ISSN:2320-0790

An International Journal of Advanced Computer Technology

Combined Methodology of PHOG and LBP to Identify Facial Expression from Videos

Ashish D. Lonare[1], Shweta V. Jain[2]

[1]Department of Computer Science and Engineering, Shri Ramdeobaba College of Engineering and Management Nagpur, India

lonareashish1@gmail.com

[2]Department of Computer Science and Engineering, Shri Ramdeobaba College of Engineering and Management Nagpur, India

jains@rknec.edu

Abstract: An interesting and one of the challenging problem for this decade is of the identification and detection of human facial expression from cluttered image sequence. In this paper we proposed a methodology of PHOG and LBP to identify facial expression from GEMEP-FERA dataset 2011 competition, a standard dataset. we compare the results of PHOG and LBP results. during results analysis it was found that by combining the features of PHOG and LBP we found improved results. For classification, we used the K-NN classifier. before this, a key frame were extracted by using the novel method. PHOG gives gradient features and LBP gives binary features by combining both features, we obtain improved results.

Keywords: Pyramid Histogram of Orientation gradients, K-NN classifier, GEMEP-FERA dataset, Facial Expression, Key Frame Extraction, gradients, binary features.

I. INTRODUCTION

One of the most challenging problems over the last few decades occur in the form of identification of facial expression from images sequence. The view of the expression can be brief by considering the shape features in the form of nose, mouth region, the area surrounding the eye and the appearance features like bulges, forehead mark can give temporal details about the expression for particular person. Facial expression analysis has many real life applications like human computer interaction in the form of human vision system, lie detection, pain detection, medical appliance etc. A lot of research is going on in this area where the classification is performed using the supervised or unsupervised learning methods.

Facial expression gain important due to the advancement in the region of face detection, face tracking and face recognition methodology. Fasel and Luettin [1] suggested different approaches for facial expression analysis in the form of model based methods, local methods, motion extraction methods and some in the form of holistic methods. They also consider the difficulty faced during expression recognition like issues in the form of occlusions such as person with specs and hair around the face region, illumination such as partly lightened face where difficulty arise in processing the face region for expression classification and pose problems such as out of head movement occur where the tracking of head become difficult. Ekman and Keltner [2] developed Facial Action Coding System(FACS) which brief facial expression in terms of Action Units. Facial Action Coding System speaks in term of total 46 AUs. This AUs are related to the Facial muscle movements like forehead changes, lower jaws movement. This all movement constitute total of 46 Action Units. Wang [3] worked on the real time facial expression recognition. For this real time facial expression recognition, a technique named Adaboost had been opted for facial expression recognition in which a face had detected using haar features. After the face had detected, expression are classified for face region like Happy, sad, fear, Angry etc.

In this paper, we propose two method which works locally on face region and obtain the facial features in the form of the Features vector. This features vectors of two method are combined and the results obtain in the form of improved over single method of each method. The dataset used for experimental purpose is of the standard form named GEMEP-FERA[4] . This dataset contained videos of length varying from 2-5 sec where different person(subject) shows the expression by uttering some significant words. These words are considered because the expression should come naturally. Detecting Facial Expression from static (still) images is of challenging due

to the ambiguity occurring which expression falls under which category of group. Donato and his group [5] performed comparison of different method for classifying different facial expression. This methods are principal component analysis(PCA), Independent component analysis(ICA), optical flow and local filters such as Gabor wavelet. Lucey and Matthews[6] proposed Active Appearance Models used for extracting facial features post fitting and machine learning technique to classify emotions.

The paper is organized as follows. Section II gives details of the system. Section III views the methods used for obtaining features vectors then this features vectors are feed to the classifier for classifying emotions of different subject into particular expression class. Section IV gives details about the results obtained for PHOG method, LBP method and combined features of both methods. Section V gives details about conclusion and future work.

