Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2017) 25: 2696 – 2707

doi:10.3906/elk-1607-113

Late fusion of facial dynamics for automatic expression recognition

Alessandra BANDRABUR[1,2,*], **Laura FLOREA**[1], **Cornel FLOREA**[1], **Matei MANCAS**[2]
[1]University "Politehnica" of Bucharest, Bucharest, Romania
[2]IT Department, Faculty of Engineering, University of Mons, Mons, Belgium

Received: 10.07.2016 • **Accepted/Published Online:** 13.10.2016 • **Final Version:** 30.07.2017

Abstract: Installment of a facial expression is associated with contractions and extensions of specific facial muscles. Noting that expression is about changes, we present a model for expression classification based on facial landmarks dynamics. Our model isolates the trajectory of facial fiducial points by wrapping them up in relevant features and discriminating among various alternatives with a machine learning classification system. The used features are geometric and temporal-based and the classification system is represented by a late fusion framework that combines several neural networks with binary responses. The proposed method is robust, being able to handle complex expression classes.

Key words: Feature extraction, machine learning, facial expression recognition

1. Introduction

Human–computer interaction has gained significant momentum. The human face is a powerful means of communication and facial expressions disseminate important cues in human interactions. Facial expressions and body movements were called by Pentland and Heibeck "honest signals" [1], as they can measure the quality of an interaction between humans by including information about the motivation, intention, and emotion of the subject.

Automatic facial expression recognition is the fundamental for multiple applications in various domains such as security, computer science, education, automotives [2], crime investigation [3], interactive gaming [4], health support appliances, research of depression [5], or pain detection [6]. However, the current state of the art indicates that only particular solutions have reached maturity and further advance is still required.

The most commonly used way to describe facial expressions is the Facial Action Coding System (FACS) proposed by Ekman et al. [7]. It is an anatomically based system that measures the facial muscle movements in terms of action units (AUs). Each AU is dynamic and has three phases: onset, apex, and offset. Because of this behavior, the systems for facial expression recognition analyze sequences of images containing the neutral face and the expression apex [8–10]. The temporal dynamics have a crucial role for categorization of psychological states like pain or shame [11]. Among applications one may include differentiation between posed and spontaneous facial expressions [12].

According to Ekman's theory [13], there are facial expressions that can be recognized in every culture and are different and distinguishable. Their automatic recognition is challenging since expressions can be categorized in macro, subtle, and micro expressions. For reviews on state-of-the-art systems for facial expression analysis we refer the reader to the works of Zeng et al. [14] and Cohn and De La Torre [15].

*Correspondence: abandrabur@imag.pub.ro

In this paper, we propose an automatic system for facial expression recognition. The system is built on the classical pattern recognition paradigm. Features are extracted from gray-scale face images and later discrimination among various inputs is achieved using a classification system.

This paper is organized as follows: Section 2 provides a brief summary of the state of the art for facial expression recognition, Section 3 provides a brief description of the used database, Section 4 describes the proposed method, Section 5 presents the achieved performance and discusses implications of the experimental results, and the paper is concluded in the last section.

2. Related work in facial expression recognition

The challenge of facial expression recognition systems is to classify a portrait image in a class of discrete facial expressions associated with certain emotions. Multiple solutions acknowledge the dynamic aspect of expression and employ pattern recognition techniques using sequences of images. The systems for expression recognition start by detecting the face and computing a set of facial features. The classification step either classifies the facial expressions into a number of discrete emotions or all the AUs detected, followed by a mapping into emotions. The latter approach is arguable as there does not exist any widely accepted mapping between AUs and emotions and for specific cases the consensus of specialists forms the ground truth.

Cohn and De La Torre [15] classified the extracted features in geometric, appearance, or motion-based. Geometric features are related to facial landmarks. The landmarks are named facial fiducial points and can evolve independently or can be connected in a mesh. Appearance features rely on changes in texture. Motion features use the dynamics of the expressions and include optical flow, volume local binary patterns, etc.

