Smile Veracity Recognition Using LBP Features for Image Sequence Processing

Karolina Nurzynska
Institute of Informatics
Silesian University of Technology
Gliwice, Poland
Email: Karolina.Nurzynska@polsl.pl

Bogdan Smolka
Institute of Automatic Control
Silesian University of Technology
Gliwice, Poland
Email: Bogdan.Smolka@polsl.pl

***Abstract*—Correct recognition of emotion veracity exhibited in facial gestures is troublesome for people. Yet, there is a belief that computer systems are able to perceive some tiny changes correlated to veracity expression, invisible for people, and therefore are able to improve proper perception of emotions. This work addresses the problem of spontaneous and posed smile recognition and suggests two approaches. The first one uses the visual cues, in which the feature vector describes the content of evenly sampled frames in the movie by applying uniform local binary patterns. The second one, describes the video sequence with smile intensity information derived from information extracted from each frame, where the feature vector is built using simple statistical measures calculated from this data. These two systems and a combination of them are tested on UVA-NEMO database and proved to deliver encouraging results.**

Keywords–emotion veracity recognition, LBP, smile intensity, classification, image processing

I. Introduction

Computer vision finds ever wider applications for understanding and monitoring the daily life of people. Systems for automatic person identification on the basis of biometric data, such as fingerprints, iris image, ear shape, and face, exist for many years and proved to be very efficient. Therefore, it should not be surprising that the computer vision algorithms are going deeper in world understanding and nowadays methods which can deal with emotion recognition and its genuineness verification are emerging.

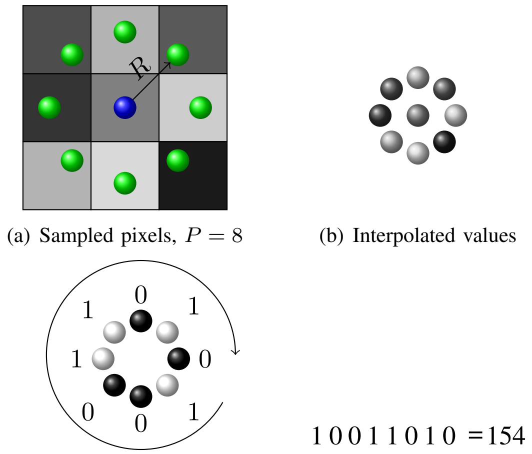
The problem of emotion recognition is not a scientific novelty, though. First approaches were noted in early 90s of the last century. The suggested solutions concentrated on facial muscle movement tracking with optical flow [1] in order to describe a particular emotion. Later some holistic and explicit measures were exploited [2], [3]. However, the more up-to-date solutions consider in their approaches the knowledge about emotion expression included in the Facial Action Coding System (FACS) [4], that analyzes which groups of muscles are activated and in which order, while presenting a particular facial gesture. This approach was firstly exploited in [5] and proved to give satisfactory results.

The literature concerning emotion recognition is very broad and many different solutions have been already employed [6]–[8]. However, the best performance was observed when applying local binary patterns operator (LBP) used for the description of texture of facial image as presented in [9]. In order to classify the obtained feature vectors, the best performance is achieved for support vector machine (SVM), as reported in the comparison [10], but also k nearest neighbors (kNN) or template matching are yielding good results [9].

When transforming the LBP operator into 3D space, where the third dimension corresponds to time, the SVM might be used for emotion expressing happiness and its veracity evaluation [11]. The authors exploited SPOS (spontaneous vs. posed) database and achieved the accuracy up to 80%, when analyzing not only visual data but also infrared readouts. The same authors in [12] investigated methods for automatic recognition of spontaneous micro-expressions using temporal interpolation model, together with comprehensive spontaneous micro-expression corpus. On the other hand, the information of motion, recorded for each part of face also seems crucial for differentiation between genuine and fake expressions. For instance, in [13] the eyelid movement is tracked with very good results achieved for the BBC dataset. Next, in [14] the movement of cheek and lip corner is additionally exploited and the system performance was verified on UVA-NEMO database [15] achieving 87% correct classification rate. Some additional improvements which incorporate the information about subject age are presented in [16].

