Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1685–1692
DOI: 10.15439/2016F390
ACSIS, Vol. 8. ISSN 2300-5963

Mouth features extraction for emotion classification

Staniucha Robert* and Wojciechowski Adam†
Institute of Information Technology
Lodz University of Technology
Wólczańska 215, 90-924 Lodz, Poland
Email: *800671@edu.p.lodz.pl, †adam.wojciechowski@p.lodz.pl

***Abstract*—Face emotions analysis is one of the fundamental techniques that might be exploited in a natural human-computer interaction process and thus is one of the most studied topics in current computer vision literature. In consequence face features extraction is an indispensable element of the face emotion analysis as it influences decision making performance. The paper concentrates on classification of human poses based on mouth. Mouth features extraction, which next to eye region features becomes one of the most representative face regions in the context of emotions retrieval. Additionally, in the paper, original mouth features extraction method was presented. It is gradient based. Evaluation of the method was performed for a subset of the Yale images database and classification accuracy for single emotion is over 70%.**

I. Introduction

VISUAL determinants revealing human emotions comprise [1]: emotional voice, body pose, gestures, gaze direction end facial expressions. Thorough and complete human emotion analysis should consider also state of human environment, both current and passed, that may originate emotions. Typical spontaneous, emotions accompanying, muscular activity last usually between 250 ms and 5s – very rarely longer [1]. As a result not only location of human action is important, but intensity and its dynamics as well. In 1971 Ekman and Friesen [2] postulated 6 basic emotions that reveal distinctive set of features with unique facial expressions. These are happiness, sadness, fear, disgust, surprise and anger. Face poses visual recognition or even monitoring may have considerable influence not only on building playful and intelligent social living environment, control employing nonexpert workers (i.e. crowdsourcing [3]) but also on our security as well (i.e. video surveillance system).

Surveying facial expressions, mouth region seems to be the most representative next to the eyes region. This paper concentrates on visual mouth features extraction which have key influence on further facial expression and emotion detection analysis. The new approach bases on local intensity gradients and subsequent dedicated impulse filter responses. Precise extraction of mouth features can increase the efficiency of a classifier by decreasing its complexity. Authors present a new, alternative method of mouth characteristics extraction and evaluate effectiveness of the method on the well known and reliable Yale faces database [4]. The paper shows also classification results based on elaborated method for common used classifiers.

II. Related work

Biologically facial expressions are generated by contractions of facial muscles, which cause face features temporal deformations. The most evident changes concern eye lids, eye brews, nose, lips or skin possible wrinkles. Facial expression intensity can be measured by geometric deformation of facial features or facial texture analysis, i.e. density of face appearing wrinkles.

Among anthropologically justified face core landmarks forming unquestionable framework for face features extraction researches [5] proposed 11 points: pronasale, alare (left & right), subnasale, chelion (left & right), endocanthion (left & right), exocanthion (left & right) and sellion (fig. 1). Subsequently some minor landmarks can be evaluated.

Face features extraction methods, presented in literature, can be classified into two main groups: appearance based and geometric based methods [6]. Though the appearance based approach seems to be currently the most popular, the geometric based methods seem to be recently neglected but still promising. Even though the geometric approach seems to be very well studied [7], [8], according to Pali [9], among the most evident aspects that can be still improved, within face features extraction, there are dimensionality reduction, features extraction techniques and features subset selection. Thus, the geometrical approach can almost automatically reduce space dimension problem and behave more reliably in demanding scenarios where pose (in-plane and out of plane face rotations) and illumination are not controlled [10].

Face features detection is a challenging task as, due to the quantity of local face image structures, classical corner detectors are useless without considering their context. Researches estimating face inherent features and edges date back to well known Kanade work [11] and were further intensively developed (i.e. [12]). Authors attempted to reconstruct horizontal and vertical lines applying Laplacian operator and evaluating horizontal and vertical integral image projection obtaining effectiveness of about 75% on dedicated, self-prepared databases.

Castrillon [10], Castille [13] and Yang [14] suggested Viola-Jones object detection algorithm for coarse face parts (i.e.: eyes, mouth, nose) localization but it did not localize precisely face landmarks and required further, more detailed features detectors. Even Panning [15] and Wang et al. [16] approaches, measuring distances between face regions, detected with originally elaborated features or Lienhart [17] extended Haar-like