2.1. Geometric-based features

The solutions falling into this approach require a first step of accurate facial landmark localization. Geometric features extracted from tracking fiducial points were used by Valstar et al. [8], Pantic and Patras [16], etc. The proposed feature descriptor is related to the one introduced by Valstar et al. [8] as further discussed. The former consists of facial landmark geometric positions, while the latter is formed by the temporal aspect of these points. In addition, we use a late fusion schema involving two descriptors. Our previous work [17] exploits the dynamics of the features and uses a MLP to select the appropriate facial expression; this work is an extension, as it differs by using the scale space for feature computation and replaces the simple classifier with an ensemble built upon the late fusion paradigm.

2.2. Appearance-based features

Tian [18] and Littlewort et al. [19] relied on Gabor wavelets to extract appearance information. Local binary patterns (LBPs) with different extensions were also used: Jiang et al. [20] compared LBPs and local phase quantization for action unit analysis, while Zhao and Pietikainen used volume LBPs [21]. The texture of the face was encoded in local directional number patterns by Rivera et al. [22]. Rudovic et al. [23] introduced a model topology of the face by a low-dimensional manifold that preserves discriminative information about facial expressions.

2.3. Hybrid features

Because each category of features has advantages and disadvantages, researchers commonly rely on hybrid methods for expression recognition. Youssif and Asker [24] fed 19 geometric features and 64 appearance features

to a radial basis function neural network in order to classify six facial expressions. Yi et al. [25] used feature points extracted by an active appearance model and extracted geometric and texture information based on the relative positions of those points. Shbib and Zhou [26] used features extracted on an active shape model and fed to a SVM to analyze the facial expressions in a one-against-one training-testing procedure.

2.4. Machine learning-oriented solutions

The late advance of machine learning allows for the derivation of solutions that replace traditional blocks with systems that have the advantage of learning adapting to the task specificity. Restricted Boltzmann machines were used by Liu et al. [27] to learn hierarchical features in order to obtain expression recognition. Zhong et al. [28] used two-stage multitask sparse learning to analyze patches on the face in order to discriminate between different expressions.

Although initially there was debate about the work-flow of achieving emotions from AUs, since the Cohn-Kanade+ (CK+) database [29] was released, there has been a general agreement about their associations. Lucey et al. [30] provided three methods for the CK+ database. The one that gives the best results uses the normalized positions of 68 facial landmarks and the canonical normalized appearance merged into an active appearance model.

3. CK+ database

The CK+ [29,30] contains 593 sequences from 123 subjects performing one of the seven discrete emotions: happiness, sadness, surprise, fear, contempt, disgust, and anger. The database provides expression labels, AUs, and the positions of 68 landmarks. Each sequence starts with a neutral expression and ends with an apex one. The apex is FACS-coded and is annotated with one of the seven emotions by gathering a consensus from multiple experienced observers, who determine whether the expression is a good representation of the specific emotion.

4. The proposed algorithm

The proposed system has the classical structure of a feature-based one: face detection, feature extraction, and facial expression classification. The classification stage consists of a late fusion scheme using a multilayer perceptron and support vector machines. Like most state-of-the-art algorithms in facial expression recognition, due to the dynamic behavior of facial expressions, our system requires for analysis sequences of gray-level images containing the neutral face and the expression apex. The algorithm schematic can be seen in Figure 1 and the facial expression classification schematics in Figure 2.

Figure 1. The schematic of the proposed algorithm.

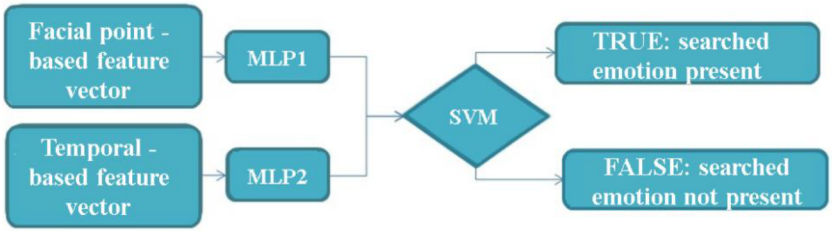

Figure 2. The facial expression classification schema into one of the seven emotion categories.

4.1. Face detection and preprocessing

The face is detected in a frame using the classical Viola and Jones method [31], as implemented in OpenCV. On each frame of the sequence, the fiducial landmarks are localized. The faces are normalized. Further landmark normalization is achieved by means of an affine transformation given by the chosen landmarks' positions.

