International Journal of Engineering & Technology, 7 (2) (2018) 568-572

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET
doi: 10.14419/ijet.v7i2.10764
Research paper

Robust hybrid framework for automatic facial expression recognition

Gunavathi H S[1]*, Siddappa M[2]

[1]*Research Scholar, Dept. of CS&E, Jain University, Bengaluru and*
Asst. Prof., Dept of CS&E, Bangalore Institute of Technology, Karnataka, India
[2]*Dean(Academics), Prof. & Head, Dept. of CS&E, SSIT, Tumkur, Karnataka, India*
**Corresponding author E-mail:gunavathihs@gmail.com*

Abstract

Over the last few years, facial expression recognition is an active research field, which has an extensive range of applications in the area of social interaction, social intelligence, autism detection and Human-computer interaction. In this paper, a robust hybrid framework is presented to recognize the facial expressions, which enhances the efficiency and speed of recognition system by extracting significant features of a face. In the proposed framework, feature representation and extraction are done by using Local Binary Patterns (LBP) and Histogram of Oriented Gradients (HOG). Later, the dimensionalities of the obtained features are reduced using Compressive Sensing (CS) algorithm and classified using multiclass SVM classifier. We investigated the performance of the proposed hybrid framework on two public databases such as CK+ and JAFFE data sets. The investigational results show that the proposed hybrid framework is a promising framework for recognizing and identifying facial expressions with varying illuminations and poses in real time.

Keywords: *Compressive sensing; facial expressions; feature extraction; HOG and LBP.*

1. Introduction

Automatic facial expression recognition is the ability of the machine to automatically recognize expressions of emotions or expressions of social signals on faces. The emotions of a person are depicted through the facial expressions and body languages that explains the mental state of mind and psychometric aspects of a person (Fernandes et al., 2017)[1]. Emotion recognition is an estimation tool to recognize the human desires and emotions for predicting the present state of mind and future interactions.

Ekman et al. [2] described six fundamental facial expressions which are happiness, surprise, anger, sadness, fear, and disgust. The significant challenges in the facial expression recognition are head pose identification, illumination of light, age, sex, race, reducing the dimensionalities of the facial image and partially occluded images because of different posture, wearing sunglasses, beard and moustache.

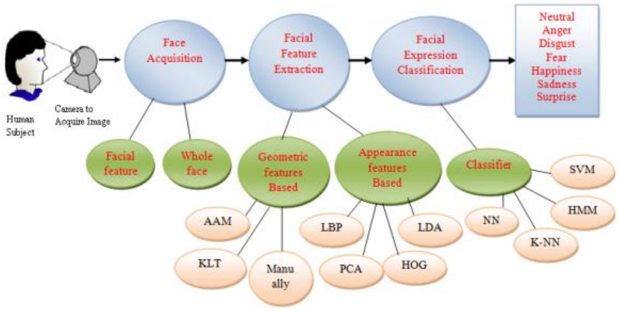


Fig. 1: General pipeline of facial expression recognition system [3].

Figure 1 explains the general pipeline of facial expression recognition system along with the most common methods in each step. The three significant phases of the facial expression recognition system are 1) Face acquisition &pre-processing 2) Facial feature extraction and 3) Facial expression classification.

The first step is to acquire the image and detect the face regions from the captured pictures or video. Once the image obtained, they are pre-processed and normalised. Later faces are identified in the photos for further processing. Face detection is done by the most widely used method such as Viola and Jones method [4].

The next step is to obtain the features from faces. Feature extraction for facial expression recognition broadly classified into three methods. 1) Geometric feature based system 2) Appearance feature based system and 3) System based on combination of the geometric and appearance based model [5]. In geometric feature based system, geometric features looks at the shape of the face, mouth, distance between eyelids, etc. There is a straight forward relation between features and distance between two facial points and the expressions. However, finding those facial points is not so simple. Choi et al. [6] explained Active Appearance Model (AAM) and its variations are used by most of the geometric feature-based models to find these facial points. After finding the facial points, we use them to obtain the shape and sizes of the essential parts of the face such as mouth, eyes, nose, and chins.

