ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 12, December 2013

A Survey on Facial Expression Analysis for Emotion Recognition

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

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

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

Abstract: Over the last decade,facial expression recognition also stated as FER has become an active research area that finds significant applications in areas such as human-computer interfaces,talking heads,image retrieval and human emotion analysis like happy,sad,neutral,disgust,anger,surprise etc. Facial expression reflects not only emotion but other mental activities such as those in cases of clinical approaches. This survey also deals with brief details of various approaches like optical flow method, local binary patterns, Pyramid of histogram of gradient (PHOG) and Local phase quantisation(LPQ) method, Facial action coding system(FACS).

Keywords: FER, Human Emotion Analysis, Optical flow, Local binary patterns, PHOG and LPQ, FACS.

I. Introduction

Human-Computer Interaction (HCI) aims at significant interaction between human and machines by body gestures, eye gaze, speech, facial expressions, cognitive modelling etc. Facial expressions can be used as an efficient way of emotion detection, thus facilitating HCI. A natural way of interaction between man and machine can be obtained by detection and classification of facial expressions. However facial expressions and its intensity vary from person to person, even the expression of the same subject (person) varies with respect to time. Hence, designing a generalized method for facial expression analysis is a very tough task. However, specified algorithms can be used to analyze facial expressions of a single person accurately.

Facial expression analysis in recent time make an advancements in face detection, face tracking, and face recognition techniques. Ekman and Keltner developed Facial Action Coding System(FACS) [1] which describes facial expressions in terms of Action Units (AUs). FACS consists of 46 AUs which are primarily related to facial muscle movements. Facial expression recognition deals with the classification of facial motion and facial feature deformation into abstract classes that are purely based on visual information. Facial expression analysis is mainly divided into two categories: images based and video based. In real world, human facial expressions are dynamic in nature. They constitute an onset, one or more apex (peaks) and an offset. All methods that have been proposed for facial expression analysis can be studied in two parts: First, methods based on the characteristics of total face (holistic method); second, methods which work in important parts of face (geometric methods) that works on Facial Action Units [FAUs]. The first group focuses on the features of whole image using different algorithms. In other methods, some points (Land-marks) are selected (manually or automatically) on the active units of face and used them in other images, then facial expression can be extracted.

This paper is organized as follows: Section 2 includes Measurement of Facial Expression, Section 3 includes different methods overview and Section 4 includes conclusion.

II. Measurement of Facial Expression

Contractions of facial muscles leads to generation of facial expressions, which results in temporally deformed facial features such as eye lids, eye brows, nose, lips and skin texture, often shown by wrinkles and bulges. Geometric deformation of facial features or the density of wrinkles appearing in certain face region help to measured facial expression intensities.

A. Judgment-based approaches

Judgment-based approaches are centered around the messages conveyed by facial expressions during facial expression analysis. However, during classification of facial expressions into a predefined number of emotion, an agreement of a group of coders is taken as ground truth. Ekman and Friesen [2] introduced a System which directly map most of the facial expressions into one of the basic emotion classes.

B. Sign-based approaches

Using revealed sign-based approaches, facial motion and deformation are coded into visual classes. Facial actions are abstracted uniquely and brief by their location

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 12, December 2013

and intensity. It is appearance-based and does not convey any other information related to the mental activity. FACS uses 44 action units (AUs) for the description of facial actions in accordance to their location and intensities. Individual expressions may be modelled by single action units or action unit combinations. Fig. 1. shows basic structure of facial expression analysis.

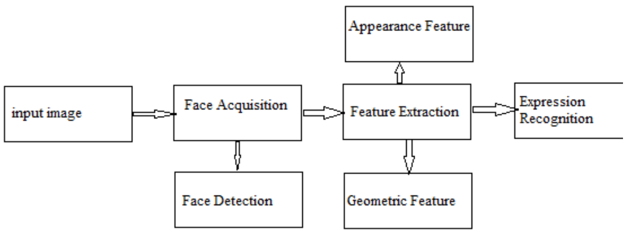

Fig. 1. Basic Structure of Facial Expression Analysis

III. Different Methods Overview

This section deals with the different approaches that can be used for the geometric features and appearance features extraction.

A. *Local Binary Patterns*

Local binary patterns (LBP) is used as an efficient method for facial expression analysis [3]. LBP features look into minor changes of facial expressions for different emotions. It extracts the local features from face region and after extracting local features, different expressions are classified to different classes of emotions. The most important part in a classification problem is selection of the features vector. The inherent problem related to image classification include the pose, scale, translation and variation in illumination level. The feature vector extracted should be invariant to these difficulties to achieve better accuracy in classification problem.

LBP method is used due to its efficient light invariance property and low computational complexity [4]. The neighbourhood values are thresholded by the center value and the result is obtained in the form of binary number. Fig. 2 shows the details of LBP feature calculation procedure. The binary weighted summation of values in 3x3 neighbourhoods are thresholded and this gives the feature vector at that location.

$$LBP = \sum_{n=0}^{7} s(i_n - i_c)2^n \quad (1)$$

Where, i_c is the central pixel value.