For landmark normalization we apply an affine transform identified by a linear combination of translation, rotation, and scaling. The transform is determined by requiring a subset of features to be in fixed positions. We chose the subset by selecting those landmarks that are not influenced by muscle changes: nose points and the inner corner of the eyes. The face is rotated so as to have the line given by the inner corners of the eyes horizontal. We scale the faces with respect to the nose length, which is set to 40 pixels. We translate faces so as to center them at the nose tip. To extract the features, we switch to a scale space and follow with actual computation. We do a postprocessing operation, which consists of a Gaussian filter applied on all fiducial points' coordinates in the image sequence to get rid of the error accumulation.

4.2. Feature extraction

4.2.1. Scale space

In the scale space [32], the original function space $x(t)$ is replaced by the scale space of a function $X(t;\sigma)$:

$$X(t;\sigma) = G(t;\sigma) * x(t), \tag{1}$$

where $*$ stands for convolution and $G(x;\sigma)$ is a Gaussian rotationally symmetric kernel with variance σ^2 (the scale parameter):

$$G(t;\sigma) = \frac{1}{2\sigma}\exp\left(\frac{t^2}{2\sigma^2}\right). \tag{2}$$

Switching to scale space not only decreases the noise of the input function but also present calculus advantages. Here, the differentiation is computed by a convolution with the derivative of the Gaussian kernel:

$$\frac{\partial}{\partial t}X(t;\sigma) = \left(\frac{\partial}{\partial t}G(t;\sigma)\right) * x(t). \tag{3}$$

4.2.2. Feature computation

This work draws its inspiration from the works of Valstar and Pantic [8] and continues in the same line as our previously proposed method for facial expression recognition [17]. The major difference from the previous work, with respect to feature computation, is the introduction of the calculus in the scale space. We choose a feature descriptor, which represents the shape modifications of the face due to the muscle contractions. This descriptor is defined by four pairs of basic geometric-feature characteristics.

The first pair is very basic and consists only of the coordinates of each face landmark:

$$\begin{aligned} f_1\left(p_i\left(t\right)\right) &= X_i\left(t\right); \\ f_2\left(p_i\left(t\right)\right) &= Y_i\left(t\right); \quad i \subset \{1,\ 2,\ ...,\ N_p\} \end{aligned} \tag{4}$$

where $p_i(t)$ is the ith face landmark within the tth frame with the coordinates ($Xi(t)$, $Yi(t)$) of the scale functions. Each frame has N_p face landmarks detected on each face.

The second pair of features captures the temporal information from the current frame from the sequence with respect to the neutral frame. The features are based on the difference between the current frame and the neutral frame from the following sequence:

$$\begin{aligned} f_3\left(p_i\left(t\right)\right) &= \left\|X_i\left(t\right) - X_i\left(1\right)\right\|; \\ f_4\left(p_i\left(t\right)\right) &= \left\|Y_i\left(t\right) - Y_i\left(1\right)\right\|; \end{aligned} \tag{5}$$

where the ith landmark coordinates with the tth frame are $X_i(t)$ and $Y_i(t)$ and the neutral frame is considered to be the first one ($t = 1$) from the sequence.

The third pair of features tracks the rate of change during the facial muscle movements. It consists of the first derivative with respect to time:

$$\begin{aligned} f_5\left(p_i\left(t\right)\right) &= \tfrac{\partial X_i(t)}{\partial t}; \\ f_6\left(p_i\left(t\right)\right) &= \tfrac{\partial X_i(t)}{\partial t}; \end{aligned} \tag{6}$$

Eq. (6) is computed using convolution with a Gaussian derivative as describe in Eq. (3).

Valstar and Pantic [8] found that a temporal window of seven frames is enough to determine the changes in neuromuscular facial action. Based on these findings we approximate the coordinates during 7 frames by a second-order polynomial function g:

$$\begin{aligned} g_x\left(X_i\left(t\right)\right) &= a\left(X_i\right)t^2 + b\left(X_i\right)t + c\left(X_i\right); \\ g_y\left(Y_i\left(t\right)\right) &= a\left(Y_i\right)t^2 + b\left(Y_i\right)t + c\left(Y_i\right); \end{aligned} \tag{7}$$

where t stands for the middle frame from a seven-frame window.

