Movie Rating Predictions Based on Audience Face Expressions using KNN

Shritika Wayker[1], Mrs. V. L. Kolhe[2]

[1](Shritika Wayker, Department of Computer Engineering,)
[2](Mrs. V. L. Kolhe, Department of Computer Engineering,)
(D.Y. Patil college of Engineering, Akurdi, Pune, Maharashtra, India)

Abstract – **Facial emotion recognition is one of the specific issues of computer vision. Emotions which can be classified like fear, happiness, joy, sadness, aggressiveness are recognizable facial expressions using computer vision. We have implemented a systemwhich can perceive four facial expressions are confused, happy, and sad and normal for prediction of movie rating. At the first step of the system, human faces are detected and located in an image. In the second step, eyes and mouth are detected in facial areas.The emotion feature database is used for face and emotion recognition of the person. For detecting face from the input image we are using Viola-Jones face detection algorithm and to evaluate the face and emotion detection KNN classifier is used.**

Keywords: Facial emotion recognition, Viola-Jones face detection algorithm, KNN classifier, Preprocessing, Feature Extraction, Prediction of Movie Rating.

I. INTRODUCTION

Among the emotional factor, facial expression is one of the most important and spontaneous element to recognize the human emotional status that gives large amounts of information about person's feelings.The automated analysis of emotions allows an approximate idea of the feelings that a person is experiencing.

While watching movies, audience members exhibit both subtle and coarse gestures like smiles, head-pose change, toying, stretching which convey sentiment i.e., engaged or disengaged during feature length movies. Noticing these behaviors using computer vision systems is a very challenging problem especially in a movie theatre environment. The environment is dark and contains views of people at different scales and viewpoints. Moreover, the physical configuration makes it difficult to observe facial expressions on all audience members. Manually identifying these events is prohibitive given the size of the dataset. As audience members are often stationary for long periods of time, the annotation of sentiment levels can be expedited by exploiting the significant redundancy in the input signal.

The proposed framework can automatically predict audience sentiment levels using K-Nearest Neighbor (KNN). Using the audience engagement levels, we predict movie ratings and demonstrate its predictive capability.

II. LITERATURE REVIEW

Victor M. Álvarez et al. [1] describes a method for facial expression detection based on a comparison between interest points of facial images showing an emotion and facial images showing a neutral expression. The resulting gradients were classified and visualized in similarity plots that will be latter processed by a classifier. The system can detect the real-time facial emotion recognition system of simple structure demanding little computational burden.

Yang Zhong et al. [2] proposes the approach of predicting face attributes using KNNs trained for face recognition. Combining with conventional face localization techniques the KNNs is get with off-the-shelf architectures and publicly available models like Google's FaceNet with the conventional pipeline to study the prediction power of different representations from the trained KNNs. Here the face descriptors are constructed from different levels of the KNNs for different attributes to best facilitate face attribute prediction. By properly leveraging these off-the-shelf KNN representations, the system can achieve accurate attribute prediction with current state-of-the-art performance using the two datasets LFWA and CelebA.

Ramón Zatarain-Cabada et al. [3] presents the building and validating of a face expression database and a face expression recognizer. The face expression recognizer uses a geometric-based technique that measures distances between the central point on the face and other 68 facial landmark points. These measures are transformed into features to train a support vector machine. The database was built inside an educational context while students' program in Java code. The tests validate the accuracy of the recognizer applying a ten-fold cross-validation.

Barbu, et al. [4] is considered that Gabor filtering is one of the most vital feature extraction systems in facial expression recognition. The basic limitation of Gabor filter is its bandwidth limitation i.e. Supreme bandwidth is limited to one octave. Gabor filters cost high and low frequency information since it is band pass in nature.

Ashwin T et al. [5] propose multiuser face detection-based eLearning system using support vector machine based supervised machine learning technique. Experimental results demonstrate that the proposed system provides the accuracy of 89% to 100% w.r.t different datasets (LFW, FDDB, and YFD).