3x3 neighbourhood

12	63	78
155	67	55
143	210	20

Thresholded

0	0	1
1		0
1	1	0

Binary Weights

1	2	4
128		8
64	32	16

Pattern 11100100

LBP = 4+32+64+128=228

Fig. 2. Calculation of LBP

Now for histogram formation of LBP, the LBP coefficients are collected in a histogram, but it results in the distortion of spatial information. The image is divided into several small blocks to retained the spatial information, after this the LBP histogram of each small region is found. In this way, LBP works on the local parts of the face region.

B. *Optical Flow*

Optical flow tries to compute a flow field that minimizes the difference between the two images involved in the process [5]. It is useful for estimating motion in an image sequence without tracking object [6]. An appropriate mask is designed using Gabor filters, and it is then convolved with first frame of video sequence image. Then oval part of the face is specified and its main components are characterized. During frames, the optical flow is used on the regions of main parts, and then it is classify to the six basic classes such as neutral, happy, sad, disgust, surprise and anger for facial expression analysis. This method has high accuracy in comparison with other methods and don't need to select landmark point manually at the beginning.

Optical flow applied on each video sequence images from first frame until the last frame is reached and maximum vectors of points are formed that moved larger than others were extracted as shown in Fig. 3. The position (co-ordinate) of extracted vectors with their arguments and phases, are features that classify them to achieve the goal of accuracy.

Fig. 3. Largest motion flow vectors

C. *PHOG and LPQ*

PHOG [7] is a spatial pyramid extension of the histogram of gradients (HOG) [8] descriptors. PHOG features are used for shape (geometric) features extraction which are permanently present on the face region and changes according to the deformation of the face portion. The HOG descriptor technique counts occurrences of gradient

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 12, December 2013

orientation in the localized portions of an image and has been used extensively in the computer vision methods. PHOG features [9] is used as an extension of HOG and had shown good performance in object recognition and static facial expression analysis [10]. In the beginning canny edge detector method is applied to the face which is normalized after the cropping is applied on original face. After this the face is divided into spatial grids at all pyramid levels. At last, the 3x3 Sobel mask is applied to the edge contours of the face for calculation of the orientation gradients and then the gradients of each grid are joined together at each pyramid level.

For appearance feature extraction, a LBP family of descriptor from Local Phase Quantization (LPQ) method is used owing to its ability for texture analysis, static and temporal facial expression analysis and face recognition[11]. LPQ is basically based on calculating short-term fourier transform (STFT) on local image window. For the image level, at each pixel coefficients for the local fourier are computed for four frequency points. For phase information, the sign of the real and the imaginary part of each coefficient is quantized using a binary scalar quantiser. The resultant eight bit binary coefficients are then represented as integers using binary coding. This step of calculation is similar to histogram construction step in LBP. This method works better than baseline method used on the GEMEP-FERA data set. The result is presented in the form of the confusion matrix which compare its outcome with the baseline method.

D. FACS (Facial Action Coding System)

The facial action coding system (FACS) [12] is the most objective and comprehensive coding system in the behavioural sciences. The total of 46 component movements are formed by decomposing facial expression, which corresponds to the 44 facial muscles. According to the manual of FACS, the facial expression analysis system estimates the measurement of actions and also classifies the actions [13]. The overview of fully automated facial action coding system is shown in Fig. 4.

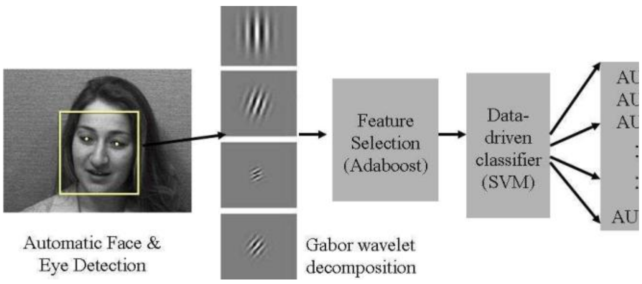

Fig. 4. Overview of fully automated facial action coding system

The facial action coding system was trained on Cohn and Kanade's dataset which contains FACS scores by two certified FACS coders in addition to the basic emotion labels. Automatic eye detection [14] was done to align the eyes in each image. Images were scaled to 192x192, passed through a bank of Gabor filters at 8 orientations and 7 spatial frequencies. There were in total of 18 action units for which there were at least 15 examples in the dataset. Separate support vector machines were trained one for each AU to perform context-independent recognition. During this recognition, the system detects the presence of a given AU regardless of the co-occurring AU's. Generalization to new subjects was tested using leave-one-subject-out cross-validation. The results are shown in table 1.

TABLE I
Performance for fully automatic recognition of 18 facial actions, generalization to novel subjects. N: Total number of examples of each AU, including combinations containing that AU. agreement: Percent agreement with Human FACS codes. Nhit:FA: Raw number of hits and false alarms.