The appearance features based systems are based on the variations in the appearance of skin, furrows, wrinkles and texture changes. In this method, features related to facial appearance changes of the whole face or specific face regions are extracted by applying image filters. There are many methods proposed for this system some of them are Local Phase Quantization, Local Binary Pattern (LBP) [7], Local Directional Patterns, Local Feature Analysis (LFA), Independent Component Analysis [8], Principal Component Analysis [9], and Speeded-up Robust Features (SURF). In the combined method both the geometric and appearance features used for facial expression recognition, for example, Valstar et al. presented an AAM which combines histogram with texture and shape parameters.

In the final phase, emotion is recognized by classifying the obtained facial features. Some of the most widely used classifiers are K-Nearest Neighbour (KNN)[10], Hidden Markov Model (HMM)[11] and Support Vector Machines (SVM) [12].

The remainder of the paper structured as follows. Section 1 focuses on an introduction to facial expression recognition. A brief overview of related work in the area of facial expression recognition, local binary pattern, histogram of oriented gradients and compressive sensing are discussed in section 2.The proposed framework for facial expression recognition is presented in Section 3. Experimental datasets and results are analyzed in section 4. Finally, the conclusion of the proposed framework for facial expression recognition along with directions for future research work is given in section 5.

2. Related work

Many methods have been studied and proposed by researchers in recent years for feature extraction techniques capable of distinctively recognizing various facial expressions. In this session, we will present few of the feature extraction techniques used in our proposed approach.

2.1. Local binary patterns (LBP)

LBP is very interesting, highly successful and one of the relatively simple methods, which can be developed quickly. Due to its simplicity in nature, it is extensively used in many applications of image & video processing and pattern recognition. Ojala et al. [13] initially proposed local binary operator, and it minimizes the variations in our features that are caused by irrelevant to facial expressions such as lighting, identity and maximizes the features that relevant to facial expressions such as edges. The local binary pattern operator detects many different texture primitives, which makes it possible to analyze facial images in real time.

The beautiful thing about LBP is that it is illumination variant. If we change the lighting on the image or scene, all the pixels values in the scene go up, but the relative difference between pixels remains same. So, the binary pattern will remain same irrespective of illumination variations.

The LBP algorithm uses a 3x3-pixel block of pixels, and it's particularly interested in the central pixel, and it has eight pixels surrounded. LBP turns 3x3 pixels into a single value; it does that by first comparing its centre intensity pixel values with its every neighbouring pixel intensity value. The pixels intensity values in this block are threshold by its central pixel intensity value. The algorithm checks all the neighbourhood pixels around the centre pixel. Any pixel intensity value greater than or equal to the centre pixel intensity value will assign binary bit 1, and pixel intensity value smaller than centre pixel intensity will assign binary bit 0. Then we are going to turn these 8 bits into one byte, and it is converted to decimal value and assigned this new value to centre pixel. The above operation repeated until the whole image is converted.

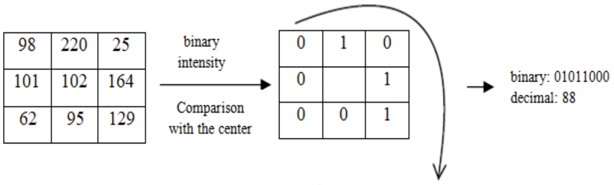
The decimal form of the generated 8-bit word (LBP code) is written as equation (1) [7].

$$\mathrm{LBP}(x_c, y_c) = \sum_{n=0}^{7} S(i_n - i_c)2^n \tag{1}$$

Where i_c represents the grey value of the centre pixel(x_c, y_c), in (n=0,…, N-1) refers to the grey values of N equally spaced pixels on a circle of Radius R(R>0), and S(x)corresponded to threshold function and defined as Equation (2).

$$S(x) = \{1 \ \text{if}\, x \geq 0 \ |0 \ \text{if}\, x < 0 \tag{2}$$

In Figure 2 Local binary operator code for centre pixel (102) is calculated by arranging outcome of the function by starting from left-top corner clockwise to get the 8-bit binary code (which is "01011000"). Then, the 8-bit binary code is converted into a decimal value (88) and is assigned as a new value to centre pixel.



98	220	25
101	102	164
62	95	129

0	1	0
0		1
0	0	1

Fig. 2: Example of LBP operator.

2.2. Histogram of oriented gradients (HOG)

HOG descriptor is an extensively used method in the domain of image processing and pattern recognition for applications such as face recognition, object recognition, image stitching, and scene recognition. David G. Lowe proposed scale-invariant feature transform (SIFT) algorithm in 1999[14], and he summarized the algorithm comprehensively in 2004[15]. Histogram of oriented gradients feature descriptor comes from the last step of Lowe's SIFT technique for image matching. It's highly helpful for detection of textured objects including distorted shapes.