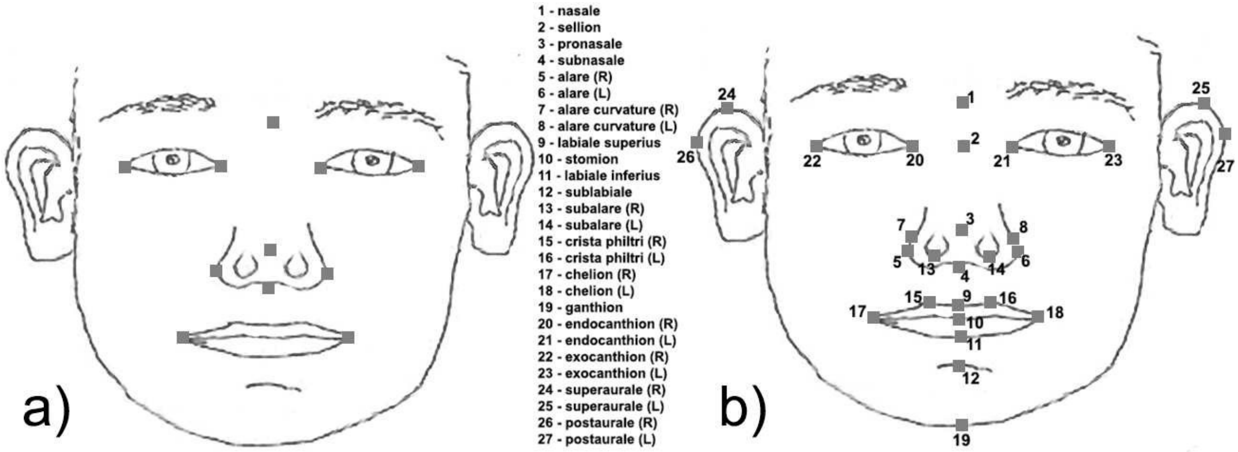


Fig. 1. a) Set of 11 anthropological landmarks (small square) used for face identification and face expression analysis b) Minor landmarks (small square) to improve accuracy

features, required consequent holistic extensive well trained classification.

The process of face landmark thorough detection should be introduced by an appropriate in-image face detection. Subsequently extracted landmarks can be used then for facial expression classification. Face localization and general face features extraction can be performed in tances between face regions, detected with originally elaborated features or Lienhart [17] different manner. Common solutions encompass deploying of a sequence of Haar-like features according to Viola-Jones algorithm [13], [18], eigenspaces [19] also referring to 3D model [20], skin color segmentation [21], statistical methods [22], [23] or active contour or shape models [24], [21].

Some authors [25], [26] exploited local binary patterns (LBP) idea for face image features extraction. Their extension: Local Direction Patterns (LDP) and locally assembled binary (LAB) features [27] appeared to be quite efficient for face edges detection and partly inspired authors of this paper for gradient distribution analysis.

If dedicated mouth features are further required, within face region analysis should be performed. Kim [28] and Chien [29] analyzed grid-based and coordinate-based lips features (width/height of outer/inner lips edges) for Korean language words recognition support. Matthews [30] exploited AAM and ASM for mouth visual shape description for lips reading. Shen [31] and Lewis [32] analyzed color space for lip features retrieving. He et al. [33] proposed modified Biologically Inspired Model of face features extraction improving the SVM classification of face smile. Su [34] suggested geometrical and Gabor filter retrieved face features fusion for better facial expression recognition.

Aforementioned approaches, though well studied and elaborated, lack of generic simplicity which lies in mouth lines detection. Presented method robustly extracts simplified lips edges by means of originally elaborated gradient-based approach which can be subsequently interpreted. Presented solution provides a representative set of features for mouth classification and consequent facial emotions recognition. It was additionally tested on a subset of Yale faces database [4]. It uses many of commonly used classifiers to show that presented method provides sufficient information which enables the detection or identification of face emotions.

III. Mouth features extraction method

To detect mouth shape features we need to find human face and relative localization of the mouth. Next, we can try extract mouth shape from it. Aggregated mouth features extraction process can be completed within subsequent steps:

1) Finding face,
2) Finding mouth on the face,
3) Mouth segmentation,
4) Features extraction.

In image face localization can be performed by means of Haar-like features method [17], but it will not be described here, because it is a part of another problem. Localization of mouth within the face region is a similar problem and can be completed with an analogical set of Haar-like features. That is why further, core mouth features analysis, assumes that face image is cropped to mouth with some border, as shown in fig. III

Mouth segmentation is then performed in several steps. These can be described as:

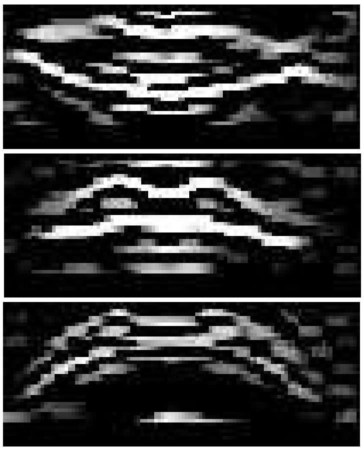
1) Gradient calculation,
2) Resulting image normalization,
3) Resulting image filtering,
4) Resulting image thresholding,
5) Noise removal.