AU	Name	N	Agreement	Nhit:FA
1	Inner brow raise	123	93%	98:15
2	Outer brow raise	83	96%	69:11
4	Brow corrugator	143	89%	103:29
5	Upper lid raise	85	92%	49:16
6	Cheek raise	93	94%	71:16
7	Lower lid tight	85	87%	37:32
9	Nose wrinkle	43	99%	35:0
11	Nasolabial furrow	23	96%	3:0
12	Lip corner pull	73	98%	62:6
15	Lip corner depress	49	95%	27:12
17	Chin raise	124	91%	91:20
20	Lip stretch	51	96%	31:6
23	Lip tighten	38	94%	10:12
24	Lip press	35	95%	14:6
25	Lips part	118	94%	94:10
26	Jaw drop	18	97%	3:0
27	Mouth stretch	51	98%	46:12
44	Eye squint	18	97%	5:6

The FACS system obtained a mean of 94.5% agreement with human FACS labels. The appearance of a facial action and the direction of movement frequently change when the action occurs in combination with other actions. Combinations are typically handled for specific AU combinations by developing separate detectors in case of more than 1 AU's in combination.

Today, most facial expression analysis systems attempt to map facial expressions directly into abstract classes and thus are unable to handle facial actions caused by non-emotion mental and physiological activities. FACS may provide a solution to this dilemma, as it allows to classify facial actions prior to any interpretation attempts. However, many facial expression analysis system have attempted to map facial expressions directly towards basic emotions, which represents an ill-posed problem. Decoupling facial expression recognition and facial expression interpretation may provide a solution to this dilemma. This solution can be achieved by first coding facial expressions with an appearance-based representation scheme such as facial action coding system (FACS) and then using facial expression dictionaries in order to translate recognized facial actions into mental activity categories.

IV. CONCLUSION

The Block LBP histogram features extract local as well as global features of face image resulting in higher accuracy. however, LBP methodology is limited to classify frontal image only. The accuracy is near about to 97% when applied to the real images.

Optical flow works on the regions of main parts during frames of image sequence, and classify them to the six basic classes such as: neutral, happy, sad, surprise, anger, disgust. this method has high accuracy in comparison with other methods and no needs to select landmarks manually at first.

For capturing shape information, PHOG features are used and for appearance, recently proposed LPQ features are used. This method had perform on GEMEP-FERA dataset and shown better results than baseline methods. For future work, this method can be implemented on other dataset.

FACS coding is the basic method which estimate the AU and classify on the basic of this AU. The expression is then classify by considering this single AU or its combination with other AU's.

REFERENCES

[1] 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.

[2] P. Ekman, W. V. Friesen, "Constants across cultures in the face and emotion", J. Personality Social Psychol. 17 (2) (1971) 124-129.

[3] S L Happy, Anjith George, Aurobinda Routray, "A Real Time Facial Expression Classification System Using local Binary patterns", IEEE Proceedings of 4th International Conference on IHCI, Kharagpur, India, December 27-29,2012.

[4] T. Ojala, M Pietikainen, and T. maenpaa, "multiresolution gray-scale and rotation invariant texture classification with local binary patterns," IEEE transactions on pattern Analysis and machine Intelligence, vol. 24, no. 7, pp 971-987, 2002.

[5] Behnam Kabirian Dehkordi, javad Haddadnia,"Facial Expression Recognition in Video Sequence Images by Using Optical Flow", IEEE Proceedings of 2nd International conference on Signal Processing Systems (ICSPS), 2010.

[6] L. Streeter, "A direct Geometric Algebra Optical Flow Solution", University of Waikato, Department of Physics and Electronic Engineering, Hamilton.

[7] 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.

[8] N .Dalal and B.Triggs, "Histograms of oriented gradients for human detection," in international Conference on Computer Vision & pattern Recognition, vol. 2,June 2005, pp.886-893.

[9] Abhinav Dhall, Akshay Asthana, Roland Goecke, Tom Gedeon, " Emotion Recognition Using PHOG and LPQ features",In IEEE AFGR2011 Workshop FERA, pages 878-883,2011.

[10] Z. Li, Imai, and M.Kaneko," Facial components-based bag of words and phog descriptor for facial expression recognition," in proceedings of the 2009 IEEE international conference on Systems, Man and Cybernetics, ser. SMC'09, 2009.

[11] V. Ojansivu and J. Heikkil, "Blur insensitive texture classification using local phase quantization," in Image and Signal Processing, ser. Lecture Notes in Computer Science, 2008.

[12] P. Ekman and W. Friesen. facial Action Coding System: "A technique for the measurement of facial movement". consulting Psychologists Press, Palo Alto, CA, 1978.

[13] Marian Stewart Bartlett, Gwen Littlewort, Claudia lainscsek,Ian Fasel, Javier Movellan, "machine Learning methods for Fully Automatic Recognition of Facial Expressions and Facial Actions", Proceedings of the IEEE Conference on Systems,man & cybernetics, the Hague, Netherlands, 2004.

[14] I. R. Fasel, B. Fortenberry, and J. R. Movellan . GBoost:"A generative framework for boosting with applications to real-time eye coding". Computer Vision and Image Understanding, in press.