SPONTANEOUS SMILE DETECTION WITH APPLICATION OF LANDMARK POINTS SUPPORTED BY VISUAL INDICATIONS

Karolina Nurzynska and Bogdan Smolka

Faculty of Automatic Control, Electronics, and Computer Science
Silesian University of Technology,
ul. Akademicka 2A, 44-100 Gliwice, Poland
Karolina.Nurzynska@polsl.pl, Bogdan.Smolka@polsl.pl

ABSTRACT

When automatic recognition of emotion became feasible, novel challenges has evolved. One of them is the recognition whether a presented emotion is genuine or not. In this work, a fully automated system for differentiation between spontaneous and posed smiles is presented. This solution exploits information derived from landmark points, which track the movement of fiducial elements of face. Additionally, the smile intensity computed with SNiP (Smile-Neutral intensity Predictor) system is exploited to deliver additional descriptive data. The performed experiments revealed that when an image sequence describes all phases of smile, the landmark points based approach achieves almost 80% accuracy, but when only onset is exploited, additional support from visual cues is necessary to obtain comparable outcomes.

KEYWORDS

Smile veracity recognition, Landmarks, Smile intensity, Classification

1. INTRODUCTION

Automatic analysis of image content finds its application in many various domains: starting from medical image analysis in order to facilitate diagnosis, going through detection of flaws in products on a production line in a factory, and finishing on human behaviour interpretation. In all these situations, similar approaches for image processing and recognition are implemented. The image content is explored in order to derive some characteristics, which later on become a feature set. These characteristics may originate from visual information, when for instance the texture is exploited as a data source, or describe abstract details, when shape and active shape features are considered.

Having an image sequence of facial gesture it is possible to recognize which emotion was presented with very high efficiency, as was proved in several research works [2, 3, 4, 5, 6]. However, the accuracy of such system performance decreases significantly, when real life scenarios are considered [7, 8, 9], because programs trained on posed emotions displayed in laboratory environment are used to very strong facial gesture presentation, which is not common

David C. Wyld et al. (Eds) : CSITA, ISPR, ARIN, DMAP, CCSIT, AISC, SIPP, PDCTA, SOEN - 2017
pp. 187– 194, 2017. © CS & IT-CSCP 2017 DOI : 10.5121/csit.2017.70118

in daily life, not to mention lighting variation or occlusions. Nevertheless, in most cases recognition of the happiness emotion is on satisfactory level. Therefore, further investigations were performed, which aimed at recognition whether the smile corresponding to this emotion is a genuine one or not [10, 11, 12].

When designing a system for smile veracity recognition, two most common approaches are identified: One exploits facial landmarks for feature calculation [10], where distances and characteristic points movement relationship are considered. The other uses data sampled from the image texture [12] and combines them with smile intensity measure. The presented solution is a combination of both. The veracity of smile is described by automatically derived characteristic points, which are later normalized before the features are computed. On the other hand, a smile intensity function derived from the image textural content, following the idea of SNIP system [1], is applied.

The paper starts with description of facial landmarks determination presented in Sec. 2. Next, Sec. 3 presents the method for smile intensity function calculation. The details of image sequence corpora selected for results verification is given in Sec. 4. Then, the results are presented in Sec. 5 and conclusions are drawn in Sec. 6.

2. CHARACTERISTIC POINTS

In order to describe an emotion presented on image, the positions of facial landmarks called also characteristic points are calculated. These landmarks correspond to characteristic regions of the face, such as eyes, nose, lips, etc. and are exploited to track their changes during the presentation of emotion. Figure 1 depicts the placement of data obtained with implementation presented in [13].

2.1 Data Normalization

The subject during emotion presentation may move the head freely in any direction, therefore the landmarks are firstly normalized as was suggested in [10, 11]. It is necessary due to differences in head size, which is mostly noticeable when the head moves in and out of the camera focus as well as anatomical differences between humans should be compensated. Because the proposed automatic method for landmark annotation returns the coordinates in XY-plane only, therefore the rotation normalization was neglected and only scaling the landmarks positions was performed. It assumes a constant distance (100 pixels) between left (eye_L) and right (eye_R) eye centres, which coordinates are computed as an average of red marks presented in Fig. 1. New characteristic points coordinates p^* are computed as follows

$$p_i^* = (p_i - \frac{(eye_L + eye_R)}{2}) \cdot \frac{100}{\varrho(eye_L, eye_R)}, \quad (1)$$

where $\varrho(a, b)$ denotes the Euclidean distance between points a and b.