In HOG we are going to divide an image into a number of small connected regions. We generate the histogram of gradient directions for all pixels in the region and the whole image. HOG descriptor represents the combination of all the histograms generated for regions in the image. The significant steps of HOG divided into three stages a) Gradient computation b) Spatial and orientation binning c) Normalization and descriptor blocks.

HOG features describe the shape information and gradient relates to the first derivative of the image. Let P_x be the gradient of X direction and let P_y be the gradient of Y direction. For an image I(x, y), gradient in a random pixel point is a vector and defined as equation (3) and magnitude ($|\nabla I(x, y)|$) as equation (4), direction angles of the gradient ($\theta(x, y)$) as equation (5) [15].

$$\nabla I(x, y) = [P_x, P_y]^T = \left[\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y}\right] \tag{3}$$

$$|\nabla I(x, y)| = \sqrt{}(P_x{}^2 + P_y{}^2) \tag{4}$$

$$\theta(x, y) = \arctan \frac{P_y}{P_x} \tag{5}$$

Also, we can represent a gradient of the image I(x, y) as equation (6)[15].

$$\begin{aligned} |\nabla I(x, y)| = \sqrt{}\{&[f(x+1, y) - f(x-1, y)]^2 \\ &+ \ [f(x, y+1) - f(x, y-1)]^2\} \end{aligned} \tag{6}$$

For feature extraction first, we divide an image into small local regions R0, R1, R2… Rm-1 and independently extract texture descriptors from each region. Then we construct a histogram with all possible labels for every region. Let n be the number of labels produced by LBP operator and m be the number of local regions in the divided image. Then we can represent a histogram (H) of labelled image d_l(x, y) as equation (7).

$$H_i = \sum_{x,y} S\{d_l(x, y) = i\}\, S\{(x, y) \in R_j\} \tag{7}$$
$$i = 0, \ldots, n-1, j = 0, \ldots, m-1$$

Where S(B) is a function and when B is true, S(B) equal to 1 and when B is false, S(B) equal to 0. Later we construct the feature vector by joining the regional histograms into one large histogram for the whole image.

2.3. Compressive sensing (CS)

According to Shannon's sampling theorem, a band-limited signal with maximum frequency “X” can be accurately reconstructed from its uniformly spaced digital sample if the rate of sampling exceeds “2X”, which is Nyquist rate and Shannon, Whitaker, Kotelnikov and Nyquist independently discovered this. Shanon's sampling theorem states that the sample needs to be uniformly spaced, sampling rate has to be very high, and original signal contains higher frequencies. Increasing sampling rate adversely affects the performance of the system. So, it is difficult to implement in real time [16]. A novel method to address these problems is proposed and named as compressed sensing (also known as compressive sampling, compressive sensing, or sparse sampling) [17]. Compressive sensing is a way of an efficiently acquiring a signal with prior knowledge that the signals of interest are sparse and rebuilding the signal by finding solutions to underdetermined linear systems.

Compressive sensing relies on two main principles. 1) Natural signals/images have a sparse representation in specific basis or transformation domain such as discrete cosine transform, Haar, wavelet domain etc. 2) the measurement matrix Φ is incoherent (poorly correlated) with the signal basis matrix ψ.

Compressive sensing can be represented as a linear algebra statement. One can find solutions to underdetermined linear system equations using the prior knowledge that the solution is sparsed.

In compressive sensing, the compressed measurements ‘y’ can be described as in equation (8).

$$y = \Phi f = \Phi\psi\theta \tag{8}$$
$$y \in R^m, \Phi \in R^{m*n}, f \in R^n, m \ll n$$

Where “f” is naturally occurring signal (input signal N*1), y is a measurement of this signal on the device, “Φ” is the measurement matrix (m * n) or the sensing matrix, “ψ” signal basis matrix/ transformation domain and “θ” is K-sparse in transformation domain. The rebuilding of the signal “f” from its measurement ‘y’ can't be solved using traditional inverse methods.