II. DETAILS OF SYSTEM

The facial expression recognition system is provided with the input in the form of video of length 2-5 sec. From this video, which consists of frames are processed to obtain key frames. Those key frames are of importance for next processing i.e face acquisition step. After the key frames are obtain, those are feed to the haar classifier for face detection step. This face detection is done automatically where the key frames are processed to obtain only the face region. After the face region is obtain, the PHOG and LBP method acts on to obtain the features vectors in the form of number values. This Features vectors then provided to the K-NN classifier for classification into respective classes. Figure I shows the diagram of PHOG and LBP method for training set during providing training to the videos. Figure II shows details of the testing set during testing of videos. This test videos are taken for classification into respective classes.

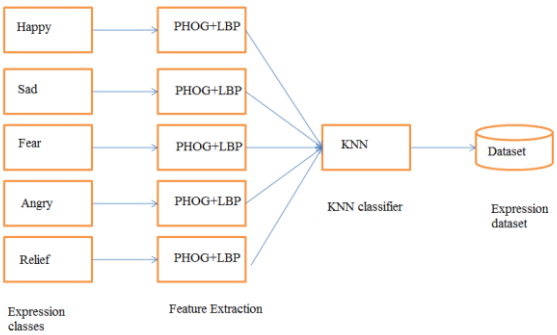

Fig. I Diagram for training set

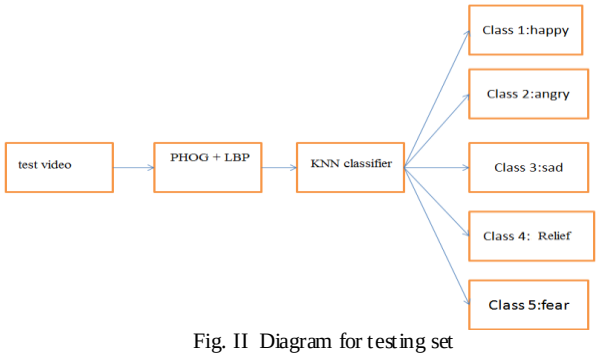

Fig. II Diagram for testing set

III. METHODOLOGY

The stated methods mainly consists of following steps: key frame detection from video, after this face acquisition from key frame, feature extraction step from face region using local methods i.e. PHOG and LBP, lastly expression classification using k-nn classifier

A. Key Frame Detection

The image segmentation is done for key frame detection using the novel method [7] of key frame detection. Each frame of video is partition into block of size mxn. After this the histogram matching difference of two consecutive frame is calculated. Mean difference and standard deviation is calculated for each histogram difference. Some standard threshold value is used by considering this threshold a key frame are detected.

B. Face Acquisition and Detection

Due to high detection accuracy, efficiency and optimal performance, a Haar classifier based method is used [8]. Haar features compute the difference in average intensity in different region of the image and form black, white connected rectangle in which the value of the feature is the difference of sum of pixel values in these region.

C. Feature Extraction

PHOG, a spatial shape descriptor method is used for image classification in particular class. PHOG [9] gives the gradient features along the edges of the region of interest. It form the histogram of pyramid along either 180 or 360. For experimental purpose, we used the orientation range of [0-360].

Step 1: In the first step, we extract the edge contours along the region of interest like forehead, eye region, mouth region and so on. For this canny edge detector is used which also brief the weak edges along with the strong edges.

Step 2: Face image is divided into cells of grid level and the gradients values for each grid and for all pyramid levels are computed. The standard pyramid level used is 3.

Step 3: The final PHOG descriptor of an face is formed by concatenation of all the Gradients vectors. This vectors of gradients are formed for all components like red, green and blue of colour formats.

LBP [10], local binary pattern feature extraction method is used due to its light independent property and low

computational complexity. The values of neighbourhood are thresholded by the center value and the result is formed as a binary number. Like this the whole face region is processed and binary pattern is obtain in the form of histogram.

D. Facial Expression Classification

For classification we trained a K-NN classifier [11] and the parameters are selected with cross validation. k-NN is trained by providing supervised learning using training videos set. The parameters for K-NN for expression classification are all fixed on training videos, by following a leave-one-video-out setting for person-independent classification.

