GU J Sci 30(3): 19-27 (2017)

Gazi University

Journal of Science

http://dergipark.gov.tr/gujs

A Study on Facial Expression Recognition

Elena BATTINI SÖNMEZ[1,*]

[1] *Department of Computer Engineering, Assistant Professor, Istanbul Bilgi University, Istanbul, Turkey*

Article Info

Received: 17/04/2017
Accepted: 27/07/2017

Keywords

Little databases
facial expression recognition
sparse representation based classifier
convolutional neural network

Abstract

This study focuses on the issue of automatic facial expression recognition on little databases of 2D faces. Convolutional Neural Networks (CNN) is a new classification technique, which reaches the state of the art on big databases; however, the use of CNN with a scarce number of samples is still an open challenge. Following the classical machine learning approach, we considered different combination of feature extraction and classifiers, and we compared their performances with special designed CNN. Our results show that CNN outperforms the other classifiers in the “close system” experiment; however, in the more challenging “open system” experimental setup the Sparse Representation based Classifier is more successful.

1. INTRODUCTION

Automatic Facial Expression Recognition (FER) is a popular area of research in computer vision due to its great number of potential application, ranging from human-computer interaction, synthetic face animation, image analysis and understanding. Moreover, the particular nature of the problem makes this research field interesting also for neuroscientists and psychologists. Available surveys on this topic are [1] and [2].

Considering the classical machine learning approach, an automatic FER system is made up of the following three steps: (1) face detection and alignment, (2) facial feature extraction, and (3) classification. Face detection is an active research field, particularly hard with real word images having several faces and a large number of details. The goal of the feature extraction step is to find compact and robust representation of the original face. Currently, appearance based approach is the one giving the most promising results, and, because it does not require any accurate and reliable face landmark detection, it is also easier to implement [3]. Another common approach is to use Gabor feature [4] to represent the facial emotion; this technique is one of the most successful, but it has high complexity, which is generally handled by applying the Gabor filter to a set of feature points. In 2002, Ojala [5] used the Local Binary Patterns (LBP) technique to extract robust and discriminative features suitable for face analysis, due to its low complexity and high discriminative power.

After projection, classification is performed using the Nearest Neighbor (NN) [6], Nearest Subspace [7], Support Vector Machine (SVM) [6], or the Sparse Representation based Classifier (SRC) [8].

Despite all contributions, the traditional approach of machine learning still fails to reach human like performance when dealing with problems which are hard to formalize, such as object detection, scene identification, natural language processing, etc. Recently, deep learning theory [9] proposed alternative approaches to challenge those issues. To date, with images, Convolutional Neural Networks (CNN) gives the state-of-the-art performance in many fields; however, CNN requires the availability of huge amount of data, and the usage of CNN with little databases is still an interesting open challenge.

We restricted our attention on the Japanese Female Facial Expression (Jaffe) [10] database, which is a little collection of emotional faces, widely used in the research community, and, thus allowing comparison with

*Corresponding author, e-mail: ebsonmez@bilgi.edu.tr

other algorithms' performance. In the following we consider the available literature on Jaffe: In 2005, Deng et al. [11] introduced a local Gabor filter bank, with the aim to decrease the complexity of the original Gabor bank and compared its performance against Gabor features compressed with Principal Component Analysis (PCA) [6], and PCA plus Linear Discriminative Analysis (LDA) [6] on selected pictures of the Jaffe database. In 2008, [12] Bashyal et al. used Gabor filters in combination with Learning Vector Quantization (LVQ) for recognition of the seven expressions on selected images of the Jaffe database. In 2009, [13] Shan et al. introduced an improved LBP feature called 'boosted LBP', and used it together with SVM on the Leave-One-Subject-Out (LOSO) experiment on the Jaffe database. In 2009, Zilu et al. [14] coupled Non-negative Matrix Factorization (NMF) with SVM to run the 7-class LOSO experiment on Jaffe. In 2010, Huang et al. [15] used the LBP feature with SRC, and compared its performance against PCA, LDA, and Gabor histogram on the LOSO experiment of the Jaffe database, on a special selection of pictures. In 2011, Zavaschi et al. [16] compared the performance of LBP and Gabor features when coupled together with the SVM classifier on the FER experiments of the Jaffe database. In 2012, Zhao et al. [17] introduced a new kernel-based supervised manifold learning algorithm, called Discriminative Kernel Locally Linear Embedding (DKLLE), and compared its performance against LDA and PCA. In 2014, Liu et al. [18] challenged the LOSO experiment of Jaffe with a Boosted Deep Belief Network (BDBN).