The coefficient of the signal f is Φ and hence the signal f itself is recovered by finding the solutions to the following constrained minimization problem as shown in equation (9).

$$\min||\theta||_0 \text{ such that } y = \Phi\psi\theta \tag{9}$$

We propose to use LBP and HOG with Compressive Sensing technique. These techniques have many advantages like illumination invariant, operating on local cells, allowing for contrast normalization, and being invariant to photometric transformations and geometric orientation. Because of these natural benefits we use LBP and HOG for feature extraction. After identifying the features, we reduce the dimensionality of feature space using compressive sensing which will lead to a reduction in computation cost. For classifying these features, we use multiclass SVM classifiers.

3. Proposed approach

In this paper, a robust hybrid facial expression recognition technique is presented. There are mainly three stages: 1) Image preprocessing & face detection stage 2) Feature extraction stage and 3) Classification and recognition stage. These three steps are described in detail in the present section. Our technique is shown in Figure 3, which is based on hybrid methods for feature extraction using LBP, HOG and Compressive sensing algorithms.

3.1. Image pre-processing and face detection stage

The objective of this phase is not the recognition of faces but face detection. Once we obtain the image, we pre-process the image and convert colour image to gray scale image, and we do cropping, resizing and noise reduction. After pre-processing, face detection performed by using Viola and Jones method. Characteristics like efficient features selection, fast features computations, location invariant detector, scale invariant, and robustness with excellent detection rate make it one of the best techniques for face detection. In this stage, we detect all the faces present in an image and use them for feature extraction in the next step.

3.2. Facial feature extraction stage

Facial feature extraction stage plays a vital function in facial expression recognition. In our proposed method, for feature extraction, we use integration of LBP, HOG and compressive sensing. This stage can further divide into two steps: i) Feature vectors are extracted using LBP and HOG in both testing and training phases. ii) Compressive sensing technique is used for projection of extracted dense features vectors to lower space.

LBP and HOG feature descriptor is a natural choice for feature extraction because of its several benefits. However, because of dense feature vectors, the computation cost is very high, and it limits us using it in the real-time applications. So we propose to minimize feature space dimensionality using compressive sensing, which will lead to the decrease in computation cost and increase in performance.

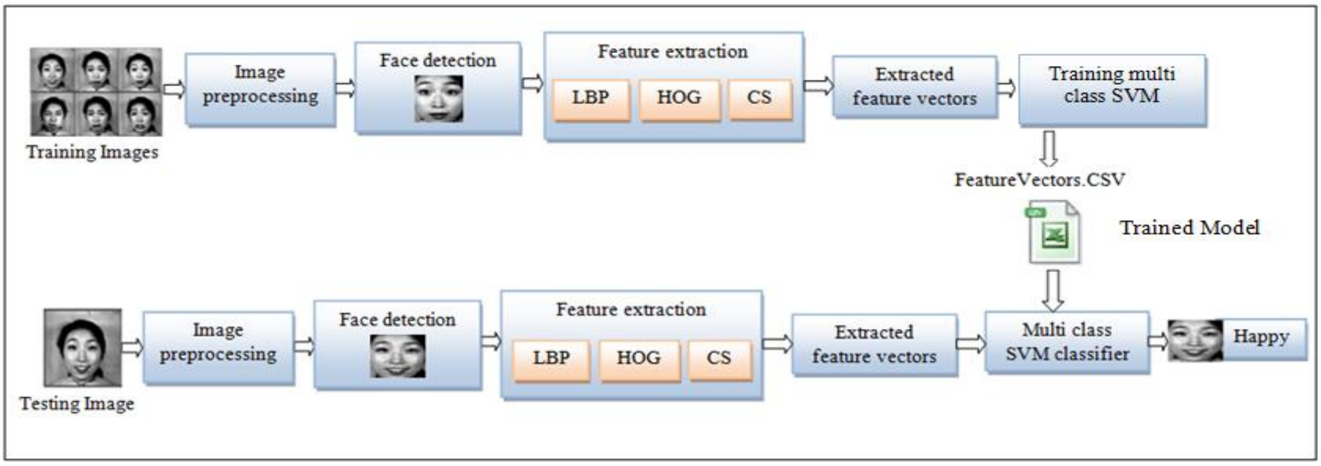


Fig. 3: Proposed framework

3.3. Classification and recognition stage

Support vector machine is a frontier which best segregates two class labels also called as a binary classifier, but several real-world applications have more than two classes (e.g. expression recognition, optical character recognition). So we use multiclass SVM classifier for facial expression recognition, as we need to classify the facial expression into seven basic emotions.