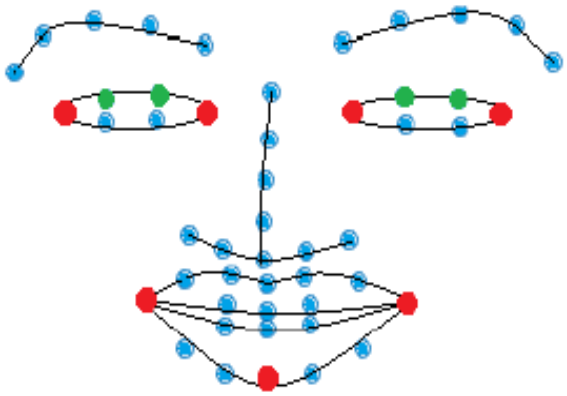

Figure 1: Characteristic points denoted on the facial image.

Table 1: Formulas for features calculation.

Feature	Definition
Duration	$\zeta(F_+), \zeta(F_-), \zeta(F)$
Duration Ratio	$\frac{\zeta(F_+)}{\zeta(F)}, \frac{\zeta(F_-)}{\zeta(F)},$
Maximum Amplitude	$\max(F)$
Mean Amplitude	$\frac{\sum(F)}{\zeta(F)}, \frac{\sum(F_+)}{\zeta(F_+)}, \frac{\sum(F_-)}{\zeta(F_-)}$
STD of Amplitude	$\mathrm{std}(F)$
Total Amplitude	$\sum(F_+), \sum(\lvert F_-\rvert)$
Net Amplitude	$\sum(F_+) - \sum(\lvert F_-\rvert)$
Amplitude Ratio	$\frac{\sum(F_+)}{\sum(F_+)+\sum(\lvert F_-\rvert))}, \frac{\sum(F_-)}{\sum(F_+)+\sum(\lvert F_-\rvert))}$
Maximum Speed	$\max(S_+), \max(S_-)$
Mean Speed	$\frac{\sum(S_+)}{\zeta(S_+)}, \frac{\sum(S_-)}{\zeta(S_-)}$
Maximum Acceleration	$\max(A_+), \max(A_-)$
Mean Acceleration	$\frac{\sum(A_+)}{\zeta(A_+)}, \frac{\sum(A_-)}{\zeta(A_-)}$
Net Ampl. Duration Ratio	$\frac{\sum(F_+)-\sum(\lvert F_-\rvert)}{\zeta(F)}$

2.2 Feature Calculation

In order to distinguish between the posed and genuine smile, its course is divided into onset, apex, and offset phases. The onset is defined as a time from the emotion start to its full presentation. The offset presents signal attenuation, while the apex depicts the full emotion facial gesture. The length ζ of each part may vary between subjects, as well as presented emotions. In order to split the data sequence into these three parts, a smile amplitude signal is considered. Figure 2 presents exemplary progress of smile amplitude change within an image sequence. Such plot is exploited for data division, with assumption that values above 1 represent smile apex. Frames preceding the apex belong to the onset phase, while those following, to the offset. In case when several smile peaks are detected, the longest one is considered in calculations.

The smile strength is given by lip movement amplitude calculated following the formula given in [10, 11]. It exploits the information of landmarks which describes the movement of lips corners and the bottom lip middle point (all pointed in red in Fig. 1). Similarly, for the eye's region, the eyelid magnitude is computed, basing on eyes corners (red colour in Fig. 1) and middle point of the upper eyelid calculated as an average of coefficients obtained for green points in Fig. 1.

Finally, the features gathered in Tab. 1 are computed separately for each smile phase as well as for each considered facial region. Moreover, information about local signal increase (denoted with '-' symbol in index) and decrease (indicated with '+' symbol in index) are exploited. For each signal F, a speed $S = dF/dt$ and acceleration $A = d^2F/dt$ are obtained. That gives 24 features for each region and phase.

Additionally, in the experiments, general approach to feature calculation was investigated. In such case, a set of features for whole smile or eyelid movement intensity was considered and as an increasing part the onset was exploited, while offset described the decreasing element in formulas in Tab. 1. Since some parameters are too strongly related to video duration and could overlap with previously calculated ones, they were removed (Duration) or redefined. For *Maximum Amplitude*, *Mean Amplitude*, *Total Amplitude*, *Maximum Speed*, *Mean Speed*, *Maximum Acceleration*, and *Mean Acceleration* only one value without distinction for increasing and decreasing parts was calculated. This generalization resulted in 14 features.