The final pair of descriptors is represented by the polynomial coordinates a, b, and c, as they best represent the movement. The resulting pair of descriptors will be:

$$\begin{aligned} f_7\left(p_i\left(t\right)\right) &= \left[a\left(X_i\right), b\left(X_i\right), c\left(X_i\right)\right], \\ f_8\left(p_i\left(t\right)\right) &= \left[a\left(Y_i\right), b\left(Y_i\right), c\left(Y_i\right)\right]. \end{aligned} \tag{8}$$

Further, we create a feature vector from the concatenation of f_1 and f_2, which has a length of $2 \times N_p$. A second feature vector that aims at describing the temporal aspect of facial expression is obtained from the concatenation of f_3 to f_8 (a $5 \times N_p$ feature-long vector).

4.3. Classification scheme

To determine the final decision given an input sequence we apply a late fusion scheme [33]. We feed each descriptor into a specifically trained multilayer perceptron (MLP). Each of those MLPs is trained for regression

on a specific facial expression. The results for each facial expression are then fed to a support-vector machine (SVM) classifier, which gives the final prediction of the searched facial expression in one of the basic emotion categories.

4.4. Training and testing

An 8-fold cross-validation procedure is used for training and testing. We select the first neutral frame in each sequence, which represents our reference frame, and the last three frames from the apex, as can be seen in Figure 3. The performance of the facial analysis system significantly increases when a neutral face is used, as discussed by Tian and Bolle [34]. Therefore, the reference frame that contains the neutral expression is a hard requirement for the proposed system. In order to perform the cross-validation of the training data, we randomly split the dataset into 8 subsets. The subjects of any two subsets do not overlap and all facial expression examples are well balanced between sets. We train 8 times, and each time we use 7 folds as training data and the remaining fold as testing data. The final accuracy is the average of the 8 runs.

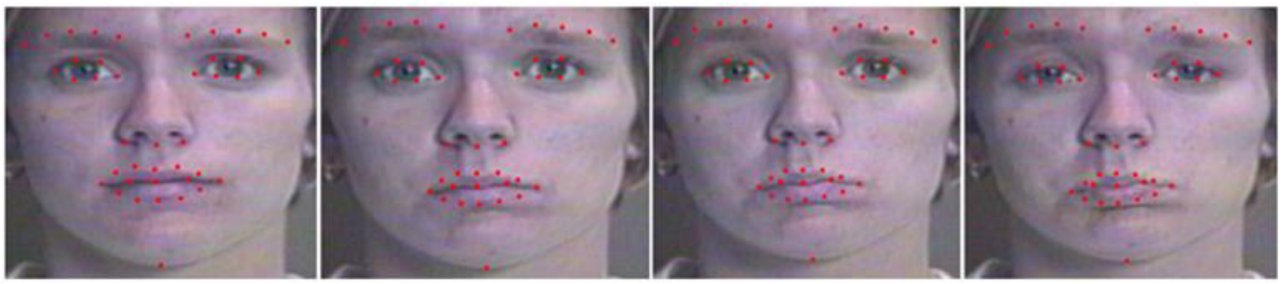

Figure 3. Modification of landmark distribution while the facial expression goes from neutral to apex (sadness). Image sequence taken from Cohn-Kanade+ database [28].

For each sequence of images there are 2 feature vectors for an image: one feature vector ($2 \times N_p$ long) contains the basic coordinates of the facial landmarks and a second one ($5 \times N_p$ long) describes the facial movements. For each of the searched facial expressions we train two MLPs, which return regression responses later fed into a SVM. It results in 14 neural networks, 2 for each facial expression, with a two-class approach representing the presence of the specific facial expression. The numbers of hidden layers and neurons were searched independently for each MLP and the best performing configurations can be seen in Table 1. Their outputs are concatenated and fed into an SVM for each facial expression classified into one of the basic emotion categories. We used the support-vector machine implemented in LibSVM [35], with an RBF kernel. The γ parameter of the kernel and respectively the cost C parameter of the SVM are found by a grid search. The test procedure respects the same work-flow.