To retrieve information from mouth images, at the first step **gradient** should be calculated. Gradient of an image is calculated with formula 1.

$$\nabla f = \begin{bmatrix} g_x \\ g_y \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} \tag{1}$$

a) b) c)

Fig. 2. Example of mouth images: (a) happy, (b) neutral, (c) sad, for subject no. 1



a) b) c)

Fig. 3. Example gradient for exemplary mouth images: (a) happy, (b) neutral, (c) sad, for subject no. 1

During the experiments, it was noticed that for descent mouth retrieval there is no need to calculate the whole gradient, but only its vertical part. Using both dimensions of gradient does not give noticeable results improvement, thus, in the segmentation process there was used only vertical component.

To calculate a vertical gradient g_y of image A we can use simple filter, as it is shown in equation 2.

$$g_y = \frac{\partial f}{\partial y} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} * A \tag{2}$$

Additionally, extended 3x3 matrix (eq. 3) was used to reduce noise, which introduces additional column on the left and on the right side of pixel position.

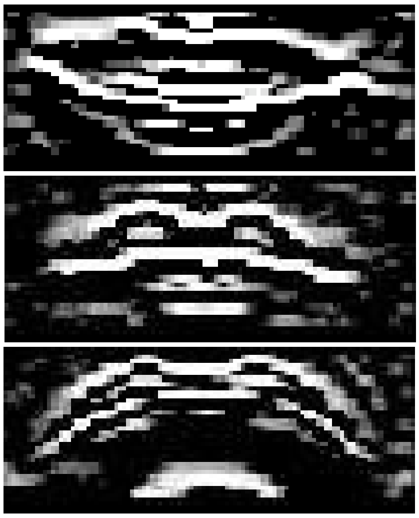
$$\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \Rightarrow \begin{bmatrix} -1 & \mathbf{-1} & -1 \\ 0 & \mathbf{0} & 0 \\ 1 & \mathbf{1} & 1 \end{bmatrix} \tag{3}$$

The first step results in images with vertical gradient calculated over the whole their area. Additionally, gradient values were squared in order to level up output and remove useless noise (fig. 3).

$$A = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \quad B = \begin{bmatrix} -3 & -10 & -3 \\ 0 & 0 & 0 \\ 3 & 10 & 3 \end{bmatrix}$$

$$C = \begin{bmatrix} -1 & -1 & -1 \\ 2 & 2 & 2 \\ -1 & -1 & -1 \end{bmatrix}$$

Fig. 4. Filter matrices used for edge extraction



a) b) c)

Fig. 5. Exemplary outputs obtained with matrix B from fig. 4: (a) happy, (b) normal, (c) sad, for subject no. 1

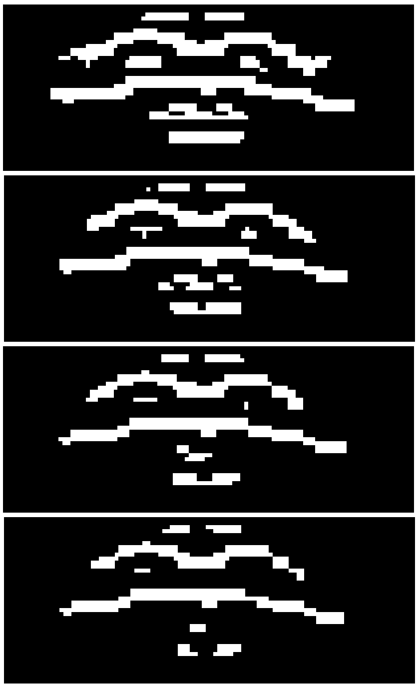
The next image processing step was the gradient **normalization**. MIN-MAX type normalization was used to adjust gradient to image full spectrum of brightness.

In the next step gradient image filtering was considered as to extract edges from it. Various filters was tested to extract shapes from images. The best results were obtained by filters represented in fig. 4. Corresponding outputs obtained with matrix B are presented in fig. 5.

After the filtering step a process of **thresholding** was applied to extract shape of lips. It was done by simple cut off image value below certain threshold of pixel brightness. Results were verified for a few threshold values, but overall the best was achieved with 250. Example results obtained with different values of threshold are shown in fig. 6

The subsequent step of the proposed method is the noise reduction. It was achieved by morphological operations performed on image. The best noise reduction was obtained for closing, which is a combination of erosion and dilation. In result shape of mouth was closed as it is shown in fig. 7.

The last stage concerned segmented image features extraction. Mouth corners were selected as initially considered features. They were found by the most extreme edges in all directions: down, left, right and up.