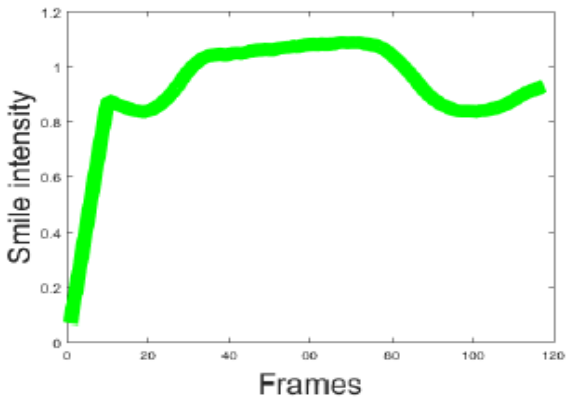

Figure 2. Smile intensity magnitude

Figure 3: Exemplary frames from image sequence in UvA-NEMO dataset.

3. SMILE INTENSITY

The smile intensity or amplitude can be also derived from the visual cues. Such a case was presented in [1], where texture operators extract feature vectors to train SVM classifier for recognition between smiling and neutral facial display. This classifier response, in other words the object distance from the division plane, is returned as a smile intensity function. In presented research, this approach for smile amplitude description is evaluated as an alternative one for these calculated exploiting lips corner movement annotated by landmarks. Similarly, the smile intensity function is divided into three phases for which parameters from Tab. 1 are computed.

4. DATABASE

The accuracy of posed and spontaneous smile recognition was verified on the UvA-NEMO database [10], which gathers images presenting only happiness emotion in scenarios allowing to record its both versions. The image sequences collected in this dataset depict happiness emotion presented by 400 subjects, whose age is in the range from 8 to 76. There are 1240 videos presenting 597 spontaneous and 643 posed facial gestures sequences. The data was recorded in controlled environment in RGB format with high resolution of 1920×1080 pixels, what results in average face resolution of 400×400 pixels. Some examples are given in Fig. 3.

5. EXPERIMENTS AND RESULTS

The goal of this experiment was to investigate not only the region influence on smile veracity detection but also the impact of the smile phase. The feature vectors built from parameters adequate for onset, apex, and offset phases of smile and eye magnitude were calculated. Moreover, a combination of all phase data was tested and the global approach for parameter calculation was exploited. The combination of all phase and global parameters was named total and tested as well. The features were fed on support vector machine (LIBSVM [14]) with linear kernel, for which the best parameter was evaluated in range $c = 10^{-5} - 10^{5}$. Figure 4 presents scores obtained for UvA-NEMO dataset.

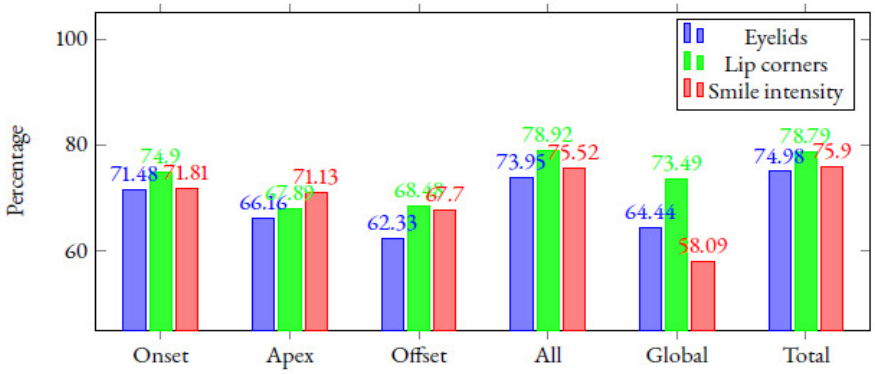

Figure 4: Influence of facial region described by features on classification rates.

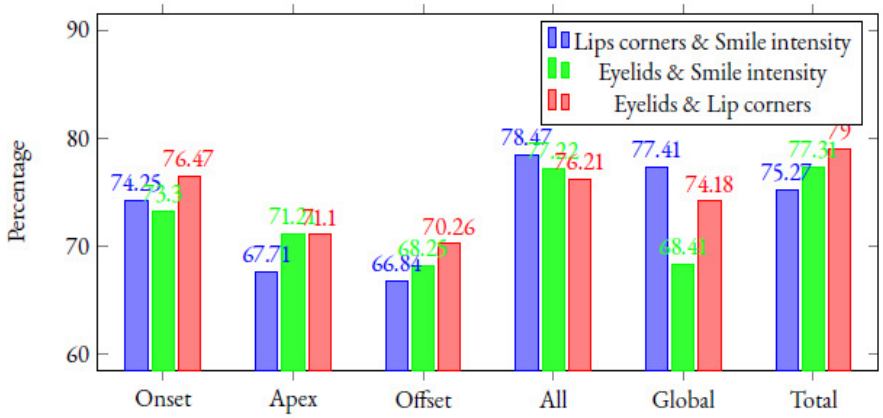

Figure 5: Classification rates when early fusion method was chosen for classification.