5. Results

5.1. Landmark localization influence

First we discuss how the achieved results are influenced by the facial landmarks' localization. In our previous work [17], in order to evaluate the impact of landmarks, true positions were corrupted by noise and we showed that facial expression recognition is robust with respect to landmark localization. Here we compare the results for facial expression recognition starting from fiducial points given by automated methods.

Table 1. Best performing MLP configurations for each emotion.

Emotion	Network	Number of hidden layers	Number of neurons on each layer
Angry	MLP1	1	20
	MLP2	1	80
Happy	MLP1	1	10
	MLP2	3	80
			70
			60
Surprise	MLP1	1	10
	MLP2	1	70
Sadness	MLP1	1	60
	MLP2	3	50
			40
			30
Disgust	MLP1	1	70
	MLP2	2	50
			40
Contempt	MLP1	3	80
			70
			60
	MLP2	1	90

Several methods for facial landmarks localization have been evaluated:

- the method by Zhu and Ramanan in [36], which provides 40 points on the given face;
- a deep convolutional neural network, DLISCD (Deep Learning in Image Segmentation, Classification and Detection), trained on BioId, LFW, and LPFW to give 15 points on the given face (https://arxiv.org/abs/1605.09612);
- the results of the Google Cloud Vision solution [37], which gives 28 points on the face.

The comparison between the performances of the three methods uses the proximity measure proposed in [38]:

$$m_e = \frac{1}{ns} \sum_{i=1}^{n} d_i, \tag{9}$$

where d_i is the Euclidean distance between each individual feature location and the ground truth location, while s is the ground truth interocular distance between the left and right eye pupils. n is the number of chosen feature locations. The results can be seen in Figure 4 and examples of localized landmarks are shown in Figure 5.

The facial points given by Google Cloud Vision are more accurately localized compared to the ones given by the other methods. However, even with this method the results are less accurate for the images that contain the expressions at the apex if compared to the neutral pose image. The main cause can be related to the lack of such images in the training set for all alternatives. Only 24.89% of the images have localization errors of less than 0.05 for the most accurate tested solution, Google Cloud Vision.

Using the resulting landmark positions, we tested the proposed algorithm. The mean obtained recognition rates can be seen in Table 2. The best result is obtained using the landmarks detected by Google Cloud Vision,

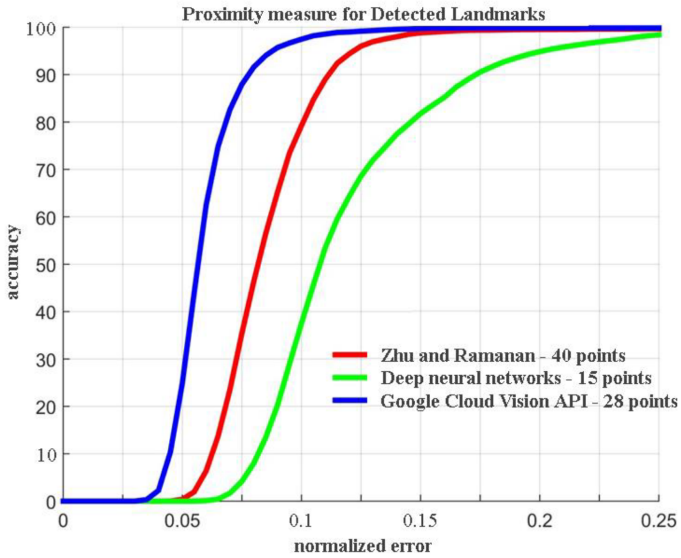

Figure 4. Proximity measure for landmarks' localization computed on CK+ database: Zhu and Ramanan [36], DLISCD, and Google Cloud Vision [37].

which are the most accurate. However, although the landmarks localized with deep neural networks are less accurate, the mean detection rate for facial expression classification is better compared to the one obtained using the 40 landmarks given by the method proposed by Zhu and Ramanan [34]. A potential explanation lies in the number and position of the detected points. For instance, the solution of Zhu and Ramanan generates high accuracies for points on the nose that are less informative while facing the task of facial expression recognition.

Table 2. The recognition rates [%] of the proposed method on the CK+ database with respect to the used landmarks' localization method.