In this paper, we compare the most popular projection, feature extraction and classification algorithms on FER with special designed CNN, so as to rank it in the list of successful classifiers for little databases. Main contributions of this work are:

- Review of the most common algorithms used for FER
- Introduction of special tricks necessary to use CNN with little databases
- Fair comparison among all algorithms

Section 2 overviews the most common algorithms used for FER. Section 3 describes the database, the experimental setups and the obtained results. Finally, conclusions are drawn in Section 4.

2. ALGORITHMS

In this section, we give a brief description of all projection methods, feature extraction techniques and classifiers normally used for FER, with emphasis on the Sparse Representation based Classifier and Convolutional Neural Networks, as they are both successful and recently introduced classification algorithms.

2.1. Subspace Analysis and Feature Extraction Methods

Principal Component Analysis (PCA): Principal Component Analysis is a dimensionality reduction method which seeks the projection that best represent the data in a least-square sense. Starting from N dimensional data, the PCA subspace is built by aligning the biggest K-eigenvectors of its covariance matrix, which are the K directions of maximum spread of the data, $K<N$. PCA subspace may not have discriminative power as it is a reconstructive method robust to noise.

Linear Discriminant Analysis (LDA): Linear Discriminant Analysis is a dimensionality reduction method, which seeks the projection that best separate the data in the least-square sense; by maximizing the between-class scatter and minimizing the within-class scatter, LDA finds a discriminative subspace which allows for successful classification.

Gabor filters: Gabor filters remove most of the variability of the image and it is robust against small shifts and deformations. Practically, the Gabor feature of every peak face is obtained by convolving the peak image with the Gabor filter bank and considering the magnitude of the Gabor wavelets representation only for particular points.

Local Binary Patterns (LBP): The original LBP operator was introduced by Ojala et al. as a powerful texture descriptor in [5]; LBP labels the pixel of an image by (1) considering a 3×3 neighborhood, (2) thresholding every neighbor pixel with the center value, and (3) converting the resulting binary string into a decimal number assigned to the center pixel. Figure 1 details these three steps:

50	40	50
0	**100**	10
120	140	250

0	0	0
0		0
1	1	1

Figure 1. *Local binary pattern operator with a 3 × 3 rectangular grid: (left) original grey-level values of an 8-bit image, the center pixel is in bold; (right) the resulting binary string, (00001110)2, corresponding to LBP label 14*

The LBP feature is the histogram of all labels; the size of the LBP feature depends to the size of the neighborhood, in the example of Figure 1 we have a total of 8 neighbors generating $2^8 = 256$ possible labels. The concept of uniform patters allows for decreasing the size of the LBP feature. In our experiments, we applied the LBP feature extraction technique to the cropped and aligned faces of Jaffe, more in details: (1) we divided the original picture of size 130×125 into 5 × 5 = 25 blocks obtaining chunks of size 26 × 26 pixels; (2) out of every block, we binned the LBP labels into 10 equally spaced containers; and (3) we concatenated the histograms of the 25 blocks obtaining a LBP feature of length 25 × 10 = 250 doubles.