Chen Cao, et al. [6] demonstrate the potential of Face Warehouse a 3D facial expression database for visual computing with four applications i.e. facial image manipulation, face component transfer, real-time performance-based facial image animation, and facial animation retargeting from video to image.

Hari Prasad et a. [7] describes facial expression recognition. When pass a surveillance video through facial expression detector model, get emotional sequences of different images. This sequence is called as single Emotional Avatar Image representation. A large rigid head motion is compensated by adopting SIFT flow technique which aligns the face images. SIFT flow technique generates the higher classification rate.

Chen S., Sun Y. and Yin B. [8] describes the recognition proportion of PCA is low and has small model scope problem. On behalf of gray facial expression detection, 2DPCA is extended to Extended 2DPCA. The drawback of this method is that it is not suitable for hue images. Consequently Sub pattern extended 2-Dimensional PCA (SpE2DPCA) is introduced for Color face recognition. PCA has the problem in PCA of small sample size is also excluded. This method has the higher recognition rate than E2DPCA, 2DPCA.

M. A. O. Vasilescu and D. Terzopoulos. [11] describes the 1-D Linear Discriminant Analysis AND 2-D Linear Discriminant Analysis are extended in Color space to improve the face recognition accuracy. A 3-D Color tensor is generated using Color Linear Discriminant Analysis (LDA) subspace . The horizontal unfolding used in this method increases the rate of recognition for 2-D LDA whereas vertical unfolding improves on the recognition proportion for 2-D Principal Component Analysis (PCA).

M. A. O. Vasilescu and D. Terzopoulos. [10] describes facial expression recognition needs different attributes such as expressions, pose, and lighting, to be taken care of. The conventional PCA addresses only difference in single factor. Multilinear image analysis is used multilinear algebra. In this, the thought of Tensor faces is functional, which isolates the several features used in the making and representation of an image. Multilinear image analysis does not incorporate the Color information. Multilinear Image Analysis method provides recognition rate higher than the PCA approach.

Songfan Yang and Bir Bhanu. [13] describes appearance based features are being used for face recognition since it encodes specific details about human faces. The facial image is separated into sub blocks and its similarities in the sub blocks is acquired.

S. M. Lajevardi and H. R. Wu. [9] describes LBP is obtained from Gabor filters for feature vector generation in Local Gabor Binary Pattern (LGBP) method. All techniques have enhanced the performance of the facial expression recognition system. The Color spaces like RGB do not provide provision for head pose and lighting variations. A significant benefit of Local Gabor Binary Pattern (LGBP) is its illumination tolerance. Local

Barbu, T. [12] describes Gabor Binary Pattern (LGBP) achieves better performance than Gabor filter technique. The Gabor filtering is considered as one of the most vital feature extraction system in facial expression recognition. The basic limitation of Gabor filter is its bandwidth limitation i.e. Supreme bandwidth is limited to one octave. Gabor filters cost high and low frequency information since it is band pass in nature [9]. Gabor filter bank performs better in terms of recognition rate than the other approaches like principal component analysis, linear discriminant analysis etc.

III.PROPOSED SYSTEM

The Aim of this project is to develop an application that can be used for monitor viewer's engagement during watching movie or video. Application has features like capturing viewer's facial expressions and based on analysis providing the graphical representation of viewer's emotion that is wholly contributing in generating movie rating which is used for maintaining quality movie or video. Figure 1 shows the structural design of the proposed system