Landmarks' localization method	Mean recognition rate for expression classification
Zhu and Ramanan [36] - 40 pts	91.99
DLISCD - 15 pts	93.57
Google Cloud Vision [37] - 28 pts	94.28

5.2. Expression-wise detection

In Table 3 the recognition rates of the proposed method are given for each of the seven searched emotions, in parallel with the number of cases for each emotion. One can easily notice that the best recognition rates are obtained for happiness and surprise, which also have the most examples in the database and are more distinct.

5.3. Comparison with the state of the art

The proposed method is closely related to the one introduced in [8] and our previous work [17]. Direct comparison with [8] is harder as the previous method does not test on the CK+ database. However, it does test on the original Cohn-Kanade database, which is included in the extended version. The total number of failures for our method is 18 while on a subset the original facial dynamics method reports 55 misclassifications.

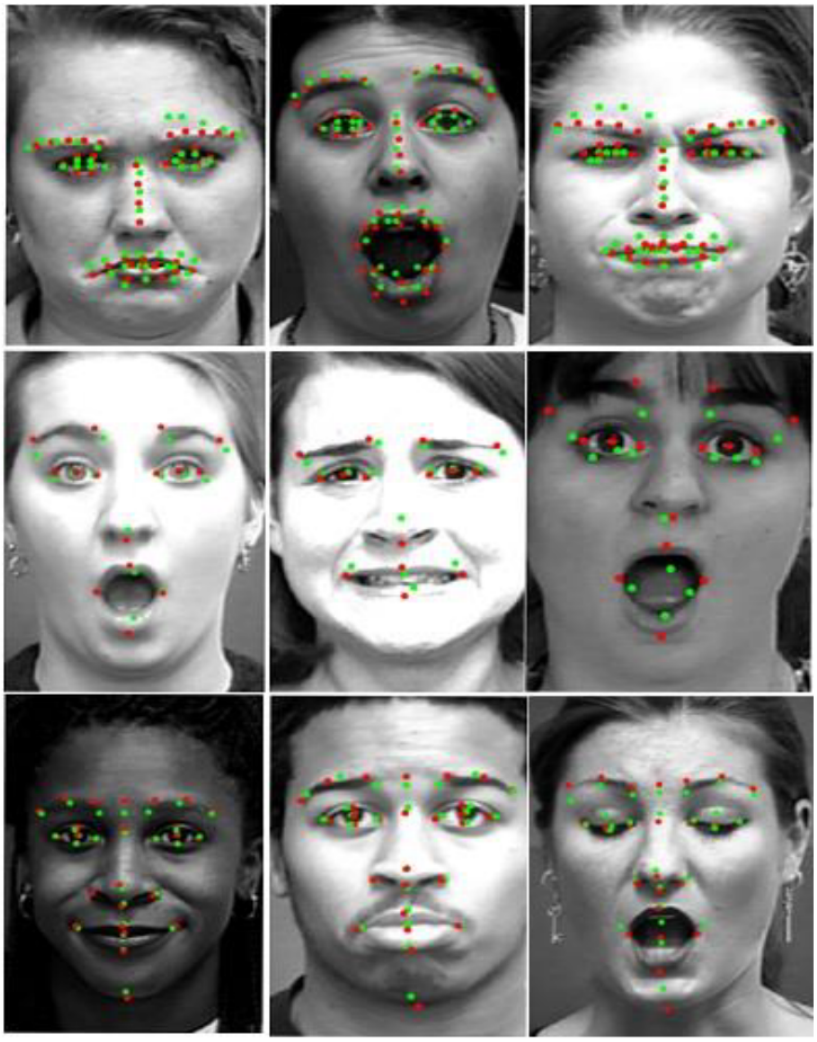

Figure 5. Examples of detected landmarks (green) vs. ground truth (red) on CK+ database: Zhu and Ramanan [36] (first row), DLISCD (second row), Google Cloud Vision [37] (third row).

Table 3. The recognition rate [%] of the proposed method on the CK+ database for each of the searched emotions.

Emotion	Cases in the database	Recognition rate [%]
Angry	45	91.57
Happy	69	97.23
Surprise	83	95.5
Sadness	28	93.89
Fear	25	93.61
Disgust	59	94.72
Contempt	18	93.43

A comparison with related methods that evaluate the performance on the CK+ database can be seen in Table 4. Various methods use various number of folds and cross-validation or not. The proposed method is tested on 8-fold cross-validation, person-independent using a stringent evaluation, and thus numerical comparison is effective.