In our study, we train multiclass SVM to perform facial expression classification using the features we proposed. The original SVM is a binary classifier. However, we can take one-Vs-rest strategy to achieve the multiclass classification. In this stage, extracted feature vectors of testing and training images are compared using multiclass SVM classification algorithm, and we make a final decision based on the comparison.

4. Experimental results

We have tested our proposed technique for facial expression recognition extensively on two commonly adopted datasets such as Japanese Female Facial Expression (JAFFE) Dataset [18] and Extended Cohn-Kanade (CK+) Dataset [19].

The JAFFE database comprises 213 images of ten Japanese female models posed for seven basic emotions such as anger, fear, sad, neutral, surprise, happy and disgust. For training, we have downloaded the dataset and manually separated them into seven folders and named based on their facial expressions. For facial expression recognition, we have randomly divided the dataset into 185 training images and 28 test images. Figure 4displays a few examples from the JAFFE database.

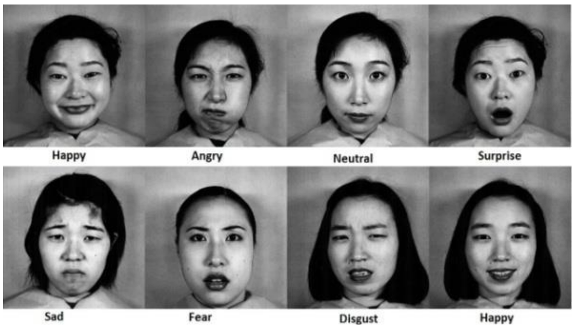


Fig. 4: Examples of JAFFE database [18].

The extended Cohn-Kanade database popularly known as CK+ database has 593 images comprising different facial expressions. We have manually separated the data sets into seven basic emotions. For facial expression recognition, we have randomly selected 56 images for testing purposes and 537 images for training purposes. Figure 5 shows some examples from the CK+ database.

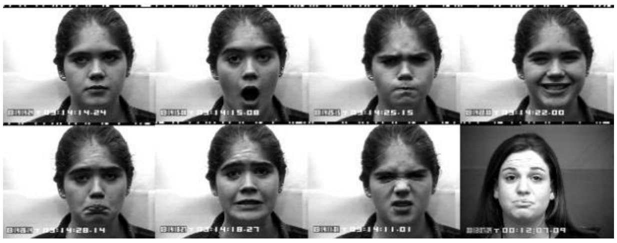


Fig.5: Examples of CK+ database [19].

The images preprocessed before extracting the features. In preprocessing stage, we convert it into a grayscale image and detect the faces using Viola and Jones method. Once the faces identified, we crop and resize it to 64 x 64 pixels. We have extracted feature vectors from training images using the proposed method and stored in a training feature vector. CSV file for further processing. The length of extracted feature vector per image is 4456. We used multiclass SVM classifier for classification.

As we discussed earlier, the equation y = Φf should be linear underdetermined system and number of rows (M) must be much smaller than the number of columns (N) in measurement matrix (Φ). We compute various values of compressed features (M) such as 10, 50, 100, 200, 500 and 800. These values are used in testing our proposed technique. The results of varying dimensions (M) are presented in Table 1for JAFFE and CK+ databases.

The test conducted on ten images from the training set, and we took the average performance for one image. In Table 1, zero means correct facial expression not found, and one means exact facial expression was detected. From the Table 1, we conclude that at least 200 feature vectors are required to recognize the facial expression correctly.

Table 1: Test Results Using JAFFE and CK+ Database

M	JAFEE		CK+	
	Expression Match	CT (Sec)	Expression Match	CT (Sec)
10	0	0.138	0	0.129
50	0	0.182	0	0.178
100	0	0.19	0	0.192
200	1	0.207	1	0.204
500	1	0.211	1	0.213
800	1	0.214	1	0.215

Figure 6 shows the comparison of execution time analysis between with and without using the compressive sensing. In case of using all feature vector from LBP+HOG and without using compressive sensing, the multiclass SVM classifier used 0.732 seconds on JAFFE and 0.698 seconds on CK+ databases, which is much higher than the classification time obtained with compressed dimension (M=200), 0.207 seconds on JAFFE and 0.204 seconds on CK+ with the same multiclass SVM classifier.