2.2. Classifiers

Nearest Subspace (NS): The Nearest Subspace classifier assigns a test sample to the nearby class; that is, having a training set divided into C classes, it calculates the distance between the test samples and the subspace spanned by every class i, for all i=1, …,C, and assigns the test sample to the nearby subspace.

Sparse Representation based Classifier (SRC): In 2009, Wright et al. [8] proposed the Sparse Representation based Classifier (SRC), which casts the classification problem as a sparse representation issue. Practically, SRC builds a dictionary whose base elements consist of training samples themselves, and searches for a parsimonious representation of the target object in terms of these samples. More in details, a dictionary, D is a collection of parameterized waveforms where each waveform, di, is a discrete time signal of length N called atom and it has unit length, $||d_i||=1$. Dictionaries are complete, if they contain exactly N linearly independent atoms, and over-complete, if they contain more than N atoms; D is an over-complete dictionary. The use of over-complete dictionaries allows a sparse representation because we can decompose the target signal in more than one ways; moreover, the non-unique representation of the observed signal gives the possibility of adaptation, the potential of choosing among many representations the one which is most suited to our purpose. Some of the goals are sparsity, discriminative power and robustness. Following the standard notation, the matrix format of the general reconstruction or projection step is:

$$y = D \cdot x \tag{2.2.1}$$

Where $D \in R^{(N\times M)}$ is the dictionary, $y \in R^N$ the observed signal, and $x \in R^M$ is the coefficient vector to be determined; that is, SRC reads equation (2.2.1) from the synthesis point of view; when $N < M$, D is a flat matrix having more columns than rows, equation (2.2.1) has infinitely many solutions at every point; among all possible solutions, we are interested in the one that minimizes the error and has the minimum number of non-zero elements. If $N > M$, matrix D has more rows than column and equation (2.2.1) has generally no solution; in this case we are interested in minimize the error in the approximation. In both case, we want to minimize the following quantity:

$$||y-D\cdot x||_2+\gamma \, ||x||_0 \tag{2.2.2}$$

where $||x||_0$ is the L0 norm, it is simply the count of non-zero elements.

The theory of Compressive Sensing [19] proved that, if the signal is sparse enough, the sparsest linear representation of the test sample, the solution x to eq. (2.2.1), can be recovered efficiently via L1 minimization. Having the solution x, SRC calculates the distance between the current test sample y and all classes and it assigns y to the nearby class.

Notice the fundamental difference of SRC from the NS method, since SRC has a first global stage, where it uses the entire training set to solve for x, and a second local one, where it considers one class at a time and it uses the sparse coefficients x as weights to calculate the distance between the test sample and every class. A detailed description of all studies done on SRC is presented in [20].

In case of emotion classification, classes are emotion; when working with the Jaffe database, we considered the 6 universal expression plus the neutral face, and the dictionary D is divided into 7 classes.

S*upport Vector Machine (SVM):* SVM is a binary classifier, a supervised learning model which attempts to find a linear function to separate all samples of the two classes by a clear gap which is required to be as wide as possible. In other words, SVM demands the score of the correct class to be higher than all other scores by at least a margin delta. Among all possible weights satisfying the above condition, SVM chooses the set of weights which are little in the L2 norm.

Convolutional Neural Networks (CNN): Convolutional Neural Networks (CNN) are a special type of Neural Networks for processing 2D data [9]. Comparing with Neural Networks, the main difference is that CNN uses convolution, instead of general matrix multiplication, in at least one of its layer. Informally, convolution is a mathematical way of combining two signals, or functions, to produce a third one; the first argument to the convolution is often referred to as the input, the second argument is the kernel, and the output function is the future map. In the field of digital image processing, the input function is the greylevel of the image, it is the raw pixels of the image to be processed; the kernel is also called filter or mark, it is a small matrix of learnable coefficients which circumscribe a neighborhood. The operation of filtering produces a new pixel in the same position of the processed one with the greylevel value equal to the weighted sum of the neighborhood pixels.