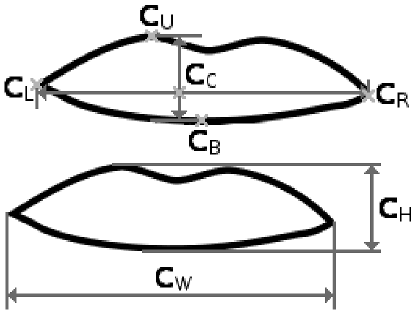


a)
b)
c)
d)

Fig. 6. Example results obtained with different threshold values: (a) 100, (b) 150, (c) 200 and (d) 250, for subject no. 1

a)
b)
c)
d)

Fig. 7. Example results with different noise reduction: (a) erosion, (b) dilatation, (c) opening, (d) closing, for subject no. 12



a)
b)

Fig. 8. Feature extraction corners: (a) positions, (b) sizes, for subject no. 1

Edge point c_e is defined as the farthest point in specified direction, as is shown in equation 4.

$$c_e = (x, y) \tag{4}$$

where

$$e = \{\ \mathtt{B} \text{ – bottom, } \mathtt{L} \text{ – left, } \mathtt{R} \text{ – right, } \mathtt{U} \text{ – topmost } \}$$

In case of the left edge $c_{\mathtt{L}}$, it is the most distant left point of the largest contour area. $c_{\mathtt{R}}$ can be defined analogously, but for the right edge. More difficult is to find the bottom and topmost edge points: $c_{\mathtt{B}}$ and $c_{\mathtt{U}}$, because often the contour is composed of several separated pieces, so we need to check if the contour size is sufficiently large in relation to the image size. Calculation of the shape size was simply done by following each white pixel connected to initially located edge points ($c_{\mathtt{L}}$ and $c_{\mathtt{R}}$). In case of contour inconsistence, averaged extreme values (bottom and topmost), retrieved from two separate edges, anchored independently from left and right corners were calculated.

The farthest edges are marked with white and light grey dashed lines. Brighter color is used to mark part of shape with features determining edges (fig. 8).

The collected information was sufficient to easily determine the next two parameters: $c_{\mathtt{W}}$ and $c_{\mathtt{H}}$. The first one determines the maximum distance stretching horizontally the mouth area, between $c_{\mathtt{L}}$ and $c_{\mathtt{R}}$, on the average height between them. Similarly value of $c_{\mathtt{H}}$ was determined.

With a **cross-section** $c_{\mathtt{C}}$ of width $c_{\mathtt{W}}$ and height $c_{\mathtt{H}}$, it is possible to interpret shape of the the mouth. Cross-section is always defined, because interpreted shape corners c_e are also determined.

Additionally, c_{h_i} as height of some part of mouth shape can be extracted similar to $c_{\mathtt{H}}$, but calculated as height between the top most point in shape and the lowest one for some i position horizontally.

IV. Features extraction

To evaluate the method a subset of the Yale's face database was used. It comes from UC San Diego Computer Vision [4]. It contains 165 grayscale images of 15 people, originally sized 320x243 pixels, but had manually cropped mouth region with size of 75x30 pixels. Each subject had several images in different face expressions, quality and light conditions, but for all people those conditions were the same. It allowed us to test elaborated method in various conditions of image quality.

Each image was marked by expert as shown in fig. 9: shape of mouth was built out of 6 points and it was marked by expert with white line, maximum width and maximum height. Marked elements were measured and noted as m_e, i.e. width as $m_{\mathtt{W}}$, analogically to computed values $c_{\mathtt{W}}$ from authors' method.

Exemplary results are shown in fig. 10. In analogy to the reference images, maximum widths (red) and heights (blue) are marked over the image.

Mouth key features evaluation method results c_e were subsequently compared with reference, expert marked points m_e (labeled data). For i-th image, each of key features c_e estimation accuracy ACC was calculated according to equation 5. For cross-section feature $c_{\mathtt{C}}$ estimation accuracy ACC was calculated according to equation 8, where `distance` is Manhattan distance between points and $\mathtt{distance}_{MAX}$ is the maximum possible distance on the image.

$$ACC_i(c_e) = \frac{\min\{c_e.x; m_e.x\} \times 100\%}{\max\{c_e.x; m_e.x\}} \text{ where } e = \{L, R\} \quad (5)$$

$$ACC_i(c_e) = \frac{\min\{c_e.y; m_e.y\} \times 100\%}{\max\{c_e.y; m_e.y\}} \text{ where } e = \{B, U\} \quad (6)$$

$$ACC_i(c_e) = \frac{\min\{c_e; m_e\} \times 100\%}{\max\{c_e; m_e\}} \text{ where } e = \{W, H\} \quad (7)$$

$$ACC_i(c_c) = 100\% - \frac{(\|c_c.x - m_c.x\| + \|c_c.y - m_c.y\|) \times 100\%}{75 + 30} \quad (8)$$

In table I selected features positioning accuracy results are presented. The results were calculated as averaged value of selected ($n = 66$) Yale database images individually estimated accuracies $ACC_i(c_e)$ (eq. 9).

$$\begin{gathered} ACC(c_e) = \frac{\sum_i ACC_i(c_e)}{n} \\ i \in \{1, 2, \ldots, n\}, \; e = \{L, R, B, U, W, H, C\} \end{gathered} \quad (9)$$

As extracted feature vector of single image was tuple of $c_L, c_U, c_R, c_B, c_W, c_H, c_C, c_{h_0} c_{h_n}$, where c_{h_i} means additionally height for a few positions, equally distributed over width of mouth. For tests n was 7.

