centered designs instead of computer centered designs 8 Introduction Application areas and this paradigm shift will have far reaching consequences, one of them being smart homes. The houses could be equipped with systems that will record different readings i.e. lighting conditions, type of music playing, room temperatures etc. and associate them with the facial expressions of the inhabitants over time. Thus, such system can later control different recorded environment parameters automatically.

Affective/social robots -For social robots it is also important that they can recognize different expressions and act accordingly in order to have effective interactions. The Social Robots Project at Carnegie Mellon University states its mission as "wanting robots to behave more like people, so that people do not have to behave like robots when they interact with them". To attain such human-robot interaction, it is of paramount importance for the robot to understand the human's facial expressions.

Detection and treatment of depression and anxiety- Research based on the FACS has shown that facial expressions can predict the onset and remission of depression, schizophrenia, and other psychopathological afflictions has also been able to identify patterns of facial activity involved in alcohol intoxication that observers not trained in FACS failed to note. This suggests there are many applications for an automatic facial expression recognition system based on FACS.

Pain monitoring of patients - Pain monitoring of patients is a very complicated but very important task. Manually monitoring of pain has some problems: first, pain cannot be recorded continuously. Secondly, some patients can under report the pain while other can do just opposite. An automatic facial expression recognition system could solve above mentioned problems. It has been shown that it is possible to derive a measure of pain and to distinguish between different types of pain from a patient's facial expressions.

DATABASE COLLECTION

Indian Spontaneous Expression Database (ISED) was developed in Indian Institute of Technology (IIT) Kharagpur, India, directed by Prof. Priyadarshi Patnaik at the Department of Humanities and Social Sciences, IIT Kharagpur.

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

ISED is composed of spontaneous facial expressions through active elicitation of emotions. The details of the database is given in Table 1:

Number of video clips	428
Number of participants	50 (29 male, 21 female)
Emotions elicited	Happiness (227 clips) Surprise (73 clips) Sadness (48 clips) Disgust (80 clips)
Clip duration	1-10 sec
Clip selection	Manual
Self-report of emotion	Yes
Emotion rating scale	0-5 (0: no emotion, 5:maximum intensity)

Table 1: Database

PROPOSED SYSTEM

The proposed system consists of the following modules: Face and mouth detection, Extraction of GLCM and LBP features, training and testing using SVM. The block diagram of the proposed system is shown in Fig. 1.

Face and mouth detection:

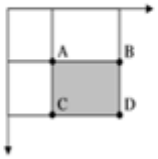
Face detection is being used in a variety of applications that identifies human faces in digital images. Face detection also refers to the psychological process by which humans locate and attend to faces in a visual scene. Face detection can be regarded as a specific case of object-class detection. In object-class detection, the task is to find the locations and sizes of all objects in an image that belong to a given class. Examples include upper torsos, pedestrians, and cars. Face-detection algorithms focus on the detection of frontal human faces. It is analogous to image detection in which the image of a person is matched bit by bit. Image matches with the image stores in database. Any facial feature changes in the database will invalidate the matching process.

The basic principle of the Viola-Jones algorithm is to scan a sub-window capable of detecting faces across a given input image. The standard image processing approach would be to rescale the input image to different sizes and then run the fixed size detector through these images. This approach turns out to be rather time consuming due to the calculation of the different size images. Contrary to the standard approach Viola-Jones rescale the detector instead of the input image and run the detector many times through the image – each time with a different size. At first one might suspect both approaches to be equally time consuming, but Viola-Jones have devised a scale invariant detector that requires the same number of calculations whatever the size. This detector is constructed using a so-called integral image and some simple rectangular features reminiscent of Haar wavelets.

The first step of the Viola-Jones face detection algorithm is to turn the input image into an integral image. This is done by making each pixel equal to the entire sum of all pixels above and to the left of the concerned pixel.

This allows for the calculation of the sum of all pixels inside any given rectangle using only four values. These values are the pixels in the integral image that coincide with the corners of the rectangle in the input image.

Sum of gray rectangle= D-(B+C)+A



Since both rectangle B and C include rectangle A the sum of A has to be added to the calculation.

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

1	1	1
1	1	1
1	1	1

Input image

1	2	3
2	4	6
3	6	9

Integral image

It has now been demonstrated how the sum of pixels within rectangles of arbitrary size can be calculated in constant time. The Viola-Jones face detector analyzes a given sub-window using features consisting of two or more rectangles. The different types of features are shown in Fig. 2.