IV. EXPERIMENTS AND RESULTS

The GEMEP-FERA dataset developed at automatic facial expression recognition competition FERA 2011 [12] consists of videos of 8-10 actors depicting different expressions, while uttering a meaningless phrase. There are 7 subjects in the training data, and 6 subjects in the test data, few of them are not present in the training set. There are in total five emotion categories in the dataset: happy, sad, fear, angry and relief. Table 1 shows the result of PHOG method and table 2 shows the result of the PHOG+LBP method.

TABLE I
RESULTS FOR EXPRESSION CLASSIFICATION of PHOG

	happy	sad	fear	angry	relief
happy	16	0	1	2	2
sad	0	17	2	1	1
fear	0	2	12	3	4
angry	2	0	1	8	3
relief	1	1	1	3	7

The overall accuracy of the expression classification for PHOG method is near about 0.64 and that for the combination of PHOG+LBP method it is near about 0.78 when the combine feature are considered as shown in table II.

TABLE II
RESULTS OF PHOG+LBP

	happy	sad	fear	angry	relief
happy	16	0	0	1	2
sad	1	17	2	2	2
fear	1	2	14	2	1
angry	0	0	1	12	0
relief	1	1	1	0	12

V. CONCLUSION AND FUTURE SCOPE

We presented a combination of PHOG and LBP method for features extraction from face region. The time required for processing is less as compare to others global descriptors. The GEMEP-FERA dataset is standard dataset where different subject gives different expressions. The combination of method gives improves accuracy as compare to individual features of the method. The proposed methodology is limited to classify frontal form of images. In the future the research will be extended to classify the 3D based images.

ACKNOWLEDGMENT

The best paper award was given for this work in previous paper. The academic mail id was required for this standard dataset. The work was greatly supported by the department.

VI. REFERENCES

[1] B.Fasel and J. Luettin,"Automatic facial expression analysis: a survey," Pattern Recognition, vol.36,no. 1,pp.259-275,2003.

[2] P. Ekman and E.L. Rosenberg, what the face reveals: Basic and applied studies of spontaneous expression using the facial action coding system. USA:Oxford university Press,1997.

[3] Y.Wang, H.Ai, B.Wu, and C.Huang."Real time facial expression recognition with adaboost," in Proceedings of the 17th international conference on pattern recognition,2004,pp.926-929.

[4] T. Banziger and K.R. Scherer,"Introducing the geneva multimodal emotion portrayal (gemep) corpus," in Blueprint for affective computing: A sourcebook, T. B. K. R. Scherer and E. R. (Eds.), Eds. Oxford England: Oxford University Press.

[5] G.Donato, M.S.Bartlett,J.C. Hager,P.Ekman, and T.J. Sejnowski,"Classification facial actions,"IEEE TPAMI,vol. 21,no. 10,pp. 974-989,1999.

[6] S. Lucey,I. matthews, C.Hu,Z. Ambadar,F.de la Torre, and J. Cohn,"Aam derived face representations for robust action recognition," in IEEE AFGR, 2006.

[7] ZHAO Guang-sheng,"A Novel Approach for shot boundary detection and key frames extraction,"2008 International Conference on multi media and information technology.

[8] P.Viola and M.J. Jones,"Robust real-time face detection,"international journal of computer vision, vol.57,no. 2,pp. 137-154,2004.

[9] A. Bosch, A. Zisserman, and X.Munoz,"Representing shape with a spatial pyramid kernel," in Proceedings of the ACM international Conference on Image and Video Retrieval,2007.

[10] T. Ojala, M Pietikainen, and T. Maenpaa,"Multiresolution gray-scale and rotation Invariant texture classification with local binary patterns,"IEEE transaction on Pattern Analysis and machine Intelligence,vol.24,no. 7,pp. 971-987,2002.

[11] T. Denoeux, "A k-nearest neighbor classification rule based on Dempster–Shafer theory," *IEEE Trans. Syst Man, Cybern.*, vol. 25, pp. 804–813, May 1995.

[12] M.Valstar,B. Jiang, M. Mehu, M. Pantic, and S. Klaus," The first facial expression recognition and analysis challenge," in Automatic Face and Gesture Recognition,2011