Results gathered in Fig. 4 revealed that it is possible to distinguish the posed from genuine smile with high accuracy. This plot shows also that the biggest influence on smile type detection has the onset phase, however for data exploiting smile intensity, similar results were obtained for the apex. Moreover, incorporating information obtained in other phases in all feature vector proved to give the best results. As presumed the global approach outcomes are slightly worse and its combination with all does not improve the accuracy (total), except for the smile intensity case. It is also worth noticing that it is easier to distinguish spontaneous emotion from posed ones, when lips magnitude is considered.

In order to create more accurate classifier, several options were considered. Figure 5 collects correct classification performances for concatenated features vectors describing all possible combinations of pairs of accessible descriptors. The best score of 79% was obtained for the total feature vector build from eyelids and lip corner amplitudes, and slightly outperformed the results when data was described by lips corners supported by smile intensity, which reached 78.47%.

Finally, Fig. 6 presents correct classification performance ratios when two different approaches fo outcome computation were investigated. The early fusion data was obtained by concatenation of feature vectors computed in first experiment and then classified with SVM. The late fusion applied three SVMs classifiers for each feature vector generated in first experiment and used voting to obtain the final score.

The conducted experiments show that it is possible to determine with high probability the veracity of smile using automatic method for landmark detection or visual approach of smile intensity description. Exploiting data derived only from landmark points with calculation of features both for local phases supported with global information enabled correct recognition with almost 80% accuracy. On the other hand, when only onset phase was accessible, combination of amplitude data derived from landmarks with those computed when visual data was exploited were not far behind.

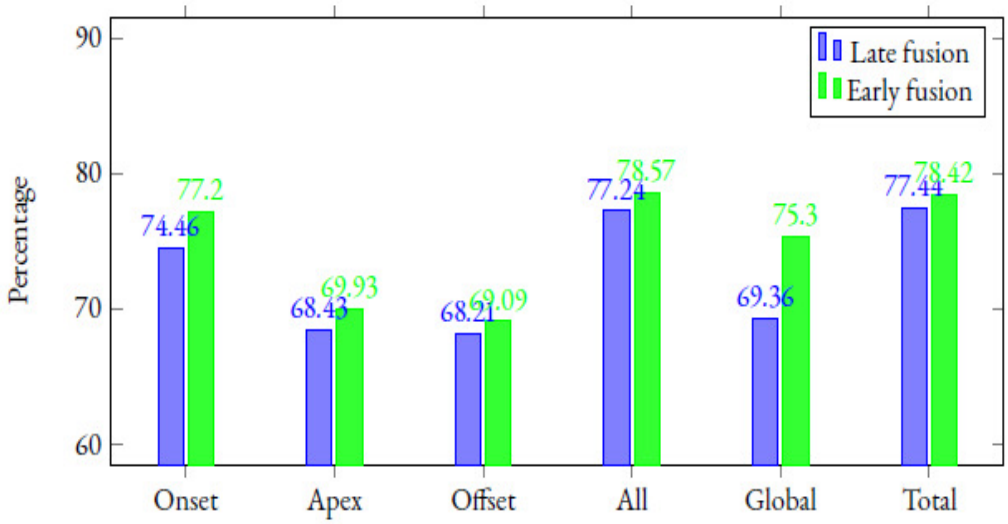

Figure 6: Classification rates when different fusion methods were chosen for classification.

6. CONCLUSIONS

This work presents fully automatic approach to smile veracity detection, which is based on landmarks movements estimation within an image sequence and is supported by information derived from visual cues. Namely, the lip corner and eyelid movements magnitudes are derived from characteristic point location and smile intensity function is achieved from the SNiP system. The feature vector was obtained as a combination of parameters computed for each input function. Using SVM for classification of parameters derived from these data, it was shown, that it is possible to achieve the accuracy of almost 80%, when all information was used. However, only slight deterioration is noticed, when parameters were reduced to describe the onset phase only. In further research, the visual content of images will be explored to improve the classification performance.

ACKNOWLEDGMENT

This work was supported by the Polish National Science Centre (NCN) under the Grant: DEC-2012/07/B/ST6/01227. K. Nurzynska was partially supported by statutory funds for young researchers (BKM/507/RAU2/2016) of the Institute of Informatics, Silesian University of Technology, Poland. B. Smolka received partial funding from statutory funds (BK/213/RAU1/2016) of the Institute of Automatic Control, Silesian University of Technology, Poland.