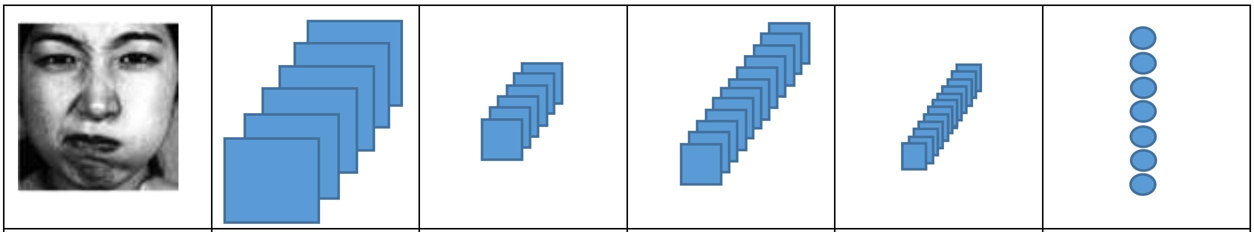
The input to a CNN is a batch of images, and a convolutional layer uses more than one filter. During the forward pass, each filter is slided across the width and height of the input images, producing an activation map. Since convolving the filter across the input is equivalent to computing the dot product between the input function and the filter, the net is looking for filters that activate in presence of specific feature of the input. The output of a convolutional level is the ensemble of all activation maps for all filters; every pixel of the output volume can also be interpreted as the response of a neuron that looks at only one small region of the input and shares parameters with neurons in the same activation map. That is, due to the filtering operation, CNN has 2 major advantages:

- sparse interaction: which is obtained by using little kernels, and stride
- parameter sharing: which refers to the use of the same kernel for more than one functions

As a result, convolution is much more efficient than matrix multiplication.

Generally, CNN has an input layer, a sequences of convolutional layers, followed by one or more fully connected layers, and one output layer. Convolutional layers may be coupled with sub-sampling layers. The initial coefficients of a CNN are fixed as random values and they are updated using the back-propagation algorithm. A good introduction to CNN is given by [6] [9].

The used CNN is made up of 5 levels: 1 input layer, 2 convolutional and subsampling layers, and one output layer. The 1st convolutional layer uses 6 random filters of size 5×5, and the zero padding technique. When the input is the entire image of size 128×124 this results in 6 feature maps of size 124×120. Subsampling is performed by smoothing every feature map with an averaging filter and down sampling by 2 in both directions. The 2nd convolutional layer uses 12 random filters of size 5×5, and the zero padding technique resulting in 12 feature maps of size 58×56. The 2nd subsampling level produces images of size 29×28=812 neurons, which are fully connected with the 7 output neurons. Figure 2 summarizes the described CNN:



The input image of size 128×124	The 1st convolutional layer produces 6 features maps of size 124×120	The 1st sub-sampling layer produces 6 feature maps of size 62×60	The 2nd convolutional layer produces 12 features maps of size 58×56	The 2nd sub sampling level produces 12 feature maps of size 29×28	The output layer with 7 neurons

Figure 2. *The architecture of the used CNN.*

Despite the exceptional performance obtained, recently, by CNN in big and challenging databases, the use of CNN with a scares amount of labeled data is not taken for granted. In this paper, we use some tricks, which allow to use CNN with little databases. The first trick is to create class-balanced batches, which is batches having a uniform distribution over all classes. That is, due to the scares number of samples per class, the random creation of batches risks to produces clusters with one or more missed classes. To avoid this problem we impose the construction of batches having a uniform distribution over all classes. Another important precaution is to work on discriminative blocks of the face; in case of emotional faces, the block of the mouth. Moreover, the original number of samples is triplicated using data augmentation techniques, and, finally, overfitting is tackled by using 2 subsampling layers. A detailed description of these tricks is available in [21].