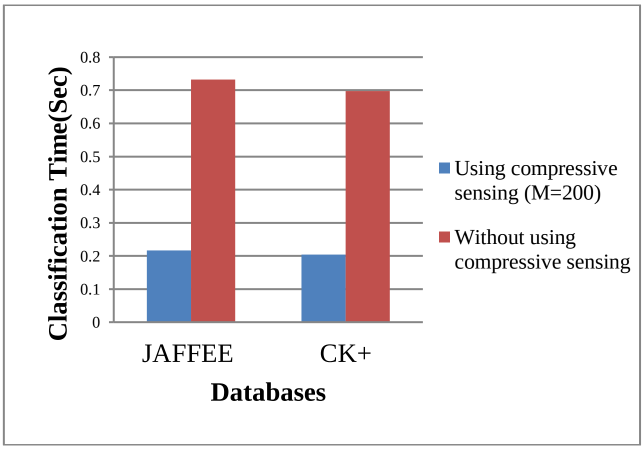


Fig. 6: Execution time analysis of proposed method.

Tables 2 and 3, shows the confusion matrix of the proposed technique on JAFFE and CK+ databases. Confusion matrix makes a more comprehensive breakdown of the results that can point out the misclassification cases and the interpretation of the possible causes.

From the confusion matrix, we can observe that the proposed hybrid approach achieved an average performance of 90% and above for JAFFE and CK+ databases. From Tables 2 and 3, we can observe that, we can obtain a decent classification rate even by reducing the feature vector using compressive sensing. It further helps in reducing the overall facial expression recognition execution time and improves the performance.

Table 2: Confusion Matrix of Facial Expression Classification on JAFFE Database

	Anger %	Fear %	Happiness %	Surprise %	Disgust %	Sadness %	Neutral %
Anger	91.8	0	0	0	5.3	2.9	0
Fear	0.9	82.6	0	0.8	8.7	7.0	0
Happiness	0	0	96.1	0.2	0	0	3.7
Surprise	2.6	0	0	90.7	4.8	2.1	0
Disgust	0.7	3.2	0	2.3	93.8	0	0
Sadness	2.1	5.6	0	0	8.4	81.7	2.2
Neutral	0	0	2.1	0	0	1.6	96.3

Table 3: Confusion Matrix of Facial Expression Classification on CK+ Database

	Anger %	Fear %	Happiness %	Surprise %	Disgust %	Sadness %	Neutral %
Anger	92.3	0	0	0	4.5	2.6	0.6
Fear	0.3	83.7	0	0	9.8	6.2	0
Happiness	0	0	96.4	0.4	0	0	3.2
Surprise	1.4	3.2	4.2	91.2	0	0	0
Disgust	1.3	0	0	0	94.3	4.4	0
Sadness	2.0	5.4	0	0	6.2	83.8	2.6
Neutral	0	0	4.7	0	0	2.2	93.1

5. Conclusion

In this paper, we presented a robust hybrid technique for facial expression recognition based on compressive sensing theory and using LBP and HOG features. LBP & HOG can efficiently extract the textures, and they are a natural choice for feature extraction because of its several advantages. However, because of large dense feature vectors, the computation cost is very high, and it limits us using it in the real-time applications. In our proposed system compressive sensing is used to reduce the dimensionality of the extracted LBP and HOG feature space which helps to reduce the computation cost. We have validated our proposed approach on JAFFE, and CK+ databases and our results illustrate an improvement in reducing computation cost, execution time and increase in accuracy. For future work, we will try to improve accuracy further by applying multimodal techniques.

References

[1] Fernandes S & Bala J, "A comparative study on various state of the art face recognition techniques under varying facial expressions", *International Arab Journal of Information Technology*, Vol. 14, No. 2, (2017), pp 254-259, available online: http://umc.edu.dz/images/A-Comparative-Study-on-Various-State-of-the-Art-Face-Recognition-Techniques-under-Varying-Facial-Expressions.pdf

[2] Ekman P, Donato G, Bartlett M, Hager J & Sejnowski T, "Classifying facial actions", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 21, No. 10, (1999), pp 974-989, available online: https://dl.acm.org/citation.cfm?id=319249

[3] Gunavathi HS & Siddappa M, "A survey of techniques for automatic facial expression recognition", *International Journal of Emerging Technology and Advanced Engineering*, Vol. 7, No.9, (2017), pp 650-655, available online: http://www.ijetae.com/files/Volume7Issue9/IJETAE_0917_96.pdf