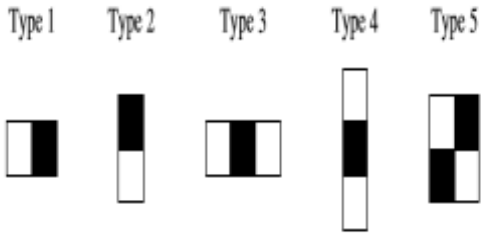


Fig.2. The different types of features.

Each feature results in a single value which is calculated by subtracting the sum of the white rectangle(s) from the sum of the black rectangle(s).

Viola-Jones have empirically found that a detector with a base resolution of 24x24 pixels gives satisfactory results. When allowing for all possible sizes and positions of the features in Fig.1 a total of approximately 160,000 different features can then be constructed. Thus, the amount of possible features vastly outnumbers the 576 pixels contained in the detector at base resolution. These features may seem overly simple to perform such an advanced task as face detection, but what the features lack in complexity they most certainly have in computational efficiency.

The results of Face and mouth detection is shown in Fig.3.

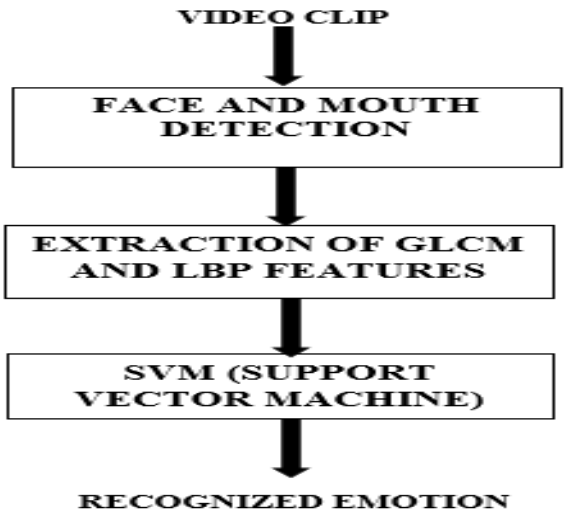


Fig.3. Block Diagram of the Proposed System.

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

Extraction of GLCM features:

In statistical texture analysis, texture features are computed from the statistical distribution of observed combinations of intensities at specified positions relative to each other in the image. According to the number of intensity points (pixels) in each combination, statistics are classified into first-order, second-order and higher-order statistics. The Gray Level Co-occurrence Matrix (GLCM) method is a way of extracting second order statistical texture features.

A GLCM is a matrix where the number of rows and columns is equal to the number of gray levels, in the image. The matrix element P(i,j | Δx,Δy) is the relative frequency with which two pixels, separated by a pixel distance (Δx,Δy), occur within a given neighborhood, one with intensity i and the other with intensity j. One may also say that the matrix element P(i,j | d,θ) contains the second order1 Consider the image Shown below. If we use the position operator "1 pixel to the right and 1 pixel down" then we get the gray-level co-occurrence matrix C

$$\begin{matrix} 0 & 0 & 0 & 1 & 2 \\ 1 & 1 & 0 & 1 & 1 \\ 2 & 2 & 1 & 0 & 0 \\ 1 & 1 & 0 & 2 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{matrix}$$

$$C=\frac{1}{16}\begin{bmatrix} 4 & 2 & 1 \\ 2 & 3 & 2 \\ 0 & 2 & 0 \end{bmatrix}$$

where an entry C_{ij} is a count of the number of times that F(x,y) = i and F(x + 1,y + 1) = j. For example, the first entry comes from the fact that 4 times a 0 appears below and to the right of another 0. The factor 1/16 is because there are 16 pairs entering into this matrix, so this normalizes the matrix entries to be estimates of the co-occurrence probabilities.

For statistical confidence in the estimation of the joint probability distribution, the matrix must contain a reasonably large average occupancy level. Achieved either by (a) restricting the number of amplitude quantization levels (causes loss of accuracy for low-amplitude texture), or (b) using large measurement window. (causes errors if texture changes over the large window). Typical compromise: 16 gray levels and window size of 30 or 50 pixels on each side. Now we can analyze C: maximum probability entry, element difference moment of order k: $\sum_i \sum_j (i-j)^k c_{ij}$.