3. EXPERIMENTS

We worked with the Japanese Female Facial Expression (Jaffe) databases; as it is a little, well know database, widely used by the research community.

3.1. The Japanese Female Facial Expression (Jaffe) database

The Japanese Female Facial Expression (Jaffe) database [10] contains 213 images of 10 Japanese female models; every subject posed 3 or 4 samples of each of the 7 facial expressions: the 6 basic emotions of happiness, sadness, surprise, anger, disgust, fear plus the neutral one; more in details, the distribution of the 7 classes is fear (32), happy (31), disgust (29), angry (30), sad (31), surprise (30) and neutral (30). The label of every image stores information of the acting subject and the dominant emotion. Each image in the database was rated by 60 Japanese female subjects on a 5 level scale, and their average asses the degree of the six basic expressions present in the face. We used all images of the database with the given label, even if there are some cases where the label of the image is not the dominant emotion, and the face does not really express the labeled emotion; just to give an example, picture number 89 was rated: happy = 1.25, sad = 2.26, surprise = 4.45, angry = 3.16, disgust = 3.03 ad fear = 2.90. The label assigned to picture 89 is “KR-FE1” corresponding to the “fear” emotion even if the dominant expression is “surprise”; moreover, as it can be observed in the following picture, the concomitant mixture of different expressions does not produce a fear face:

Figure 3. *Picture no. 89*

Since we did not make any selection of pictures, some error is expected and justified.

3.2. Pre-processing

Jaffe is a little collection of emotional faces widely used in the literature with manually pointed faces landmarks (FL); we repeated the two classical experiments run on this database, but we used automatically detected fiducial points. Figure 4 shows the normalization steps performed on Jaffe: having aligned and cropped the faces, we imposed fix inter-ocular distance and photometric normalization.

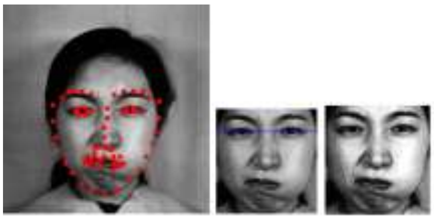


Figure 4. *(left) the original image with automatically detected face landmarks, (center) the cropped and aligned face, (right) the zoomed and photometric normalized image*

Finally, we decimated all faces to a common size, and we extracted the block of the mouth, which is one of the most discriminative area of the face [21].

3.3. Experimental Setup and Results

We repeated the 2 classical experiments run on Jaffe: the 10-fold cross-validation and the leave-one-subject-out (LOSO) experiment; in the first experiment, having 213 images, we randomly assigned 21 faces in the first 7 folders, and 22 pictures to the last 3 folders; due to the random selection, we repeated the experiment 10 times, and the given performance is the average of the resulting 100 trials. More in details, at every trial, the total number of training samples is 213-21=192 (or 193, when the test set is one of the last 3 folders); by adding Gaussian 7 and 14, we triplicated the samples, and we randomly selected 500 of them to be used for training.

Because the database has 10 subjects, in both cases the pictures are divided into 10 subsets and the experiment has 10 trials; at every run all pictures of one subset are used for testing and the training set is made up of all pictures of the remaining subsets. However, while in the first experiment, 10-fold cross validation, we have a “close system”, where the subject acting in the test samples may be present in the training set; in the second experiment, LOSO, we have an “open system”, where the subject acting the test faces is not present in the training set.

Table 1 compares the performances of the close system experiment run by different authors. We did not report the results of papers making special selection of pictures. From the 2nd row of Table 1 we may say that, when applied to the entire face, SVM is still the most successful classifier, reaching the top performance of 91.6%. However, in row 7, we see that CNN exceed SVM and reaches the top accuracy of 94.23%, when it is applied to the block of the mouth, which is a high discriminative area of an emotional face [21]. In row 8, the combination of raw pixel and SRC, positions SRC in the 3th place, with a performance of 90%.