[4] Viola P & Jones M, "Robust real-time face detection", *International Journal of Computer Vision*, Vol. 57, No. 2, (2004), pp 137-154, available online: https://link.springer.com/article/10.1023/B:VISI.0000013087.49260.fb

[5] Valstar MF, Mehu M, Jiang B, Pantic M & Scherer K, "Meta-analysis of the first facial expression recognition challenge", *IEEE Transactions on Systems, Man, and Cybernatics-B*, Vol. 42, No. 4, (2012), pp 966-979, available online: http://ieeexplore.ieee.org/ abstract/document/6222016/

[6] Choi HC, & Oh SY, "Realtime facial expression recognition using active appearance model and multilayer perceptron", *Proceeding of the International Joint Conference SICE-ICASE*, Busan, Korea, Vol. 2, (2006), pp 18-21, available online: http://ieeexplore.ieee.org/ document/4108639/

[7] Moore S & Bowden R, "Local binary patterns for multi-view facial expression recognition", *Computer Vision and Image Understanding*, Vol.115, No. 4, (2011), pp 541-558, available online: https://dl.acm.org/citation.cfm?id=1951262

[8] Bartlett MS, Moverllan JR & Sejnowski TJ, "Face recognition by independent component analysis", *IEEE Transactions on Neural Networks*, Vol. 13, No. 6, (2002), pp 1450-1464, available online:http://ieeexplore.ieee.org/document/1058079/

[9] En.wikipedia.org. (2017). Principal component analysis. Retrieved December 10, 2017, from https://en.wikipedia.org/wiki/Principal_component_analysis

[10] Yousefi S, Nguyen MP, Kehtarnavaz N & Cao Y, "Facial expression recognition based on diffeomorphic matching", *Proceedings of 17th IEEE International Conference on Image Processing (ICIP)*. Hong Kong, China, (2010), pp 4549-4552, available online: https://uthsc.pure.elsevier.com/en/publications/facial-expression-recognition-based-on-diffeomorphic-matching

[11] En.wikipedia.org. (2017). Hidden markov model. Retrieved December 10, 2017, from https://en.wikipedia.org/wiki/Hidden_markov_model

[12] Kotsia I, & Pitas I, "Facial expression recognition in image sequences using geometric deformation features and support vector machines", *IEEE Transactions on Image Processing*, Vol. 16, No. 1, (2007), pp 172-187, available online: http://www.eecs.qmul.ac.uk/~ioannisp/pubs/ecopies/kotsiatip.pdf

[13] Ojala T, Pietikainen M, & Maenpaa T, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 7, (2002), pp 971-987, available online: https://dl.acm.org/citation.cfm?id=628808

[14] Lowe DG, "Object recognition from local scale-invariant features", *Proceeding of the International Conference on Computer Vision*, Washington, USA, Vol. 2, (1999), pp 1150-1157, available online: https://dl.acm.org/citation.cfm?id=851523

[15] Lowe DG, "Distinctive omage features from scale-invariant keypoints", *International Journal of Computer Vision*, Vol.60, No. 2, (2004), pp 91-110, available online: https://www.cs.ubc.ca/ ~lowe/papers/ijcv04.pdf

[16] Baraniuk RG (2007), Compressive sensing [Lecture notes], *IEEE Signal Processing Magazin*, Vol. 24, No. 4, pp 118-121, available online: http://ieeexplore.ieee.org/abstract/document/4286571/

[17] En.wikipedia.org. (2017). Compressed sensing. Retrieved December 10, 2017, from https://en.wikipedia.org/wiki/Compressed_sensing

[18] Lyons MJ, Akamatsu S, Kamachi M, Gyoba J & Budynek J, "The japanese female facial expression (JAFFE) database", (1998), availble online: http://www.kasrl.org/jaffe.html

[19] Lucey P, Cohn JK, Kanade T, Saragih J, Ambadar Z & Matthews I, "The extended Cohn-Kanade dataset (CK+) A complete dataset for action unit and emotion-specified expression", *Proceedings of the Third International Workshop on CVPR for Human Communicative Behaviour Analysis(CVPR4HB)*, San Francisco, (2010), pp 94-101, available online: http://ieeexplore.ieee.org/document/5543262/