International Journal of Computer Applications (0975 – 8887)
Volume 181 – No. 35, January 2019

Local Appearance based Novel Facial Feature Extraction Method for Human Expression Recognition

Mohammad Shahidul Islam
IIT
Jahangirnagar University
Savar, Dhaka, Bangladesh

Tarikuzzaman Emon
CSE
Stamford University Bangladesh
Dhaka-1209

Md. Zahid Hasan
CSE
Daffodil International University
Bangladesh

ABSTRACT

A novel approach to extract the light invariant local feature for facial expression recognition is presented in this paper. It is robust in monotonic gray-scale changes caused by illumination variations. Proposed method is easy to perform and time effective. The local strength for a pixel is calculated by finding the decimal value of the neighboring of that pixel with the particular rank in term of its gray-scale value among all the nearest pixels. When eight neighboring pixels are considered, the gradient direction of the neighboring pixel with the mix of second minima and maxima of the gray scale intensity can capture more local details and yield the best performance for the facial expression recognition in our experiment. CK+ dataset is used in this experiment to find out the facial expression classification. The classification accuracy rate achieved is 92.1 ± 3.2%, which is not the best but easier to compute. The results show that the experimented feature extraction technique is fast, accurate and efficient for facial expression recognition.

Keywords

Emotion Classification, Expression Recognition, Image Analysis, Local Descriptor, Pattern Extraction.

1. INTRODUCTION

Human facial expression is an important role in human-to-human contact. It allowing people to communicate each other beyond the verbal world and understand each other from various modes. Some expressions helps human for actions, and others enhance the interaction meaning. Human computer interfaces must detect crucial changes in the user's behavior than simply responding to the user's instructions. Facial expression recognition is a challenging field in human computer interaction, human computer interface and computer vision. Due to its prospective vital applications, many young researchers chose this area as their research interest (Z. Zeng *et al*, 2009). Among the different methods, appearance-based methods have been heavily employed in this research field with huge success. Some of them are LGP- local gradient pattern, LMn- local minima, LTP-local ternary pattern, Gabor filters, local binary patterns (LBP) descriptors, Haar wavelets and subspace learning methods. A. Mehrabian (1968) wrote in his paper that the verbal part of a communication contributes only 7 percent of its overall meaning, the voice part helps 38 percent while facial movement and the expression helps 55 percent of the meaningful communication, see Fig. 1.

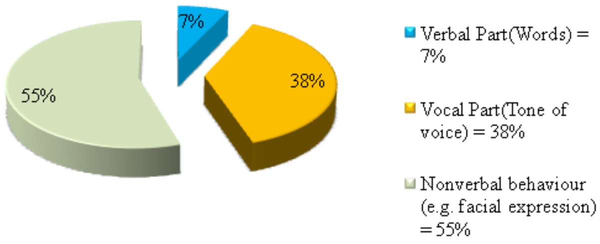

Fig. 1: 7%-38%-55% rule by A. Mehrabian (1968)

This highly meaningful expression has seven basic subdivisions,

1. contempt,
2. fear,
3. sadness,
4. disgust, 5. anger,
6. surprise and
7. happiness

MA. Peter et al. (2018) developed a instrument called Product Emotion Measurement instrument (PrEmo), which can detect many emotion types. It is found in many past research papers that most of the facial expression recognition systems (FERS) are based on the Facial Action Coding System (FACS), Y.L. Tian *et al.* (2001), Y. Tong *et al.* (2007), M. Pantic *et al.* (2000). 44 different action units (AUs) are formed using FACS is above research papers. Up-to 7000 dissimilar combinations are possible using these AU's, with wide variations of human e.g. age, size and ethnicity. M. Pantic *et al.* (2000) compared many methods on facial expression recognition. Still images are easy to obtain and good for initial learning. This is a reason for young researchers to choose still image data sets. One of the psychological experiment by J.N. Bassili (1979) has proposed that facial expressions are recognized better with higher accuracy from video. His research encouraged many researchers to work with video than still images. I. Kotsia *et al.* (2007) created a facial wire frame model from still image in their paper which can capture more details but takes more space and also slower than some other methods. Y. Zhang *et al.* (2005) introduce a novel way by using IR camera to capture facial features in details. Y.L. Tian *et al.* (2001) experimented on multi state face components and created a model of multi state AUs which was classified by neural network. M. Yeasin *et al.* (2007) enhanced the Markov model for facial expression recognition. K. Anderson *et al.* (2006) used video as dataset and created a model named the multichannel gradient model (MCGM) to find the facial optical flow. The motion signatures obtained from the MCGM are then used as input for Support Vector Machines to classify. I. Cohen *et al.* (2003) also worked on videos. They used Naive–Bayes as classifier and hidden

Markov models (HMMs) to collect facial expression features. M. Pantic *et al.* (2006) tried contour tracking and rule-based analysis to find 20 AUs from facial front view and side view. T. Ahonen *et al.* (2006) proposed a facial expression recognition model using Local Binary Pattern (LBP). The original LBP operator by T. Ojala *et al.* (1996), compares each pixel with the center pixel value in a 3x3 local area and transforming the result as a binary number, see Fig. 2.

a(1)	a(2)	a(3)
a(8)	e	a(4)
a(7)	a(6)	a(5)

Fig. 2: 3x3 pixels, local image region, e is the center and a(1-8) are the neighboring pixels.