Table 1. *Performance comparison of the 10-fold cross validation experiment*

Row no	Article [ref]	Feature	Classification	Perf (%)	Comments
1	Shan [13]	LBP	SVM	81.0	size(face)=110×150
2	Zavaschi [16]	Gabor	SVM	91.6	Best performance with 1 classifier on faces
3	Zhao [17]	DKLLE	NN	84.05	size(face)=110×150
4		LDA	NN	80.81	size(face)=110×150

5		PCA	NN	78.09	size(face)=110×150
6	Proposed study	CNN		24.08	Size(face)= 128×124 Balanced batches
7	Proposed study	CNN		94.23	Size(mouth)= 44 × 64 Balanced batches
8	Proposed study	Raw pixels	SRC	90	Size(face)= 130 × 125
9	Proposed study	Block-based LBP	SRC	83	Size(face)= 130 × 125

More in details, the top performance of row 7 is reached by using a block-based CNN, and by imposing the construction of balanced batches. That is, when working with the entire face the initial performance of 14.85% increases to 24.08% after the construction of class-balanced batches; with the block of the mouth, the initial performance of 91.23% reaches 94.23% with balanced batches. In all experiments, the batch size is equal to 25, and the number of epochs is equal to 100.

Table 2 compares the performances on the open system experiment run by several authors. We did not report the results of papers making special selection of pictures. The top performance is reached, in row 5, by a boosted DBN working in a block-based fashion, and running, alternatively, strong classifiers and weak learner until they converge. Considering only one classifier, SRC reaches the top performance, in row 8, when coupled with raw pixels.

Table 2. *Performance comparison of the LOSO experiment*

Row no	**Article [ref]**	**Feature**	**Classifier**	**Perf (%)**	**Comments**
1	Zilu [14]	NMF	SVM	66.2	size(face) = 32×32
2		PCA	SVM	53.8	size(face) = 32×32
3		LDA	SVM	55.7	size(face) = 32×32
4	Zavaschi [16]	LBP	SVM	60.6	Best performance with 1 classifier on faces
5	Liu [18]	Boosted DBN		91.8	Blocks of faces
6	Proposed study	CNN		17.06	size(face) = 120×112 Balanced batches
7	Proposed study	CNN		58.06	Size(mouth) = 24×40 Balanced batches
8	Proposed study	Raw pixels	SRC	68.08	Size(face)=64×64
9	Proposed study	LBP	SRC	52	Size(face)=64×64

4. CONCLUSION AND FUTURE WORK

We worked on automatic FER; we challenged two common experiments run on a little, widely used database of emotional faces. The limited number of samples requires the use of special designed CNN. Overall, the close system experiment is more successful than the open system one, and this is expected, since in the 10-fold cross validation setup there is high probability to find a training sample similar to the test one, that is, a sample in which the same subject is acting the same emotion. Despite all normalization, it seems that classification results are still too much affected by disturbance elements, such as identity of the acting subject, distance from the camera and presence of light. Special designed CNN reaches the top performance of 94.23% in the close-system experiment. However, in the more challenging open system experiment, the classical machine learning approaches are still more desirable.

Future work includes the theoretic and empirical study of special designed CNN, so as to make them robust to scarce amount of data, as well as studies on data augmentation techniques, and automatic identification of the most discriminative regions of the image.

CONFLICT OF INTEREST

No conflict of interest was declared by the authors

REFERENCES

[1] Caleanu, C., “Face expression recognition: A brief overview of the last decade”, IEEE Eighth International Symposium on applied Computational Intelligence and Informatics (SACI), Timisoara, Romania, 157-161, (2013).

[2] Sariyanidi, E., Gunes, H., Cavallaro, A., “Automatic analysis of facial affect: A survey of registration, representation, and recognition”, Journal of IEEE Transactions on Pattern Analysis and Machine Intelligence 37(6), 1113-1133, (2015).