In this work we concentrate on classification between the spontaneous and posed smile expression. This emotional expression was chosen because it is the most frequently used and according to previous psychological research, proved to be the easiest recognized emotion by others. However, recognition of its veracity proved to be very difficult for human beings.

The paper is structured as follows. Sections II describes the chosen LBP method for image description, while Sec. III gives a short overview of a system for smile intensity estimation. Next, the system configuration is given in Sec. IV. Then, the testing database is presented in Sec. V. The performed experiments, achieved results and its discussion are given in Sec. VI. Finally, Section VII draws the conclusions.



(a) Sampled pixels, $P = 8$ (b) Interpolated values

(c) Binary coded values (d) LBP code

Figure 1. LBP operator calculation.

II. Local binary patterns

Local binary patterns operator [17]–[20] is a texture descriptor which transforms an image into a histogram of codes, where each bin corresponds to the number of similar patterns describing texture patches. The binary code is calculated for P points sampled evenly on a circumference of a circle with radius R around the considered pixel $g_c = I(x_c, y_c)$ in a monochromatic image I as depicted in Fig. 1. This operation is repeated for each pixel in the image and the LBP code is calculated as:

$$\mathrm{LBP}_{P,R}(x_c, y_c) = \sum_{p=0}^{P-1} s(g_p - g_c) \cdot 2^p, \tag{1}$$

where x_c and y_c are coordinates of the central pixel in the considered patch, g_c is gray scale value of the central pixel, g_p is the intensity of the neighboring pixels, $p = 1, \ldots, P$ is the order of the sampled points on the circumference and s is a threshold function based on the difference sign defined as follows:

$$s(z) = \begin{cases} 1, & z \geq 0, \\ 0, & z < 0. \end{cases} \tag{2}$$

The codes obtained for each patch are collected in a LBP histogram, which length is 256 elements for $P = 8$.

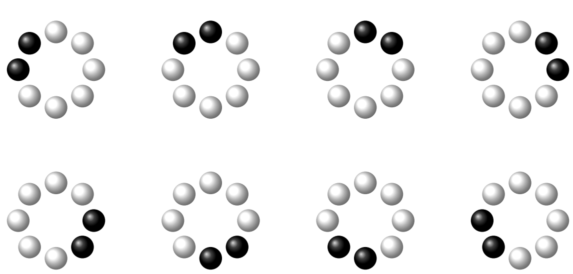


Figure 2. Some examples of uLBP.

Concatenation of several LBP histograms is necessary to describe the image content in detail and in consequence the final feature vector becomes very long. Therefore, it is indispensable to reduce the number of histogram elements. It was pointed out in [17] that some codes are more descriptive than others. The general rule says, that codes which have less than three transitions between 0s and 1s in binary notation of the operator code, describe the data more efficiently. These codes are called uniform LBP (uLBP) (see Fig. 2 for an example) and have separate bins, while all other codes are combined in one common bin. This observation allowed to shorten the length of the LBP histogram to 59 elements while maintaining similar or sometimes better accuracy in data description for image classification purposes.

III. Smile intensity estimation

The uLBP operator was previously applied in our research [10] for classification between smiling and neutral facial display. Moreover, it was noticed [21] that in trained SVM, the distance from the plane dividing data into two classes for considered facial image, corresponds to the intensity of displayed emotion. That enabled to create SNIP: Smile - Neutral Intensity Predictor. In current research this information is exploited as a mean to smile veracity verification.