The highest results of extraction accuracy were reported for left (81.42%), bottom (88.29%) and right (95.53%) part of mouth features detection. More problematic was upper part, what is strongly related to differences received in gradient of mouths. Resulting emotion characteristic determinants: mouth width (92.78%), height (72.77%) and cross-section point (86.18%) revealed also high evaluation accuracy. Weak mouth upper part features estimation influences negatively effectiveness of mouth shape determination, but can be counterbalanced by adding a number of measuring points for the lower edge and the height of shape characteristics measured for several positions, not only centrally located.

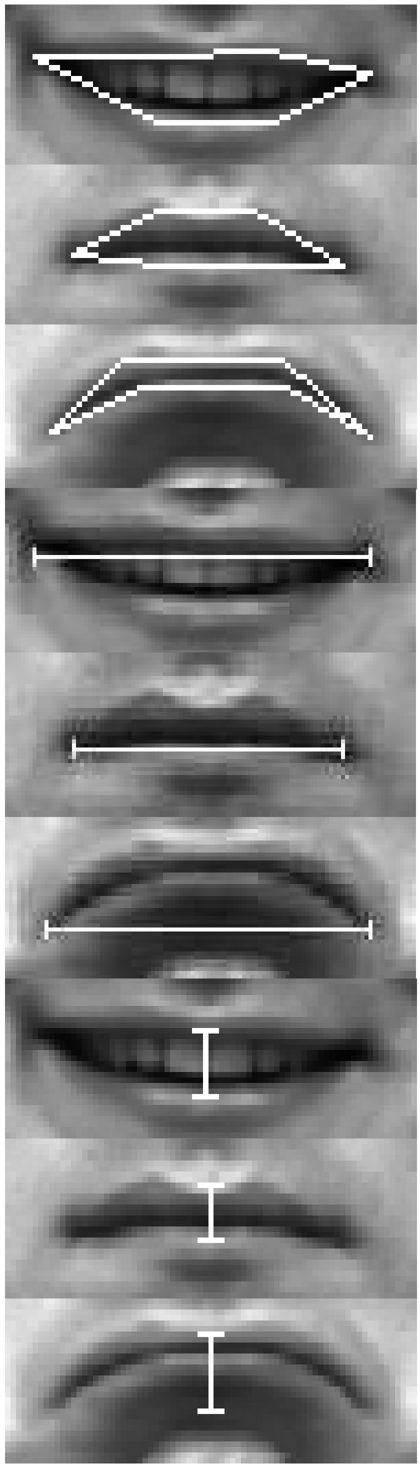


Fig. 9. Example labeled dataset with marked with white lines: shape, width and height

TABLE I
MOUTH FEATURES RECOGNITION RESULTS

Measurement	Size		Position				
	$ACC(e_W)$	$ACC(e_H)$	$ACC(e_L)$	$ACC(e_R)$	$ACC(e_U)$	$ACC(e_B)$	$ACC(e_C)$
Result [%]	92.8	72.8	81.4	95.5	43.6	88.3	86.2
Std dev. [%]	7.1	17.9	18.4	3.8	25.5	13.1	9.4

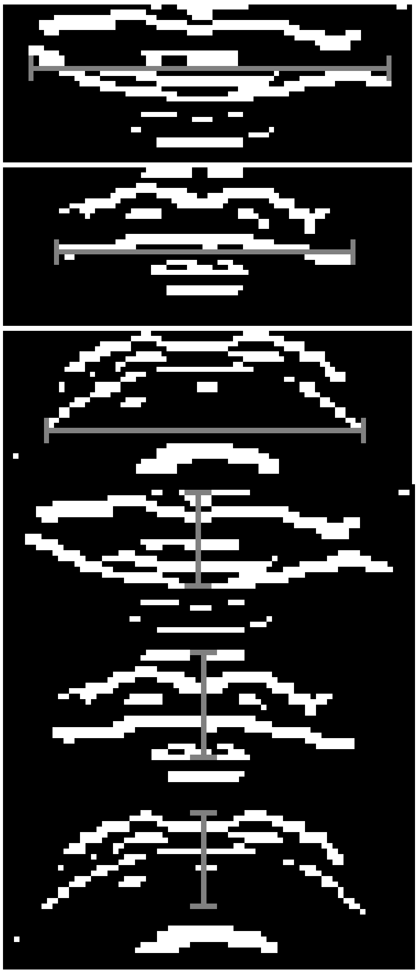


Fig. 10. Example results with marked with white lines: width and height

V. Emotion classification

After features' set extraction their classification is performed. Classifier predicts a label – one class from a few available. It is based on mathematical evaluation of input vector, due to selected classifier, and as output it has received a number, which represents assignment of input vector to some class [35], [36].