Table 4. The recognition rate [%] of the proposed method on the CK+ database compared to state-of-the-art methods.

Method	Emotions tested	Recognition rate [%]
Rudovic et al. [23]	6	86.8
Youssif and Asker [24]	6	93.31
Yi et al. [25]	6	88.7
Shbib and Zhou [26]	6	92.1
Zhong et al. [28]	6	89.89
Rivera et al. [22]	7	89.3
Liu at al. [27]	7	92.05
Lucey et al. [30]	7	83.32
Bandrabur et al. [17] - older work	7	90.96
Proposed	7	94.28

Compared to our older method [17], the additional refinements of the current proposal increased the performance to more than 3%. We emphasize that the proposed method outperforms all other solution, as reported in Table 3.

6. Conclusion

We have presented a novel algorithm to classify facial expressions in image sequences. We model the dynamics of facial muscle movement into a set of feature descriptors, which are then fed into a late fusion system. We have shown that our method outperforms the state-of-the-art algorithms published in high-profile publications.

Future work will involve using a bigger database with spontaneous facial expressions. This database contains also in-house acquisitions, which are to be validated from a psychological point of view. We also aim to apply the method to 3D facial points, provided by IR camera, since our feature descriptor is geometric-based.

7. Acknowledgment

The work was partially funded by the Sectoral Operational Programme Human Resources Development 2007-2013 of the Ministry of European Funds through the Financial Agreement POSDRU/159/1.5/S/134395 and partially by a grant of the Romanian National Authority for Scientific Research and Innovation, CNCS UEFISCDI, number PN-II-RU-TE-2014-4-0733.

References

[1] Pentland A, Heibeck T. Honest Signals: How They Shape Our World. Cambridge, MA, USA: MIT Press, 2010.

[2] Yang Q, Li C, Li Z. Application of FTGSVM algorithm in expression recognition of fatigue driving. Journal of Multimedia 2014; 9: 527-533.

[3] Basu N, Nag S, Bandyopadhyay SK. Retrieval of facial expressions for facilitating crime investigation. Asian J Sci Technol 2016; 7: 2381-2387.

[4] Hazlett RL. Measuring emotional valence during interactive experiences: boys at video game play. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; 2006. pp. 1023-1026.

[5] Girard J, Cohn J, Mahoor M, Mavadati S, Hammal Z, Rosenwald D. Nonverbal social withdrawal in depression: evidence from manual and automatic analyses. Image Vision Comput 2014; 32: 641-647.

[6] Florea C, Florea L, Vertan C. Learning pain from emotion: transferred hot data representation for pain intensity estimation. In: European Conference on Computer Vision; 2014. pp. 778-790.

[7] Ekman P, Friesen W. Facial Action Coding System: A Technique for the Measurement of Facial Movement. Palo Alto, CA, USA: Consulting Psychologists Press, 1978.

[8] Valstar MF, Pantic M. Fully automatic recognition of the temporal phases of facial actions. IEEE T Syst Man Cy B 2012; 42: 28-43.

[9] Liu P, Han S, Meng Z, Tong Y. Facial expression recognition via a boosted deep belief network. In: IEEE Conference on Computer Vision and Pattern Recognition; 2014. pp. 1805-1812.

[10] Mery D, Bowyer K. Recognition of facial attributes using adaptive sparse representations of random patches. In: European Conference on Computer Vision; 2014. pp. 778-792.

[11] Williams A. Facial expression of pain: an evolutionary account. Behav Brain Sci 2002; 25: 475-480.

[12] Ekman P, Rosenberg EL. What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System. Oxford, UK: Oxford University Press, 1997.

[13] Ekman P, Cordaro D. What is meant by calling emotions basic. Emotion Review 2011; 3: 364-370.

[14] Zeng Z, Pantic M, Roisman GI, Huang TS. A survey of affect recognition methods: audio, visual, and spontaneous expressions. IEEE T Pattern Anal 2009; 31: 39-58.