This descriptor has relatively low values when the high values of C are near the main diagonal. For this position operator, high values near the main diagonal would indicate that bands of constant intensity running "1 pixel to the right and 1 down" are likely. When k = 2, it is called the contrast:

Contrast = $\sum_i \sum_j (i-j)^2 c_{ij}$

Entropy = $\sum_i \sum_j c_{ij} \log c_{ij}$

This is a measure of randomness, having its highest value when the elements of C are all equal. In the case of a checkerboard, the entropy would be low.

$$\begin{matrix} 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \end{matrix} \quad \Rightarrow \quad \begin{matrix} 8 & 0 \\ 0 & 8 \end{matrix}$$

Uniformity (also called Energy) = $\sum_i \sum_j c_{ij}^2$ (smallest value when all entries are equal).

Homogeneity = $\sum_i \sum_j \frac{c_{ij}}{1+|i-j|}$ (large if big values are on the main diagonal).

In GLCM feature extraction, we have totally extracted 13 features Autocorrelation, Contrast, Correlation, Cluster Prominence, Cluster Shade, Dissimilarity, Energy, Entropy, Homogeneity, Maximum probability, Sum of Squares: Variance, Sum average, Sum variance, Sum entropy, Difference variance, Difference entropy ,

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

Information measure of correlation, Inverse difference (INV), Inverse difference normalized (INN) ,Inverse difference moment normalized.

Problems associated with the co-occurrence matrix methods:
1. They require a lot of computation.
2. Features are not invariant to rotation or scale changes in the texture.

Fig.3. Face and Mouth Detection

Extraction of LBP features:
Local binary patterns (LBP) is a type of visual descriptor used for classification in computer vision.

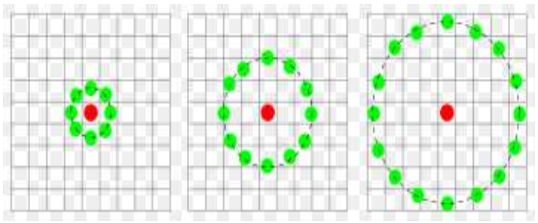


Fig.4. Extraction of Local Binary Pattern

The LBP feature vector, in its simplest form, is created in the following manner:

Divide the examined window into cells (e.g. 16x16 pixels for each cell). For each pixel in a cell, compare the pixel to each of its 8 neighbors (on its left-top, left-middle, left-bottom, right-top, etc.). Follow the pixels along a circle, i.e. clockwise or counter-clockwise. Where the center pixel's value is greater than the neighbor's value, write "0". Otherwise, write "1". This gives an 8-digit binary number (which is usually converted to decimal for convenience).

Compute the histogram, over the cell, of the frequency of each "number" occurring (i.e., each combination of which pixels are smaller and which are greater than the center). This histogram can be seen as a 256-dimensional feature vector. Optionally normalize the histogram. Concatenate (normalized) histograms of all cells. This gives a feature vector for the entire window.

The feature vector can now be processed using the Support vector machine or some other machine-learning algorithm to classify images. Such classifiers can be used for face recognition or texture analysis. A useful extension to the original operator is the so-called uniform pattern, which can be used to reduce the length of the feature vector and implement a simple rotation invariant descriptor. This idea is motivated by the fact that some binary patterns occur more commonly in texture images than others. A local binary pattern is called uniform if the binary pattern contains at most two 0-1 or 1-0 transitions. For example, 00010000 (2 transitions) is a uniform pattern, 01010100 (6 transitions) is not. In the computation of the LBP histogram, the histogram has a separate bin for every uniform pattern, and all non-uniform patterns are assigned to a single bin. Using uniform patterns, the length of the feature vector for a single cell reduces from 256 to 59.

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

Training and testing using SVM:

In machine learning, support vector machines (SVM) are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on.

In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces.

When data are not labeled, supervised learning is not possible, and an unsupervised learning approach is required, which attempts to find natural clustering of the data to groups, and then map new data to these formed groups. The clustering algorithm which provides an improvement to the support vector machines is called support vector clustering and is often used in industrial applications either when data is not labeled or when only some data is labeled as a preprocessing for a classification pass.

More formally, a support vector machine constructs a hyperplane or set of hyperplanes in a high- or infinite-dimensional space, which can be used for classification, regression, or other tasks. Intuitively, a good separation is achieved by the hyperplane that has the largest distance to the nearest training-data point of any class (so-called functional margin), since in general the larger the margin the lower the generalization error of the classifier.