4 types of classical one-label classifiers were used for classification: Random Trees, *k*–Nearest Neighbors, Multi–Layer Perceptrons and Support Vector Machines. These are well known and widely used classifiers. There also others approaches like statistical methods based on Bayesian modelling combined with a Markov chain Monte Carlo (MCMC) sampling algorithm [37] etc., but was not considered in this paper. Weka was used for implementation, as a good base to train and test various classifiers with different options – explorer module [38].

At first, Random Trees (RT) was used, which was originally described to resolve classification and regression problems [39]. It constructs a tree that considers $log_2($ predicators $)+1$ randomly chosen attributes at each node, it doesn't perform pruning. It has used estimation of class probabilities basing on a hold-out set (backfitting).

K-nearest neighbors (*k*NN) classifier calculates distance between feature vector of input vector and others feature vectors from training set [40]. It has used Euclidean distance metrics, $k = 3$ and distance weighting with equation: $1/$distance.

Multi-Layer Propagation is a classifier that uses backpropagation to classify instances. This network is built by an algorithm: input layer = input features vector, hidden layer = input layer / 2 and output layer = number of classes. The nodes in this network are all sigmoid, except for output when the class is numeric in which case the nodes become unthresholded linear units.

Support Vector Machine (SVM) is a discriminative classifier formally defined by a separating hyperplanes. It has used linear kernel and C–SVC, because it is commonly used for n-class classification. In this research the libsvm library [41] was used.

The effectiveness of all methods was measured by Classification Accuracy (CA) as the basic evaluation measure. This measure provides a very precise evaluation because it reflects relation between set of correct labeled instances and all existing ones in the training set. It is defined as:

$$CA = \frac{1}{N}\sum_{N=0}^{N} y_i = f(x_i) \tag{10}$$

where: x_i are instances, i = $1..N$, N is their total number in the test set, y_i denotes the label of x_i and $f(x_i)$ is predicted label during classification process.

The experiments were carried out on popular database, which is often used to test methods working on face images. From selected images features vectors were extracted, which were used for classification.

Extracted features from previous section were divided into three groups of datasets:

1) Single – which was used for binary classification. All instances of features were labeled to belong to single class or not, for example instance can be classified as

TABLE II
EMOTION BINARY CLASSIFICATION WITH MOUTH FEATURES

	RT [%]	KNN [%]	MLP [%]	SVM [%]
happy	80.3	87.9	88.9	89.4
normal	71.2	75.8	75.8	83.3
sad	77.3	86.4	80.3	83.3
sleepy	75.8	77.3	81.8	83.3
surprised	72.7	86.4	84.8	83.3
wink	74.2	77.3	78.8	83.3

TABLE III
EMOTION CLASSIFICATION WITH MOUTH FEATURES

	RT [%]	KNN [%]	MLP [%]	SVM [%]
Three	53.0	37.9	37.9	45.5
Three2	30.3	31.8	28.8	30.3
Six	24.8	27.3	25.8	18.2

happy or unhappy (opposite emotion). There were six classes: happy, normal, sad, sleepy, surprised, wink.

2) Three – instances were grouped into three simple emotions: happy, normal and sad, where surprised was added to happy and wink to sad, because their vectors were similar.
3) Six – all instances were joined together in one dataset, but original labels were left for classification. This was the hardest dataset, which shows bad separation of instances.

Additionally, second Three dataset (named Three2) was used, which contains features of image space from the last step of features extraction, to better compare results.

Results of binary classification were shown in tab. II and from multi class classification in tab. III. For this context, the best recognized pose was happy, with almost 90% of accuracy detection for single emotion detection, over 70% True Positive (TP) detection in three–poses dataset and over 50% TP in six-poses dataset, with over 50% overall detection. It was high detectable emotion in all experiments.

The next poses which were well diagnosed these are surprised and sad. Detection rate was over 86% and 50% TP in six-poses dataset respectively. Other poses had much worse results, because individual classes were tightly mixed to each other, so they were hard to detect. Almost the same situation exists for binary classification, where some of the poses were hard detectable, but overall results is based on not classified non-poses, e.g. normal emotion was non detected for few of classifiers. SVM almost perfect detects non-poses for all emotions. Generally False Positive ration was very low or 0 for all tests.

VI. CONCLUSION

In this paper, it was proposed method to extract a few facial features from mouth image to expression recognition. The experiments were carried out on popular database, which is often used to test methods working on face images. As was shown in the last section, it gives overall good results, especially in horizontal measurement and they should be enough to right description of many facial expressions. In the future work it will be possible to collect other facial features and merge them together to better understand face and identify poses.

