Feature-Based Detection of Facial Landmarks from Neutral and Expressive Facial Images

Yulia GUIZATDINOVA and Veikko SURAKKA

Abstract--**Feature-based method for detecting landmarks from facial images was designed. The method was based on extracting oriented edges and constructing edge maps at two resolution levels. Edge regions with characteristic edge pattern formed landmark candidates. The method ensured invariance to expressions while detecting eyes. Nose and mouth detection was deteriorated by happiness and disgust.**

Index Terms-- **I. Computing Methodologies, I.4 Image Processing and Computer Vision, I.4.6 Segmentation, I.4.6.a Edge and feature detection.**

I. Introduction

Automated detection and segmentation of a face have been active research topics for the last few decades. The motivation behind developing systems of face detection and segmentation is a great number of its applications. For example, detection of a face and its features is an essential requirement for face and facial expression recognition [1]-[4].

Due to such factors as illumination, head pose, expression, and scale the facial features vary greatly in their appearance. Yacoob, Lam, and Davis [5] demonstrated that facial expressions are particularly important factors affecting automated detection of facial features. They aimed to compare the recognition performance of template- and feature-based approaches to face

recognition. Both approaches resulted in worse recognition performance for expressive images than for neutral ones.

Facial expressions as emotionally or otherwise socially meaningful communicative signals have been intensively studied in psychological literature. Ekman and Friesen [6] developed the Facial Action Coding System (FACS) for coding all visually observable changes in the human face. According to FACS, a muscular activity producing changes in facial appearance is coded in the terms of action units (AU). Specific combinations of AUs represent prototypic facial displays: neutral, happiness, sadness, fear, anger, surprise, and disgust [7]. At present, there is good empirical evidence and good theoretical background for analyzing how different facial muscle activations modify the appearance of a face during emotional and social reactions [8]-[10].

Studies addressing problem of automated and expression-invariant detection of facial features have been recently published. In particular, to optimize feature detection some attempts have been made to utilize both profound knowledge on human face and its behavior and modern imaging techniques. Comprehensive literature overviews on different approaches to face and facial feature detection have been published by Hjelmas and Low [11] and Yang, Kriegman, and Ahuaja [12].

Liu, Schmidt, Cohn, and Mitra [13] investigated facial asymmetry under expression variation. The analysis of facial asymmetry revealed individual differences that were relatively unaffected by changes in facial expressions. Combining asymmetry information and conventional template-based methods of face identification they achieved a high rate of error reduction for face classification.

Tian, Kanade, and Cohn [14] developed a method for recognizing several specifically chosen

V. Surakka is with the Research Group for Emotions, Sociality, and Computing, Tampere Unit for Computer-Human Interaction, Department of Computer Sciences, University of Tampere, Tampere, FIN-33014 (telephone: +358-03-2158551, e-mail: Veikko.Surakka@uta.fi). He works

AUs and their combinations. They analyzed both stable facial features as landmarks and temporal facial features like wrinkles and furrows. The reported recognition rates were high for recognizing AUs from both upper and lower part of a face.

Golovan [15] proposed a feature-based method for detecting facial landmarks as concentrations of the points of interest. The method demonstrated high detection rate and invariance to changes in image view and size while detecting facial landmarks. However, the method was not tested with databases of carefully controlled facial expressions. We extended the method introduced by Golovan to detect facial landmarks from expressive facial images. In this framework, the aim of the present study was to experimentally evaluate the sensitivity of the developed method while systematically varying facial expression and image size.

II. DATABASE

The Pictures of Facial Affect database [16] was used to test the method developed for detection of facial landmarks. The database consists of 110 images of 14 individuals (i.e. 6 males and 8 females) representing neutral and six prototypical facial expressions of emotions: happiness, sadness, fear, anger, surprise and disgust [7]. On average, there were about 16 pictures per expression. In order to test the effects of image resizing on the operation of the developed method, the images were manually normalized to three preset sizes (i.e. 100×150, 200×300, and 300×450 pixels). In sum, $110 \times 3 = 330$ images were used to test the method.

III. FACIAL LANDMARK DETECTION

The regions of eyebrow-eyes, lower nose, and mouth were selected as facial landmarks to be detected. There were two enhancements to the method proposed in previous works [15], [17]. The first enhancement is the reduction of number of edge orientations used for constructing edge maps of the image. In particular, the orientations ranging from 45° to 135° and 225° to 315° in

step of 22.5° were used to detect facial landmarks (Fig. *1*). The chosen representation of edge orientations described relatively well facial landmarks and reduced a computational load of the method. The second enhancement is the construction of the orientation model of facial landmarks. The landmark model was used to verify the existence of a landmark in the image.
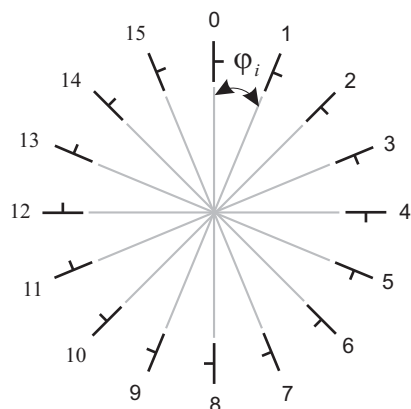
Fig. 1. Orientation template for extracting local oriented edges, $\varphi_i = i \cdot 22.5°$, $i = 0 \div 15$. Edge orientations used for detecting facial landmarks were marked as numbers $2 \div 6$ and $10 \div 14$.