[3] Ashraf, A.B., Lucey, S., Cohn, J.F., Chen, T., Ambadar, Z., Prkachin, K.M. and Solomon, P.E., “The painful face - pain expression recognition using active appearance models”, Image and Vision Computing, 27(12), 1788–1796, (2009).

[4] Daugman, J.G., “Complete discrete 2-d gabor transforms by neural networks for image analysis and compression”, IEEE Trans. Audio Speech., 36, 1169–1179, (1988).

[5] Ojala, T., Pietikinen, M. and Harwook, D., “A Comparative Study of texture measures with classification based on feature distribution”, Pattern Recognition Journal, 29(1), 55-59, (1996).

[6] Duda, R, Hart, P. and Stork, D., Pattern Classification, 2nd Edition, John Wiley, New York, (2001).

[7] Lee, K.C., Ho, J., Kriegman, D.J., “Acquiring linear subspaces for face recognition under variable lighting”, IEEE Trans. Pattern Anal. Mach. Intell., 27, 684–698, (2005).

[8] Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S. and Ma, Y., “Robust face recognition via sparse representation”, Transactions on Pattern Analysis and Machine Intelligence, 31(2), 210–227, (2009).

[9] LeCun, Y., Bengio, Y. and Hinton, G., “Deep learning”, Journal of Nature, 521, 436-444, (2015).

[10] Lyons, Akamatsu, M.S., Kamachi, M. and Gyoba, J., “Coding facial expressions with Gabor wavelets”, IEEE Int. Conf. on Autom. Face and Gesture Recognition, Nara, Japan, 200–205, 1998.

[11] Deng, H.B., Jin, L.W., Zhen, L.X. and Huang, J.C, “A New Facial Expression Recognition Method Based on Local Gabor Filter Bank and PCA plus LDA”, International Journal of Information Technology, 11, 86-96, (2005).

[12] Bashyal S., and Venayagamoorthy, G.K., “Recognition of Facial Expressions using Gabor wavelets and learning vector quantization”, Eng. Applications of Artificial Intelligence, 21(7), 1056-1064, (2008).

[13] Shan, C., Gong, S. and McOwan, P. W., “Facial expression recognition based on local binary patterns: A comprehensive study”, Image and Vision Computing, 27, 803–816, (2009).

[14] Zilu, Y. and Guoyi, Z., "Facial Expression Recognition Based on NMF and SVM", International Forum on Information Technology and Applications, Chengdu, China, 612-615, (2009).

[15] Huang, M.W., Wang, Z. and Ying, Z.L., “A New Method for Facial Expression Recognition based on Sparse Representation plus LBP”, 3th Int. Congress on Image and Signal Processing (CISP), Yantai, China, (2010).

[16] Zavaschi, T.H.H, Koerich, A.L. and Oliveira, L.E.S., “Facial Expression Recognition using ensemble of Classifiers”, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Prague, Czech Republic, 1489-1492, (2011).

[17] Zhao, X. and Zhang, S., “Facial expression recognition using local binary patterns and discriminant kernel locally linear embedding”, EURASIP Journal on Advances in Signal Processing, 20, DOI: 10.1186/1687-6180-2012-20, (2012).

[18] Liu, P., Han, S., Meng, Z. and Tong, Y., "Facial Expression Recognition via a Boosted Deep Belief Network", IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, 1805-1812, doi: 10.1109/CVPR.2014.233, (2014).

[19] Donoho, D.L., “Compressed sensing”, IEEE Trans. Inform. Theor., 52, 1289–1306, (2006).

[20] Battini Sönmez, E., “Robust Classification based on Sparsity”, Lambert Academic Publishing, Germany, (2013).

[21] Battini Sönmez, E. and Cangelosi, A., “Convolutional neural networks with balanced batches for facial expressions recognition”, 9th International Conference on Machine Vision (ICMV), Nice, France, (2016).