REFERENCES

[1] B. Fasel and J. Luettin, "Automatic facial expression analysis: a survey," *Pattern recognition*, vol. 36, no. 1, pp. 259–275, 2003.
[2] P. Ekman and W. Friesen, "Constants across cultures in the face and emotion," *Journal of Personality and Social Psychology*, vol. 17, no. 2, pp. 124–129, 1971.
[3] C. Chen, P. W. Woźniak, A. Romanowski, M. Obaid, T. Jaworski, J. Kucharski, K. Grudzień, S. Zhao, and M. Fjeld, "Using crowdsourcing for scientific analysis of industrial tomographic images," *ACM Trans. Intell. Syst. Technol.*, vol. 7, no. 4, pp. 52:1–52:25, Jul 2016. doi: 10.1145/2897370. [Online]. Available: http://doi.acm.org/10.1145/2897370
[4] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE TRANS. PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, vol. 19, no. 7, pp. 711–720, 1997.
[5] S. Liang, J. Wu, S. M. Weinberg, and L. G. Shapiro, "Improved detection of landmarks on 3d human face data," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, July 2013. doi: 10.1109/EMBC.2013.6611039. ISSN 1094-687X pp. 6482–6485. [Online]. Available: http://dx.doi.org/10.1109/EMBC.2013.6611039
[6] S. Mishra and A. Dhole, "A survey on facial expression recognition techniques," *International Journal of Science and Research*, vol. 4, no. 4, pp. 1247–1250, 2015.
[7] E. Hjelmås and B. K. Low, "Face detection: A survey," *Computer vision and image understanding*, vol. 83, no. 3, pp. 236–274, 2001.
[8] C. Zhang and Z. Zhang, "A survey of recent advances in face detection," Tech. rep., Microsoft Research, Tech. Rep., 2010.
[9] V. Pali, S. Goswami, and L. Bhaiya, "An extensive survey on feature extraction techniques for facial image processing," in *Sixth International Conference on Computational Intelligence and Communication Networks*, 2014. doi: 10.1109/CICN.2014.43 pp. 142–148.
[10] M. Castrillón-Santana, D. Hernández-Sosa, and J. Lorenzo-Navarro, "Combining face and facial feature detectors for face detection performance improvement," in *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, ser. Lecture Notes in Computer Science, L. Alvarez, M. Mejail, L. Gomez, and J. Jacobo, Eds., 2012, vol. 7441, pp. 82–89.
[11] T. Kanade, *Computer recognition of human faces*. Birkhäuser, 1977, vol. 47.
[12] R. Brunelli and T. Poggio, "Face recognition: Features versus templates," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, no. 10, pp. 1042–1052, 1993.
[13] M. Castrillón, O. Déniz, D. Hernández, and J. Lorenzo, "A comparison of face and facial feature detectors based on the viola–jones general object detection framework," *Machine Vision and Applications*, vol. 22, no. 3, pp. 481–494, 2011.
[14] M.-T. Yang, Y.-J. Cheng, and Y.-C. Shih, *Facial Expression Recognition for Learning Status Analysis*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 131–138. ISBN 978-3-642-21619-0. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-21619-0-18
[15] A. Panning, A. K. Al-Hamadi, R. Niese, and B. Michaelis, "Facial expression recognition based on haar-like feature detection," *Pattern Recognition and Image Analysis*, vol. 18, no. 3, pp. 447–452, 2008.
[16] Q. Wang, C. Zhao, and J. Yang, "Robust facial feature location on gray intensity face," in *PSIVT 2009*, ser. LNCS, T. Wada, F. Huang, and S. Lin, Eds., vol. 5414. Springer, 2009. doi: 10.1007/978-3-540-92957-4-47 p. 542–549. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-92957-4-47
[17] R. Lienhart and J. Maydt, "An extended set of haar-like features for rapid object detection," in *Image Processing. 2002. Proceedings. 2002 International Conference on*, vol. 1, 2002, pp. I–900.
[18] P. Viola and M. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, pp. 137–154, 2004.

[19] A. Pentland, B. Moghaddam, and T. Starner, "View-based and modular eigenspaces for face recognition," in *IEEE Conference on Computer Vision and Pattern Recognition, Seattle, NA, USA*, 1994, pp. 84–91.

[20] W. Yang, C. Sun, W. Zheng, and K. Ricanek, "Gender classification using 3D statistical models," *Multimedia Tools and Applications*, Mar 2016. doi: 10.1007/s11042-016-3446-7. [Online]. Available: http://dx.doi.org/10.1007/s11042-016-3446-7

[21] Y.-H. Lee, C. G. Kim, Y. Kim, and T. K. Whangbo, "Facial landmarks detection using improved active shape model on android platform," *Multimedia Tools and Applications*, vol. 74, no. 20, p. 8821–8830, Jun 2013. doi: 10.1007/s11042-013-1565-y. [Online]. Available: http://dx.doi.org/10.1007/s11042-013-1565-y

[22] H. Schneiderman and T. Kanade, "A statistical method for 3d object detection applied to faces and cars," in *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*, vol. 1. IEEE, 2000, pp. 746–751.