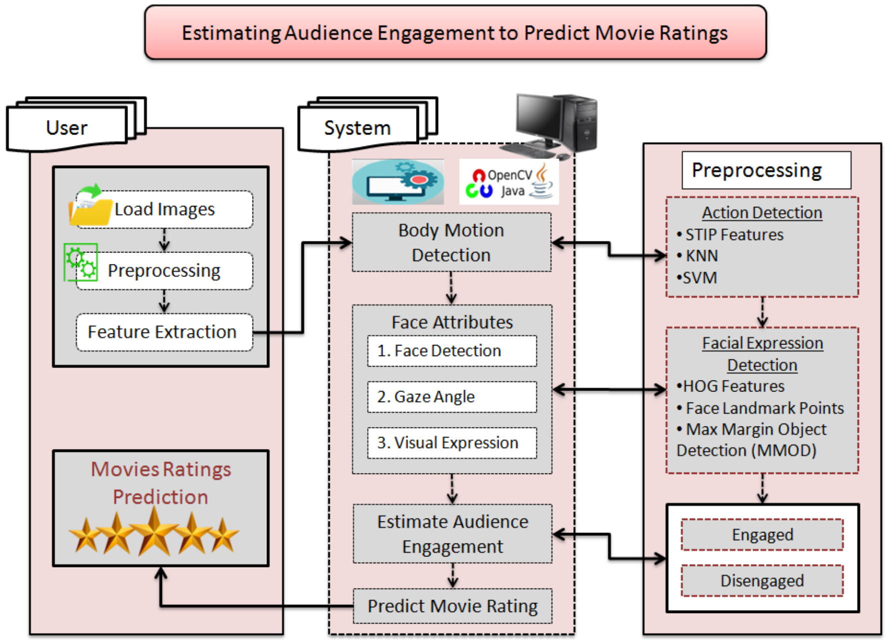

Figure 1: System Architecture

1. **Frame Extraction / Live Camera:** User uploads a video / grabs images using live camera on the application, application then extracts frames from the video. These frames are saved on local machine. Frames are usually in 640x480 formats.
2. **Pre-Processing on Images:** Once we get the faces apply the pre processing on images like noise removal, normalization etc.
   a. RGB to Gray Scale Image: Convert the image into Gray scale by taking the average of the each pixel RGB.
   b. Image Normalization: Normalization is a process that changes the range of pixel intensity values to avoid mental distraction or fatigue from the images.
   c. Noise Removal: Removing errors in the image acquisition process that result in pixel values that do not reflect the true intensities of the real scene.
3. **Body Motion Detection :** In this phase we detect the person body motion using STIP features and KNN algorithm. Here feature is extract with the help of STIP.
4. **STIP:** It is based on the detection of spatio-temporal corners. Spatio-temporal corners are located in region that exhibits a high variation of image intensity in all three directions (x, y, t). This requires that spatio-temporal corners are located at spatial corners such that they invert motion in two consecutive frames (high temporal gradient variation). They are identified from local maxima of a cornerness function computed for all pixels across spatial and temporal scales.
5. **Face Attributes :** For estimating the engagement level of the person, faces provide the useful information such as gaze angle & expression.

i. **Face Detection :** Two main techniques are used for face detection. Max-Margin Object Detection (MMOD) optimize over all sub windows to detect objects in image. Implements for train models and detect faces. HOG Features + SVM algorithm train all sub windows in every training images. Training done on 800 images.

ii. **Gaze Angle:** Detect head position (i.e. frontal/near frontal, looking away from the screen and looking down) of the audience.

iii. **Visual Expression:** In this phase audience facial motion or expression are detected. For detection facial expression HOG descriptor is used. Our visual expression feature consists of representing the input face image via HOG descriptor using 9 orientation bins with overlapping regions with block size of 2*2, and cell size of 8*8.

Feature	Description
$\mathbf{f}_{angle}$	Visual focus for movie screen
$\mathbf{f}_{hog}$	Facial expressions
$\mathbf{x}_{face}$	Face features: visual focus + facial expressions
$\mathbf{b}_{w}$	Body motion feature
$\mathbf{x}_{facebody}$	Combination of face and body features

Table 1: Feature Description

6. **Estimating Audience Engagements:** Visual appearance helps to estimating audience engagement during various segments of the movie. Audience engagement estimate using following two parameters i.e. Engaged, Disengaged.
7. **Predicting Movie Ratings:** In this phase movie rating predict according to engagement analysis results.