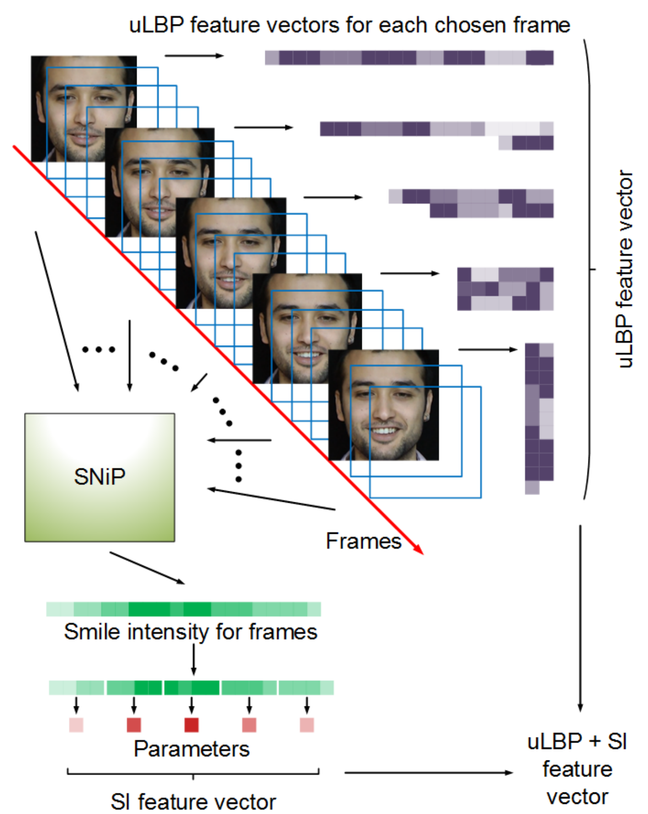


Figure 3. Feature vector computation.

IV. Feature vector computation

The system for automatic estimation of smile veracity analyzes input video sequence concentrating on the changes in facial display as well as in the estimated intensity of presented emotion as depicted in Fig. 3. The visual data inspection assumes accurate analysis of F frames sampled evenly through the whole movie. The face is detected and normalized to resolution of 300×300 pixels. Such image is divided into a mesh of 10×10 sub-images and for each of them the uLBP histogram is calculated. The uLBP histograms from one frame are concatenated and create a partial feature vector. In this way, the final visual content descriptor ties the partial feature vectors.

Additionally, the movie is fed into the SNIP system, and for each frame the smile intensity (SI) is obtained giving an ordered set of values. These values are divided into N equal parts and for each of them three measures are calculated: mean, median and standard deviation, which are used as features describing the data content and constitute a second feature vector exploited for classification.

Finally, the feature vector obtained from the analysis of the visual part is concatenated with the one calculated for smile intensities. Obtained in this way feature vectors are fed to the linear SVM. The classification is made in 10-fold manner, where 9 folds are used for training and the 10th one for testing.

V. Database

UVA-NEMO database [14] contains 1240 movies presenting 597 spontaneous and 643 posed smiles. The movies are recorded in color in resolution of 1920×1080 pixels and stored in MP4 format with 50 fps. The data is collected from 400 subjects (185 females and 215 males), whose age varied from 8 to 76. Single recording took up to 5 minutes and it was assumed that the presentation of an expression should start and finish with neutral facial display. The movies prepared for experiments have different length which vary from 41 to 717 frames.

The authors of [14] support the database with an experiment protocol which divides the dataset into 10 subsets, which assures that movies describing one subject are in the same group. Moreover, there are five different protocols, which divide the data considering age (YOUNG – 378 movies, ADULT – 660), gender (FEMALE – 528, MALE – 510) and collect all movies (ALL – 1038).

VI. Results and discussion

The goal of the first experiment was to check whether it is possible to recognize spontaneous from posed smiles using evenly sampled frames, where the feature vector was a concatenation of vectors obtained from a single frame. The feature vector describes the visual part of the movie by calculating uLBP histograms for selected frames as described in Sec. IV. The total length of the feature vector used for classification was $5900 \times F$.

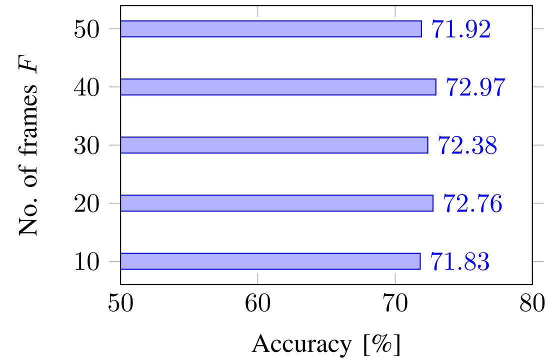


Figure 4. The impact of the number of sampled frames F on spontaneous versus posed smiles classification (Dataset: ADULT).