Whereas the original problem may be stated in a finite dimensional space, it often happens that the sets to discriminate are not linearly separable in that space. For this reason, it was proposed that the original finite-dimensional space be mapped into a much higher-dimensional space, presumably making the separation easier in that space. To keep the computational load reasonable, the mappings used by SVM schemes are designed to ensure that dot products may be computed easily in terms of the variables in the original space, by defining them in terms of a kernel function selected to suit the problem.

EXPERIMENTAL RESULTS

The proposed method is evaluated using the database shown in Table.1. Face and mouth regions are detected for all the frames in each video clip. 13 dimensional GLCM and 59 dimensional LBP features are extracted as described in section III. 300 video clips are used for training and 100 video clips are used for testing. The features extracted from training video clips are used to create an SVM for each emotion. The features extracted from test. Video clips are used for testing the SVM. Emotion recognition system achieving an accuracy of about 45.0 % , 73.0 % using GLCM and LBP features, respectively.

CONCLUSION

In this work, a method was proposed which identify person facial emotions in a video. Face and mouth regions are detected by using Viola Jones Method. The Gray Level Co-occurrence Matrix and Local Binary Patterns are extracted from the mouth region. Support Vector Machine was used for recognizing the emotions. GLCM and LBP based emotion recognition system achieves an accuracy of about 45.0 % and 73.0 %, respectively.

REFERENCES

[1] P. Ekman, Emotions Revealed: Recognizing Faces and Feelings to Improve Communication and Emotional Life, New York: Times Books, 2003.
[2] A. Kleinsmith and N. Bianchi-Berthouze, "Affective body expression perception and recognition: A survey," IEEE Transactions on Affective Computing, vol. 4, no. 1, pp. 15-33, 2013.
[3] Y. Tian, T. Kanade and J. F. Cohn, "Facial expression recognition," in Handbook of face recognition, London, Springer , 2011, pp. 487-519.
[4] E. G. Krumhuber, A. Kappas and A. S. Manstead, "Effects of dynamic aspects of facial expressions: A review," Emotion Review, vol. 5, no. 1, pp. 41-46, 2013.
[5] S. Wang, Z. Liu, S. Lv, Y. Lv, G. Wu, P. Peng, F. Chen and X. Wang, "A natural visible and infrared facial expression database for expression recognition and emotion inference," IEEE Trans. Multimedia, vol. 12, no. 7, pp. 682-691, 2010.

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

[6] P. Ekman, "Universals and Cultural Differences in Facial Expressions of Emotion," in Proc. Nebraska Symp. Motivation, 1971.
[7] P. Ekman, "Strong Evidence for Universals in Facial Expressions:A Reply to Russell's Mistaken Critique," Psychological Bull., vol. 115, no. 2, pp. 268-287, 1994.
[8] Z. Zeng, M. Pantic, G. I. Roisman and T. S. Huang, "A survey of affect recognition methods: Audio, visual, and spontaneous expressions," IEEE Trans. Pattern Anal. Mach. Intell. , vol. 31, no. 1, pp. 39-58, 2009.
[9] A. Dhall, R. Goecke, S. Lucey and T. Gedeon, "Collecting large, richly annotated facial-expression databases from movies," IEEE MultiMedia, vol. 19, no. 3, pp. 34-41, 2012.
[10] K. Mase, "Recognition of Facial Expression from Optical Flow," IEICE Trans., vol. 74, no. 10, pp. 3474-3483, 1991.
[11] H. Kobayashi and F. Hara, "The Recognition of Basic Facial Expressions by Neural Network," in Proc. IEEE Int Joint Conf. Neural Networks, 1991.
[12] M. J. Lyons, M. Kamachi and J. Gyoba, "Japanese Female Facial Expressions (JAFFE)," Database of digital images, 1997.
[13] T. Kanade, J. F. Cohn and Y. Tian, "Comprehensive database for facial expression analysis," in 4th IEEE Int. Conf. on Automatic Face and Gesture Recognition, 2000.
[14] M. Kamachi, V. Bruce, S. Mukaida, J. Gyoba, S. Yoshikawa and S. Akamatsu, "Dynamic properties influence the perception of facial expressions," Perception, vol. 30, pp. 875-887, 2001.
[15] S. Afzal and P. Robinson, "Natural affect data-collection & annotation in a learning context," in 3rd Int. Conf. on Affective Comput. and Intell. Interaction and Workshops, 2009.