Fig. 3 shows an example of obtaining an LBP from a given 3×3 pattern.

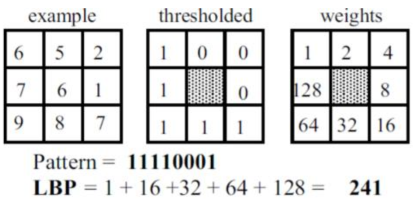

Fig. 3: Example of obtaining LBP from a 3x3 local region.

Experiments using FACS are more complex while the appearance-based calculations are less complex. Hence, a new and easy appearance-based feature extraction method is experimented in this paper that would be more efficient for facial expression recognition.

2. PROPOSED METHODOLOGY

2.1 Local Minima and Maxima

Images are converted into gray scale first. From the image 3x3 region, shown in Fig. 2, is taken at a time. The center pixel of the region is surrounded by 8 neighboring pixels. The position of the neighboring pixels with the minimum in value and maximum in value can be considered as the local feature for the given center pixel. In the previous work, MS. Islam *et al.* (2013), only the minima was considered. In Fig. 2, a(1) pixel is in position 1, a(2) pixel is in position 2 and so on. Each position is considered as a bin. E.g. Bin 1-8. Therefore if the pixel at position 1 e.g. a(1) has the minimum/lowest gray value among a(1-8) pixels, bin-1 will be increased by one. Similarly another eight bins for the maxima starting from bin-9 to bin-16.

253	87	23
56	240	172
125	210	25

Fig. 4: a(1)=253, a(2)=87,...... a(8)=56. a(3)=25 is the minimum in gray color scale and a(1)=253 is the maximum in gray color scale.

In fig. 4, a(3)=25 is the minimum in gray value and a(1)=253 is the maximum in gray value. Therefore bin-3 and bin-9 will be increased by one for pixel 240. Similar way local feature is extracted for each pixel of the image.

2.2 Experimental Setup

The full experimental setup is subdivided into three steps:

a) facial feature calculation(bin value),

b) SVM training and (classifier training)

c) facial expression determination.(classification)

All through the experiment, the Extended Cohn-Kanade Dataset (CK+) (P.Lucey *et al.*, 2010) is used. There are 123 subjects and 326 peak facial expressions of those subjects. The expressions in this dataset are subdivided into 7 categories, Anger, Contempt, Disgust, Fear, Happy, Sadness and Surprise. Each subject has several same expression but for the experiment it is collected only once. The data distribution of the dataset is shown in Table 1**.**

Table 1: CK+ Dataset, 7 expressions and numbers of instances of each expression

Expression	Numbers of instances
contempt,	18
fear,	25
sadness,	28
disgust,	59
anger,	45
surprise and	82
happiness	69

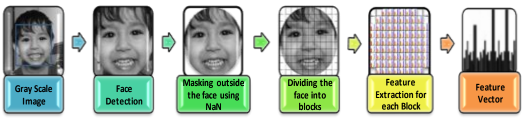

Fig. 5: Facial feature calculation step by step

Facial feature calculation phase is shown in Fig. 5. In the experiment **fdlibmex**, a free code for Matlab is used for face detection. Face is then resized into lower resolution to ease the calculation. Also masked to remove unnecessary area, thus the calculation get less and more easier. The 180x180-sized masked face (3rd block of Fig. 5) is subdivided into 9x9=81 blocks of 20 by 20 pixels each. Proposed method is used for feature calculation. The histogram of all the blocks are concatenated into a unique feature vector. Therefore, the length of the feature vector is (8+8)x9x9= 1296. LIBSVM, by C.C. Chang *et al.* (2011) is trained using feature vectors of all images in the training phase. Using multiclass SVM module, a 10-fold cross validation is performed.

3. EXPERIMENTAL RESULTS AND ANALYSIS

Table 1 shows the achieved classification accuracy rates. The 2nd minima (LMn-2) and the 1st maxima (LMx-1) resulted better with accuracy of 92.1%±3.2. This is obtained by averaging the accuracy of 10-fold cross validation. The maximum accuracy resulted by any fold from the ten folds is 94.8% and the minimum is 89.4%. Same experiment is performed with different block sizes as shown in Table 3. Though the block size of 15x15 pixels gives the highest accuracy rate but there is a penalty in feature vector length.