IV. ALGORITHM USED

A. Support Vector Classification Algorithm :

- Support vector machine (SVM) proposed by vapnik and cortes have been successfully applied for gender classification problems by many researchers. An SVM classifier is a linear classifier where the separating hyper plane is chosen to minimize the expected classification error of the unseen test patterns.
- SVM is a strong classifier which can identify two classes. SVM classifies the test image to the class which has the maximum distance to the closest point in the training.
- SVM training algorithm built a model that predict whether the test image fall into this class or another.SVM require a huge amount of training data to select an affective decision boundary and computational cost is very high even if we restrict ourselves to single pose (frontal) detection.
- The SVM is a learning algorithm for classification. It tries to find the optimal separating hyper plane such that the expected classification error for unseen patterns is minimized.
- For linearly non-separable data the input is mapped to high-dimensional feature space where they can be separated by a hyper plane. This projection into high-dimensional feature space is efficiently performed by using kernels. More precisely, given a set of training samples and the corresponding decision values {-1, 1} the SVM aims to find the best separating hyper plane given by the equation WT x+b that maximizes the distance between the two classes.

B. KNN (K Nearest Neighbors) Algorithm:

K nearest neighbors is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure (e.g., distance functions).

- A case is classified by a majority vote of its neighbors, with the case being assigned to the class most common amongst its K nearest neighbors measured by a distance function. If K = 1, then the case is simply assigned to the class of its nearest neighbor.
- Let (Xi, Ci) where i = 1, 2., n be data points. Xi denotes feature values & Ci denotes labels for Xi for each i
- Assuming the number of classes as c.
- Ci 1, 2, 3, , c for all values of i. Let x be a point for which label is not known, and we would like to find the label class using k-nearest neighbor algorithms.

Step 1:- Calculate d(x, xi) i =1, 2, .., n; where d denotes the Euclidean distance between the points.
Step 2:- Arrange the calculated n Euclidean distances in non-decreasing order.
Step 3:- Let k be a +ve integer, take the first k distances from this sorted list.
Step 4:- Find those k-points corresponding to these k-distances.
Step 5:- Let ki denotes the number of points belonging to the i th class among k points i.e. k 0

V.EXPERIMENTAL RESULTS

The images are captured and tested under the various contrast values to get the accuracy. System tested on 100 images and also gives the time required for detection of face, eyes and mouth for different method.
The accuracy of the proposed system is calculated by using True Negative (TN), True Positive (TP), False Positive(FP), False Negative(FN) values. We can classify the emotions in class like engage, disengage, happy, surprise, sad.

TABLE I:TP, FP,TN and FN

Classes	TP	TN	FP	FN
Engage	61	315	39	30
Disengage	70	306	30	39
Happy	85	291	15	20
Surprise	82	294	18	15
Sad	78	298	22	20

Here we calculate the accuracy of the proposed system with the TP,TN,FP and FN using two algorithm SVM and KNN. First, in the face detection SVM having less accuracy than KNN, because it is considering other parts as a face except face.

$$Accuracy = \frac{No.of\ Images\ Detected}{Total\ \ No.of\ Images} * 100$$

Second, in eye detection only eye get detected so points get calculated on one part where SVM having better accuracy than KNN. Third, in mouth detection both algorithms giving similar accuracy result. So, comparison is based on above equation. Below table give the total time required for each feature to get the accuracy for 100 images using above formula

TABLE II: Algorithm Comparison

Algorithm Comparison			
Features	**Algorithms**	**Time Required (ms)**	**Accuracy**
Face Detection	SVM	130	61
	KNN	125	75
Eye Detection	SVM	22	80
	KNN	18	62
Mouth Detection	SVM	19	72
	KNN	15	75

The graphical representation of the above table is shown in below figure.