[15] De la Torre F, Cohn JF. Facial expression analysis. In: Moeslund TB, Hilton A, Krüger V, Sigal L. Visual Analysis of Humans. Berlin, Germany: Springer, 2011. pp. 377-409.

[16] Pantic M, Patras I. Dynamics of facial expression: recognition of facial actions and their temporal segments from face profile image sequences. IEEE T Syst Man Cy B 2006; 36: 433-449.

[17] Bandrabur A, Florea L, Florea C, Mancas M. Emotion identification by facial landmarks dynamics analysis. In: IEEE International Conference on Intelligent Computer Communication and Processing; 2015. pp. 379-382.

[18] Tian YL. Evaluation of face resolution for expression analysis. In: Computer Vision and Pattern Recognition Workshop; 2004. p. 82.

[19] Littlewort G, Bartlett MS, Fasel I, Susskind J, Movellan J. Dynamics of facial expression extracted automatically from video. Image Vision Comput 2006; 24: 615-625.

[20] Jiang B, Valstar MF, Pantic M. Action unit detection using sparse appearance descriptors in space-time video volumes. In: Automatic Face and Gesture Recognition and Workshops; 2011. pp. 314-321.

[21] Zhao G, Pietikainen M. Dynamic texture recognition using local binary patterns with an application to facial expressions. IEEE T Pattern Anal 2007; 29: 915-928.

[22] Rivera AR, Castillo JR, Chae OO. Local directional number pattern for face analysis: face and expression recognition. IEEE T Image Process 2013; 22: 1740-1752.

[23] Rudovic O, Pavlovic V, Pantic M. Multi-output Laplacian dynamic ordinal regression for facial expression recognition and intensity estimation. In: Computer Vision and Pattern Recognition Conference; 2012. pp. 2634-2641.

[24] Youssif AA, Asker WAA. Automatic facial expression recognition system based on geometric and appearance features. Stud Comp Intell 2011; 4: 115.

[25] Yi J, Mao X, Chen L, Xue Y, Compare A. Facial expression recognition considering individual differences in facial structure and texture. IET Comput Vis 2014; 8: 429-440.

[26] Shbib R, Zhou S. Facial expression analysis using active shape model. International Journal of Signal Processing, Image Processing and Pattern Recognition 2015; 8: 9-22.

[27] Liu M, Li S, Shan S, Chen X. AU-aware deep networks for facial expression recognition. In: Automatic Face and Gesture Recognition International Conference and Workshops; 2013. pp. 1-6.

[28] Zhong L, Liu Q, Yang P, Liu B, Huang J, Metaxas DN. Learning active facial patches for expression analysis. In: Computer Vision and Pattern Recognition Conference; 2012. pp. 2562-2569.

[29] Kanade T, Cohn J, Tian Y. Comprehensive database for facial expression analysis. In: Automatic Face and Gesture Recognition Proceedings; 2000. pp. 46-53.

[30] Lucey P, Cohn JF, Kanade T, Saragih J, Ambadar Z, Matthews I. The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression. In: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops; 2010. pp. 94-101.

[31] Viola P, Jones M. Robust real-time face detection. Int J Comput Vision 2004; 57: 137-154.

[32] Lindeberg T. Scale-space theory: a basic tool for analysing structures at different scales. J Appl Stat 1994; 21: 225-270.

[33] Snoek CG, Worring M, Smeulders AW. Early versus late fusion in semantic video analysis. In: Proceedings of the 13th Annual ACM International Conference on Multimedia; 2005. pp. 399-402.

[34] Tian YL, Bolle, RM. Automatic detecting neutral face for face authentication and facial expression analysis. In: AAAI-03 Spring Symposium on Intelligent Multimedia Knowledge Management; 2003. pp. 24-26.

[35] Chang CC, Lin CJ. LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology 2011; 2: 27.

[36] Zhu X, Ramanan D. Face detection, pose estimation, and landmark localization in the wild. InL Computer Vision and Pattern Recognition Conference; 2012. pp. 2879-2886.

[37] Google Cloud Vision API. Image Content Analysis, Google Cloud Platform.

[38] Cristinacce D, Cootes TF. Feature detection and tracking with constrained local models. In: BMVC; 2006. p. 3.