Table 2. Classification in terms of Accuracy of the 1st minima to the 1-8 Local Maxima, 2nd minima to the 1-8 Local Maxima and so on.

Local Minima	Local Maxima	Classification Accuracy
1st (LMn-1)	1st (LMx-1)	89.8%±3.2
	2nd (LMx-2)	91.1%±3.2
	3rd (LMx-3)	88.8%±3.2
	4th (LMx-4)	91.9%±3.2
	5th (LMx-5)	88.9%±3.2
	6th (LMx-6)	90.2%±3.2
	7th (LMx-7)	90.5%±3.2
	8th (LMx-8)	91.4%±3.2
2nd (LMn-2)	**1st (LMx-1)**	**92.1%±3.2**
	2nd (LMx-2)	90.1%±3.2
	3rd (LMx-3)	91.8%±3.2
	4th (LMx-4)	90.9%±3.2
	5th (LMx-5)	91.9%±3.2
	6th (LMx-6)	90.2%±3.2
	7th (LMx-7)	89.5%±3.2
	8th (LMx-8)	89.4%±3.2
3rd (LMn-3).	......	....

Table 5 shows comparisons of the individual expression accuracy and the average all 7 class expression accuracy achieved by the proposed method and the other recent methods using shape or combination of shape and texture information.

Table 3. Feature length Vs Block Dimension.

Resized Face (Pixels)	Total Blocks	Block Dimension (pixels)	Classification Accuracy (%)	Feature Vector Length
180 x 180	6x6=36	30 x 30	90.1%±3.2	576
180 x 180	**9x9=81**	**20 x 20**	**92.1%±3.2**	**1296**
180 x 180	10x10=100	18 x 18	91.8%±3.2	1600
180 x 180	**12x12=144**	**15 x 15**	**94.9%±3.2**	**2304**
180 x 180	15x15=225	12 x 12	94.1%±3.2	3600
180 x 180	18x18=324	10 x 10	90.2%±3.2	5184

Table 4. Confusion Matrix for LMn-2 + MLx-1

LMn-2 + LMx-1 = 10-fold validation
Feature Extraction time for 326 Images = 192 Seconds
Average Classification Accuracy = 92.1 ± 3.2%
Kernel parameter: = (-s 0 -t 1 -c 100 -g 0.0015 -b 1)
Confusion Matrix:
C :

Actual

prediction	Angry	Contempt	Disgust	Fear	Happy	Sad	Surprise
Angry	39	0	2	0	0	4	0
Contempt	1	16	0	0	0	1	0
Disgust	2	0	54	2	1	0	0
Fear	1	1	0	21	1	0	1
Happy	0	0	0	0	69	0	0
Sad	4	1	1	1	0	20	1
Surprise	0	1	0	1	0	0	80

Table 5. Comparison the average accuracy (∑ (Accuracy of all 7 expressions/7)) of different methods. [S: shape based method, T: texture based method. S + T: both shape and texture based method (CLM-Constrained Local Model, AAM-Active Appearance Model, Avg.=Accuracy of all expressions/7)]

Authors	Method	T/S	Avg.
P. Lucey *et al.* (2010)	AAM + SVM	S	50.3
	AAM + SVM	T	66.7
	AAM + SVM	T + S	83.3
S.W. Chew *et al.* (2011)	CLM + SVM	T	74.4
L.A. Jeni *et al.* (2012)	CLM + SVM (AU0 norm.)	S	77.6
	CLM + SVM (personal mean shape)	S	86.8
Proposed Method (LMn-2+LMx-1)	**No Registration + SVM**	**T**	**92.1**

Due to different experimental setups and version differences of the CK (T. Kanade *et al.*, 2000) dataset , the results are not in a straight line comparable. L.A. Jeni *et al.* (2012) mentioned in their research that straight aligned faces can give an extra 5% to 10% boost in the expression classification accuracy. Leave-one-subject-out type validation can boost the accuracy by another 1-2% , (M.S. Bartlett *et al.*, 2003).

4. CONCLUSION

A novel local appearance based facial feature extraction method for human expression recognition in experimented in this paper. It obtains some crucial features from a gray scale image. The neighboring pixels with local minimum and maximum value on the gray scale color are used to identify those local features for the pixel in the center. Eight possible minima and eight possible maxima can be considered as a local feature for a given pixel; however, the second minima along with the 1st maxima results the highest recognition in the experiment. As its simplicity, it can be incorporated with other boosting methods to increase the accuracy of recognition.

5. REFERENCES

[1] Desmet, Pieter. "Measuring emotion: Development and application of an instrument to measure emotional responses to products." In Funology 2, pp.391-404. Springer,Cham, 2018.