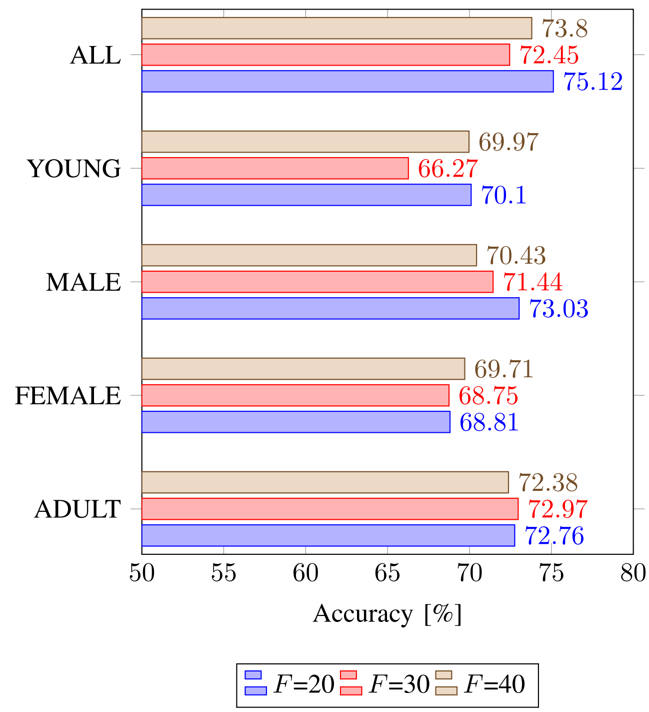


Figure 5. Spontaneous vs. posed smile classification ratio using uLBP.

Figure 4 presents the impact of the number of frames, which were evenly sampled in the movie, on the classification between spontaneous and posed smiling expressions. In this test a subset ADULT of UVA-NEMO database was used. As it is depicted, using frames amount between 20 to 40 gives comparable results, which are a little bit better than for 10 or 50 frames. We did not evaluate more frames than 50, because the shortest movie found in the experiments protocol has around 70 frames, and we wanted to keep removing at least some frames from each movie.

Next, we have compared the correct classification ratio

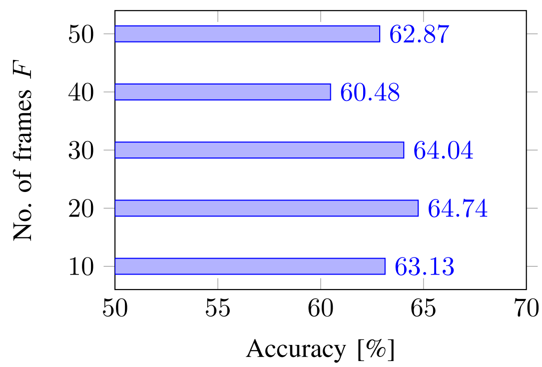


Figure 6. The impact of the number of frames F of the sequence of smile intensity on spontaneous versus posed smiles (Dataset: ADULT).

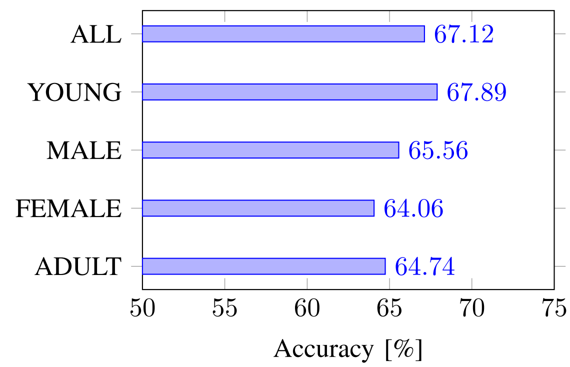


Figure 7. Spontaneous versus posed smile correct classification ratio using smile intensity as a descriptor ($N = 20$).

between spontaneous and posed smiles for all exemplary groups: ADULT, FEMALE, MALE, YOUNG, and ALL. The result of this experiment is presented in Fig. 5. Here the best results were observed for ALL group with $F = 20$. For this number of frames also the best results are obtained when YOUNG, and MALE dataset are considered, while FEMALE scores better when $F = 40$ and ADULT for $F = 30$. What is interesting, the results of smile veracity classification when age or gender is considered, lowers the classifier accuracy. Moreover, it proved to be more complicated to correctly recognize between smiles types in case of FEMALE and YOUNG dataset.