[23] M. A. Berbar, "Three robust features extraction approaches for facial gender classification," *Vis Comput*, vol. 30, no. 1, p. 19–31, Jan 2013. doi: 10.1007/s00371-013-0774-8. [Online]. Available: http://dx.doi.org/10.1007/s00371-013-0774-8

[24] M. Kass, A. Witkin, and D. Terzepulos, "Snakes: Active contour models," in *First IEEE International Conference on Computer Vision*, 1987, pp. 259–268.

[25] A. Hussain, M. S. Khan, M. Nazir, and M. A. Iqbal, "Survey of various feature extraction and classification techniques for facial expression recognition," in *Proceedings of the 11th WSEAS international conference on Electronics, Hardware, Wireless and Optical Communications, and proceedings of the 11th WSEAS international conference on Signal Processing, Robotics and Automation, and proceedings of the 4th WSEAS international conference on Nanotechnology*, 2012, pp. 138–142.

[26] T. Jabid, M. H. Kabir, and O. Chae, "Robust facial expression recognition based on local directional pattern," *ETRI journal*, vol. 32, no. 5, pp. 784–794, 2010.

[27] S. Yan, S. Shan, X. Chen, and W. Gao, "Locally assembled binary (lab) feature with feature-centric cascade for fast and accurate face detection," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, 2008, pp. 1–7.

[28] Y.-K. Kim, J. G. Lim, and M.-H. Kim, "Comparison of lip image feature extraction methods for improvement of isolated word recognition rate," Aug 2015. doi: 10.14257/astl.2015.107.14. [Online]. Available: http://dx.doi.org/10.14257/astl.2015.107.14

[29] S.-I. Chien and I. Choi, "Face and facial landmarks location based on log-polar mapping," *Lecture Notes in Computer Science*, p. 379–386, 2000. doi: 10.1007/3-540-45482-9-38. [Online]. Available: http://dx.doi.org/10.1007/3-540-45482-9-38

[30] I. Matthews, T. Cootes, J. Bangham, S. Cox, and R. Harvey, "Extraction of visual features for lipreading," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 24, no. 2, p. 198–213, 2002. doi: 10.1109/34.982900. [Online]. Available: http://dx.doi.org/10.1109/34.982900

[31] X.-g. Shen and W. Wu, "An algorithm of lips secondary positioning and feature extraction based on ycbcr color space," in *International Conference on Advances in Mechanical Engineering and Industrial Informatics*. Atlantis Press, 2015. doi: 10.2991/ameii-15.2015.271. [Online]. Available: http://dx.doi.org/10.2991/ameii-15.2015.271

[32] T. W. Lewis and D. M. W. Powers, "Lip feature extraction using red exclusion," in *Selected Papers from the Pan-Sydney Workshop on Visualisation - Volume 2*, ser. VIP '00. Darlinghurst, Australia, Australia: Australian Computer Society, Inc., 2001. ISBN 0-909-92580-1 pp. 61–67. [Online]. Available: http://dl.acm.org/citation.cfm?id=563752.563761

[33] C. He, H. Mao, and L. Jin, "Realistic smile expression recognition using biologically inspired features," in *AI 2011*, ser. LNAI, D. Wang and M. Reynolds, Eds., vol. 7106. Springer, 2011, p. 590–599.

[34] C. Su, J. Deng, Y. Yang, and G. Wang, "Expression recognition methods based on feature fusion," in *BI 2010*, ser. LNAI, Y. Yao *et al.*, Eds., vol. 6334. Springer, 2010, p. 346–356.

[35] M. Krzyśko, W. Wołyński, T. Górecki, and M. Skorzybut, "Systemy uczące się," *Rozpoznawanie wzorców, analiza skupień i redukcja wymiarowości. WNT, Warszawa*, 2008.

[36] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd ed. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2011. ISBN 0123748569, 9780123748560

[37] K. Grudzien, A. Romanowski, and R. A. Williams, "Application of a bayesian approach to the tomographic analysis of hopper flow," *Particle & Particle Systems Characterization*, vol. 22, no. 4, pp. 246–253, 2005. doi: 10.1002/ppsc.200500951. [Online]. Available: http://dx.doi.org/10.1002/ppsc.200500951

[38] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The weka data mining software," *ACM SIGKDD Explorations Newsletter*, vol. 11, no. 1, p. 10, Nov 2009. doi: 10.1145/1656274.1656278. [Online]. Available: http://dx.doi.org/10.1145/1656274.1656278

[39] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001. doi: 10.1023/A:1010933404324. [Online]. Available: http://dx.doi.org/10.1023/A:1010933404324

[40] D. W. Aha, D. Kibler, and M. K. Albert, "Instance-based learning algorithms," *Machine Learning*, vol. 6, no. 1, p. 37–66, 1991. doi: 10.1023/a:1022689900470. [Online]. Available: http://dx.doi.org/10.1023/A:1022689900470

[41] C.-C. Chang and C.-J. Lin, "Libsvm - a library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, p. 1–27, Apr 2011. doi: 10.1145/1961189.1961199. [Online]. Available: http://dx.doi.org/10.1145/1961189.1961199