REFERENCES

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

[2] K. Mase, “An application of optical flow - extraction of facial expression,” in IAPRWorkshop on Machine Vision Applications, pp. 195–198, 1990.

[3] J. Cohn, A. Zlochower, J. Lien, and T. Kanade, “Automated face analysis by feature point tracking has high concurrent validity with manual facs coding,” Psychophysiology, vol. 36, no. 2, pp. 35–43, 1999.

[4] J. Cohn, A. Zlochower, 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), pp. 396 – 401, April 1998.

[5] B. Fasel and J. Luettin, “Automatic facial expression analysis: a survey,” Pattern Recognition, vol. 36, no. 1, pp. 259 – 275, 2003.

[6] C. Shan, S. Gong, and P. W. McOwan, “Facial expression recognition based on local binary patterns: A comprehensive study,” Image Vision Comput., vol. 27, pp. 803–816, May 2009.

[7] A. Dhall, R. Goecke, J. Joshi, M. Wagner, and T. Gedeon, “Emotion recognition in the wild challenge (EmotiW) challenge and workshop summary,” in Proceedings of the 15th ACM on International Conference on Multimodal Interaction, ICMI ’13, (New York, NY, USA), pp. 371–372, ACM, 2013.

[8] J. M. Girard, J. F. Cohn, L. A. Jeni, M. A. Sayette, and F. De la Torre, “Spontaneous facial expression in unscripted social interactions can be measured automatically,” Behavior Research Methods, vol. 47, no. 4, pp. 1136–1147, 2015.

[9] J. Chen, Y. Ariki, and T. Takiguchi, “Robust facial expressions recognition using 3D average face and ameliorated AdaBoost,” in Proceedings of the 21st ACM International Conference on Multimedia, MM ’13, (New York, NY, USA), pp. 661–664, ACM, 2013.

[10] H. Dibeklio˘glu, A. A. Salah, and T. Gevers, Are You Really Smiling at Me? Spontaneous versus Posed Enjoyment Smiles, pp. 525–538. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012.

[11] H. Dibeklio˘glu, A. A. Salah, and T. Gevers, “Recognition of genuine smiles,” IEEE Transactions on Multimedia, vol. 17, pp. 279–294, March 2015.

[12] M. Kawulok, J. Nalepa, K. Nurzynska, and B. Smolka, “In search of truth: Analysis of smile intensity dynamics to detect deception,” in Advances in Artificial Intelligence - IBERAMIA 2016 - 15th Ibero-American Conference on AI, San José, Costa Rica, November 23-25, 2016, Proceedings, pp. 325–337, 2016.

[13] A. Asthana, S. Zafeiriou, S. Cheng, and M. Pantic, “Incremental face alignment in the wild,” in 2014 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2014, Columbus, OH, USA, June 23-28, 2014, pp. 1859–1866, 2014.

[14] C.-C. Chang and C.-J. Lin, “LIBSVM: A library for support vector machines,” ACM Transactions on Intelligent Systems and Technology, vol. 2, pp. 27:1–27:27, 2011. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

AUTHORS

Karolina Nurzynska received her M.E. and Ph.D. degree in the field of computer science from the Silesian University of Technology, Poland in 2005 and 2009, respectively. From 2009 to 2011 she was a Postdoc Researcher in the School of Electrical and Computer Engineering at the Kanazawa University, Japan. From 2011 to 2014 she was leading a project concerning visualisation of underground coal gasification in Central Mining Institute in Katowice, Poland, where in 2012 she was appointed the Assistant Professor position. Since 2013 she is an Assistant Profesor at the Department of Automatic Control, Electronics, and Computer Science of the Silesian University of Technology, Poland. Her research interests include image processing and understanding, data classification and 3D surface reconstruction.

Bogdan Smolka received the Diploma in Physics degree from the Silesian University, Katowice, Poland, in 1986 and the Ph.D. degree in computer science from the Department of Automatic Control, Silesian University of Technology, Gliwice, Poland, in 1998. From 1986 to 1989, he was a Teaching Assistant at the Department of Biophysics, Silesian Medical University, Katowice, Poland. From 1992 to 1994, he was a Teaching Assistant at the Technical University of Esslingen, Germany. Since 1994, he has been with the Silesian University of Technology. In 1998, he was appointed as an Associate Professor in the Department of Automatic Control. He has also been an Associate Researcher with the Multimedia Laboratory, University of Toronto, Canada since 1999. In 2007, Dr. Smolka was promoted to Professor at the Silesian University of Technology. He has published over 300 papers on digital signal and image processing in refereed journals and conference proceedings. His current research interests include low-level color image processing, human-computer interaction and visual aspects of image quality.