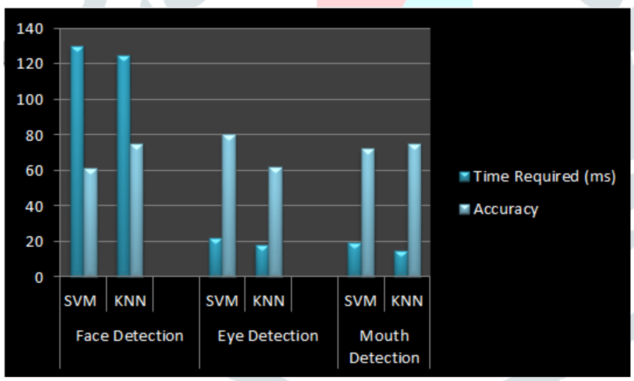

Figure 2: Algorithm Comparison

VI. CONCLUSION

An approach to estimating audience engagement for movie rating has been introduced using a SVM classification algorithm. The audience engagement is based on the audience facial expression and body movement. And the movie rating is predicted based on the audience engagement level. An efficient method to create face and emotion feature database and then this will be used for face and emotion recognition of the person. To evaluate the face and emotion detection KNN classifier is used. Accuracy is increased using SVM algorithm.

REFERENCES

[1] Victor M. Álvarez et al. "A Method for Facial Emotion Recognition Based on Interest Points ",©2018 IEEE .

[2] Yang Zhong, Josephine Sullivan, Haibo Li, "Face Attribute Prediction Using Off-the-Shelf KNN Features", 2016 International Conference on Biometrics (ICB), 2016.

[3] Ramon Zatarain- Cabada, Maria Lucia Barron- Estrada, Francisco Gonzalez- Hernandez, "Building a face expression recognizer and a face expression database for an intelligent tutoring system", IEEE, 17th International Conference on Advanced Learning Technologies, 2017

[4] Barbu, T, Gabor filter-based face recognition technique, Proceedings of the Romanian Academy, vol.11, no. 3, pp. 277 283, 2010.

[5] Ashwin T S, Jijo Jose, Raghu G, G Ram Mohana Reddy "An E-learning System with Multifacial Emotion Recognition Using Supervised Machine L

[6] Chen Cao, YanlinWeng, Shun Zhou, Yiying Tong, and Kun Zhou "Face Warehouse: A 3D Facial Expression Database for Visual Computing", IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, VOL. 20, NO. 3, MARCH 2014.

[7] Hari Prasad Mal and Dr. Swarnalatha P. Facial Expression Detection using Facial Expression Model. IEEE 20178.

[8] Chen S., Sun Y. and Yin B., A Novel Hybrid Approach Based on Subpattern Technique and Extended 2DPCA for Color Face Recognition, 11th IEEE International Symposium on Multimedia, pp. 630-634, 2009.

[9] S. M. Lajevardi and H. R. Wu, Facial Expression Recognition in Perceptual Color Space IEEE Transactions on image Processing, vol. 21, no. 8, pp. 3721-3732, 2012.

[10] M. Thomas, C. Kambhamettu and S. Kumar, (2008), Face recognition using a color subspace LDA approach, Proceedings International Conference on Tools with Artificial Intelligence.

[11] M. A. O. Vasilescu and D. Terzopoulos, Multilinear image analysis for facial recognition, in Proc. Int. Conf. 511514 Pattern Recognit. Quebec City, QC, Canada, Aug. 2002.

[12] Barbu, T, Gabor filter-based face recognition technique, Proceedings of the Romanian Academy, vol.11, no. 3, pp. 277 283, 2010.

[13] Songfan Yang and Bir Bhanu, Understanding Discrete Facial Expressions in Video Using an Emotion Avatar Image, IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics, Vol. 42, No. 4, pp.980-992,2012.