The second experiment exploited the smiling intensity data obtained for each movie using the SNIP system. Figure 6 shows the performance achieved for ADULT dataset when the data was divided into different number of parts for which the features are calculated. As one can see, the best performance was achieved when $N = 20$, which is not a surprise since the bigger N, the shortest the vector of data describing the chosen stage of smile and more difficult to calculate representative features. Figure 7 depicts correct classification ratio achieved for all experimental datasets. Here, the best result was gathered for the YOUNG set, while the worst ones were achieved for FEMALE and ADULT. Although this approach is not very accurate, it is worth pointing out that in order to describe the image content, only one value is used. Having this in mind, makes this results worth considering as a part of a more complex system.

Finally, the third experiment verifies whether a combination of feature vectors describing the visual part and smile intensity improves the classification between spontaneous and posed expression. Figure 8 exhibits the correct classification results depending on the experimental protocol dataset. It shows that combining these two different types of information does not improve the classification much. It is a little better in case of ALL, MALE and ADULT, but visible deterioration of accuracy is noticeable for FEMALE and YOUNG.

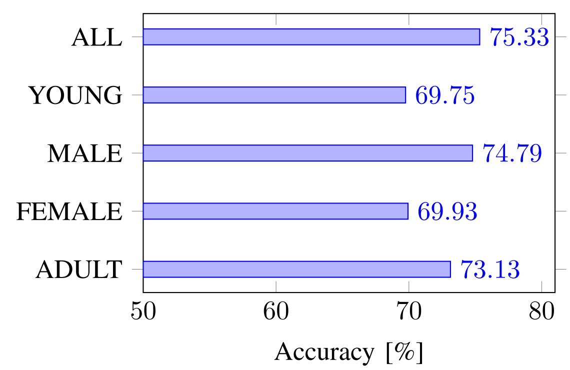


Figure 8. Spontaneous versus posed smile correct classification ratio using visual data and smile intensity as a descriptor ($N = 20$, $F = 20$).

Comparison of achieved results with experiments performed by others is difficult. The work [13] reports better accuracy in veracity detection using UVA-NEMO dataset, but it is not clear whether the description techniques exploit the information about movie length, which is strongly correlated with the type of smile. Moreover, in order to achieve accurate positions of feature points on the faces, they are manually annotated on the initial frame. Also better results were recorded in [22] for fake and true smiles which are collected between others from the BBC dataset. The presented method uses Gabor filters with various parameters not specified in the work. Finally, the approach using CLBP-TOP for input content description is suggested in [11], where for SPOS corpus better results are achieved. Yet, the amount of data is very small and difficult to assume its generality when only short image sequences from several subjects are accessible.

VII. Conclusions

This work presents three approaches for designing a system for spontaneous and posed smile automatic classification. First of them concentrates on the visual aspect, where for evenly sampled frames in the movie, the uLBP feature vector is calculated. Using linear SVM for such data classification enables classification accuracy of 75%. Second solution exploits the information about smile intensity calculated by the SNIP system with additional statistical features. The accuracy of 67% seems very promising as only one feature per image is exploited. Finally, a combination of the described approaches was evaluated, but it did not improve much the best results obtained using the uLBP features.

The performed experiments show that application of visual information is a promising starting point for further investigation of classification. Moreover, using simple linear SVM for recognition, gives very good results, however the authors are aware that using more sophisticated classification approaches such as this one presented in [11], should improve significantly the results. In further research we are going to verify this hypothesis.

Acknowledgment

This work has been supported by the Polish National Science Centre (NCN) under the Grant: DEC-2012/07/B/ST6/01227.

References

[1] K. Mase, "An application of optical flow – extraction of facial expression," in *IAPR Workshop on Machine Vision Applications*, 1990, pp. 195–198.

[2] I. A. Essa and A. P. Pentland, "Coding, analysis, interpretation, and recognition of facial expressions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 757–763, Jul. 1997.