[2] Mehrabian, Albert. "Communication without words." Communication theory (2008): 193-200.

[3] Chang, Chih-Chung, and Chih-Jen Lin. "LIBSVM: a library for support vector machines." ACM transactions on intelligent systems and technology (TIST) 2.3 (2011): 27.

[4] Islam, Mohammad Shahidul, and Surapong Auwatanamongkol. "A novel feature extraction technique for facial expression recognition." International Journal of Computer Science Issues (IJCSI) 10, no. 1 (2013): 9.

[5] Cohen, Ira, Nicu Sebe, Ashutosh Garg, Lawrence S. Chen, and Thomas S. Huang. "Facial expression recognition from video sequences: temporal and static modeling." Computer Vision and image understanding 91, no.1-2 (2003):160-187.

[6] Kotsia, Irene, and Ioannis Pitas. "Facial expression recognition in image sequences using geometric deformation features and support vector machines." IEEE transactions on image processing 16, no. 1 (2007): 172-187.

[7] Bassili, John N. "Emotion recognition: the role of facial movement and the relative importance of upper and lower areas of the face." Journal of personality and social psychology 37, no. 11 (1979): 2049.

[8] Anderson, Keith, and Peter W. McOwan. "A real-time automated system for the recognition of human facial expressions." IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 36, no. 1 (2006): 96-105.

[9] Jeni, LáSzló A., András Lőrincz, Tamás Nagy, Zsolt Palotai, Judit Sebők, Zoltán Szabó, and Dániel Takács. "3D shape estimation in video sequences provides high precision evaluation of facial expressions." Image and Vision Computing 30, no. 10 (2012): 785-795.

[10] Pantic, Maja, and Ioannis Patras. "Dynamics of facial expression: recognition of facial actions and their temporal segments from face profile image sequences." IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 36, no. 2 (2006): 433-449.

[11] Pantic, Maja, and Leon J. M. Rothkrantz. "Automatic analysis of facial expressions: The state of the art." IEEE Transactions on pattern analysis and machine intelligence 22, no. 12 (2000): 1424-1445.

[12] Yeasin, Mohammed, Baptiste Bullot, and Rajeev Sharma. "Recognition of facial expressions and measurement of levels of interest from video." IEEE Transactions on Multimedia 8, no. 3 (2006): 500-508.

[13] Bartlett, Marian Stewart, Gwen Littlewort, Ian Fasel, and Javier R. Movellan. "Real Time Face Detection and Facial Expression Recognition: Development and Applications to Human Computer Interaction." In Computer Vision and Pattern Recognition Workshop, 2003. CVPRW'03. Conference on, vol. 5, pp. 53-53. IEEE, 2003.

[14] Lucey, Patrick, Jeffrey F. Cohn, Takeo Kanade, Jason Saragih, Zara Ambadar, and Iain Matthews. "The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression." In Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on, pp. 94-101. IEEE, 2010.

[15] Chew, Sien Wei, Patrick J. Lucey, Simon Lucey, Jason Saragih, Jeffrey Cohn, and Sridha Sridharan. "Person-independent facial expression detection using constrained local models." Proceedings of FG 2011 Facial Expression Recognition and Analysis Challenge (2011): 915-920.

[16] Ahonen, Timo, Abdenour Hadid, and Matti Pietikainen. "Face description with local binary patterns: Application to face recognition." IEEE Transactions on Pattern Analysis & Machine Intelligence 12 (2006): 2037-2041.

[17] Kanade, Takeo, Yingli Tian, and Jeffrey F. Cohn. "Comprehensive database for facial expression analysis." In fg, p. 46. IEEE, 2000.

[18] Ojala, Timo, Matti Pietikainen, and Topi Maenpaa. "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns." IEEE Transactions on pattern analysis and machine intelligence 24, no. 7 (2002): 971-987.

[19] Tian, Y-I., Takeo Kanade, and Jeffrey F. Cohn. "Recognizing action units for facial expression analysis." IEEE Transactions on pattern analysis and machine intelligence 23, no. 2 (2001): 97-115.

[20] Tong, Yan, Wenhui Liao, and Qiang Ji. "Facial action unit recognition by exploiting their dynamic and semantic relationships." IEEE transactions on pattern analysis and machine intelligence 29, no. 10 (2007).

[21] Zhang, Yongmian, and Qiang Ji. "Active and dynamic information fusion for facial expression understanding from image sequences." IEEE Transactions on pattern analysis and machine intelligence 27, no. 5 (2005): 699-714.

[22] Zeng, Zhihong, Maja Pantic, Glenn I. Roisman, and Thomas S. Huang. "A survey of affect recognition methods: Audio, visual, and spontaneous expressions." IEEE transactions on pattern analysis and machine intelligence 31, no. 1 (2009): 39-58.