[3] M. Bartlett, J. Hager, P. Ekman, and T. Sejnowski, "Measuring facial expressions by computer image analysis," *Psychophysiology*, vol. 36(2), pp. 253–263, 1999.

[4] P. Ekman and W. Friesen, "Facial action coding system: a technique for the measurement of facial movement," *Consulting Psychologists Press*, 1978.

[5] J. Cohn, A. Zlochower, J.-J. J. Lien, and T. Kanade, "Feature-point tracking by optical flow discriminates subtle differences in facial expression," in *Proceedings of the 3rd IEEE International Conference on Automatic Face and Gesture Recognition (FG '98)*, April 1998, pp. 396 – 401.

[6] S. Liao, W. Fan, A. Chung, and D.-Y. Yeung, "Facial expression recognition using advanced local binary patterns, Tsallis entropies and global appearance features," in *Image Processing, 2006 IEEE International Conference on*, Oct 2006, pp. 665–668.

[7] M. Szwoch and P. Pieniek, "Facial emotion recognition using depth data," in *2015 8th International Conference on Human System Interaction (HSI)*, June 2015, pp. 271–277.

[8] A. Kołakowska, A. Landowska, M. Szwoch, W. Szwoch, and M. R. Wróbel, "Emotion recognition and its application in software engineering," in *2013 6th International Conference on Human System Interactions (HSI)*, June 2013, pp. 532–539.

[9] C. Shan, S. Gong, and P. McOwan, "Facial expression recognition based on local binary patterns: A comprehensive study," *Image Vision Comput.*, vol. 27, no. 6, pp. 803–816, May 2009.

[10] K. Nurzynska and B. Smolka, "Smiling and neutral facial display recognition with the local binary patterns operator," *Journal of Medical Imaging and Health Informatics*, vol. 5, no. 6, pp. 1374–1382, 2015.

[11] T. Pfister, X. Li, G. Zhao, and M. Pietikäinen, "Differentiating spontaneous from posed facial expressions within a generic facial expression recognition framework," in *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*, Nov 2011, pp. 868–875.

[12] T. Pfister, X. Li, G. Zhao, and M. Pietikäinen, "Recognising spontaneous facial micro-expressions," in *2011 International Conference on Computer Vision*, Nov 2011, pp. 1449–1456.

[13] H. Dibeklioğlu, R. Valenti, A. A. Salah, and T. Gevers, "Eyes do not lie: Spontaneous versus posed smiles," in *Proceedings of the 18th ACM International Conference on Multimedia*, ser. MM '10. New York, NY, USA: ACM, 2010, pp. 703–706.

[14] H. Dibeklioğlu, A. A. Salah, and T. Gevers, *Computer Vision – ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part III*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, ch. Are You Really Smiling at Me? Spontaneous versus Posed Enjoyment Smiles, pp. 525–538.

[15] "Uva-nemo database," http://www.uva-nemo.org/, accessed: 2016-04-27.

[16] H. Dibeklioğlu, A. A. Salah, and T. Gevers, "Recognition of genuine smiles," *IEEE Transactions on Multimedia*, vol. 17, no. 3, pp. 279–294, March 2015.

[17] M. Pietikäinen, A. Hadid, G. Zhao, and T. Ahonen, *Computer Vision Using Local Binary Patterns*, ser. Computational Imaging and Vision. Springer, 2011, vol. 40.

[18] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51–59, 1996.

[19] T. Ojala, M. Pietikäinen, and T. Mäenpää, "A generalized local binary pattern operator for multiresolution gray scale and rotation invariant texture classification," in *Proc. 2nd Intern. Conf. on Advances in Pattern Recognition*. London, UK: Springer, 2001, pp. 397–406.

[20] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 7, pp. 971–987, Jul 2002.

[21] K. Nurzynska and B. Smolka, *Computational Vision and Medical Image Processing V*. Taylor & Francis Group, 2016, ch. SNIP: Smile–Neutral facial display Intensity Predictor, pp. 347–353.

[22] H. Liu and P. Wu, "Comparison of methods for smile deceit detection by training au6 and au12 simultaneously," in *2012 19th IEEE International Conference on Image Processing*, Sept 2012, pp. 1805–1808.