The method was implemented through three stages: preprocessing, edge map constructing, and orientation matching. These stages will be described in details in the following sections.

A. Preprocessing

First, an image is transformed into the grey-level representation. To eliminate noise edges and remove small details the grey-level image is then smoothed by the recursive Gaussian
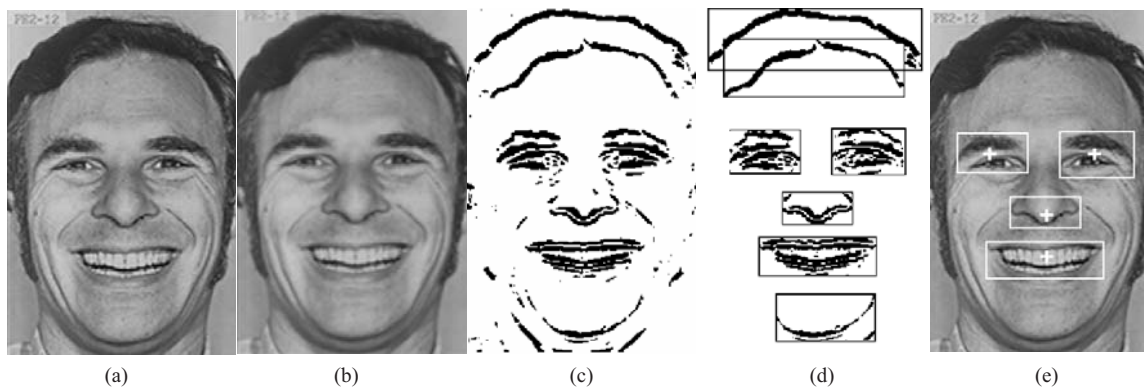
Fig. 2. Landmark detection: (a) image of happiness; (b) smoothed image ($\sigma = 0.8$); (c) extracted oriented edges ($\sigma = 1.2$); (d) landmark candidates; (e) facial landmarks and their centres of mass. Image from Pictures of Facial Affect. Copyright © 1976 by Paul Ekman. Reprinted with permission of Paul Ekman.

transformation. The smoothed images are used to detect all possible candidates for facial landmarks, and no smoothed images - to analyse the landmark candidates in details (Fig. *2a, b*). In that way, the amount of information that is processed at high resolution level is significantly reduced.

B. Edge Map Constructing

Local oriented edges are extracted by convolving smoothed image with a set of 10 kernels. Each kernel is sensitive to one of 10 chosen orientations. The whole set of 10 kernels results from differences between two oriented Gaussians with shifted kernels.

$$G_{\varphi_k} = \frac{1}{Z}(G^-_{\varphi_k} - G^+_{\varphi_k}),\qquad(1)$$

$$Z = \sum_{p,q}(G^-_{\varphi_k} - G^+_{\varphi_k}),\quad G^-_{\varphi_k} - G^+_{\varphi_k} > 0,\qquad(2)$$

$$G^-_{\varphi_k} = \frac{1}{2\pi\sigma^2} e^{-\frac{(p-\sigma\cos\varphi_k)^2+(q-\sigma\sin\varphi_k)^2}{2\sigma^2}},\qquad(3)$$

$$G^+_{\varphi_k} = \frac{1}{2\pi\sigma^2} e^{-\frac{(p+\sigma\cos\varphi_k)^2+(q+\sigma\sin\varphi_k)^2}{2\sigma^2}},\qquad(4)$$

where σ is a root mean square deviation of the Gaussian distribution; φ_k is angle of the Gaussian rotation; $\varphi_k = k\cdot22.5°$; $k = 2,3,4,5,6,10,11,12,13,14$; $p, q = -3,-2,-1,0,1,2,3$.

The maximum response of all 10 kernels defines the contrast magnitude of a local edge at its pixel location. The orientation of a local edge is estimated with orientation of a kernel that gave the maximum response.

$$g_{ij\varphi_k} = \sum_{p,q} b^{(l)}_{i-p,j-q} G_{\varphi_k},\qquad(5)$$

where b denotes the grey level of the image at pixel (i,j); $i = 0\div W\text{-}1$; $j = 0\div H\text{-}1$; W, H are respectively width and height of the image, $l = 1,2$.

The threshold for contrast filtering of the extracted edges is determined as an average contrast of the whole smoothed image. Edge grouping is based on the neighborhood distances between oriented edges and is limited by a possible number of neighbors for each edge. The optimal thresholds for edge grouping are determined using small image set taken from the database. In such a way, the edge map of the smoothed image (i.e. $l = 2$) consists of the regions of edge concentrations presumed to contain facial landmarks. Fig. *2c* presents the primary feature map that was constructed by detecting local edges of 10 chosen orientations. Fig. *2d* shows the primary map after contrast thresholding and grouping extracted edges into the candidates for facial landmarks.

To get more detailed description of the extracted edge regions, edge extracting and edge grouping are applied to high resolution image (i.e. $l = 1$) within the limits of these regions. In this case, the threshold for contrast filtering is determined as a double average contrast of the high resolution image.

C. Orientation Matching

We analyzed orientation portraits of edge regions extracted from 12 expressive faces of the same person. On the one hand, expressions do not affect specific distribution of the oriented edges contained in regions of facial landmarks (Fig. *3a*). On the other hand, noise regions have arbitrary distribution of the oriented edges (Fig. *3b*).

Finally, we created four average orientation portraits for each facial landmark. Average orientation portraits keep the same specific pattern of the oriented edges as individual ones (Fig. *4*).
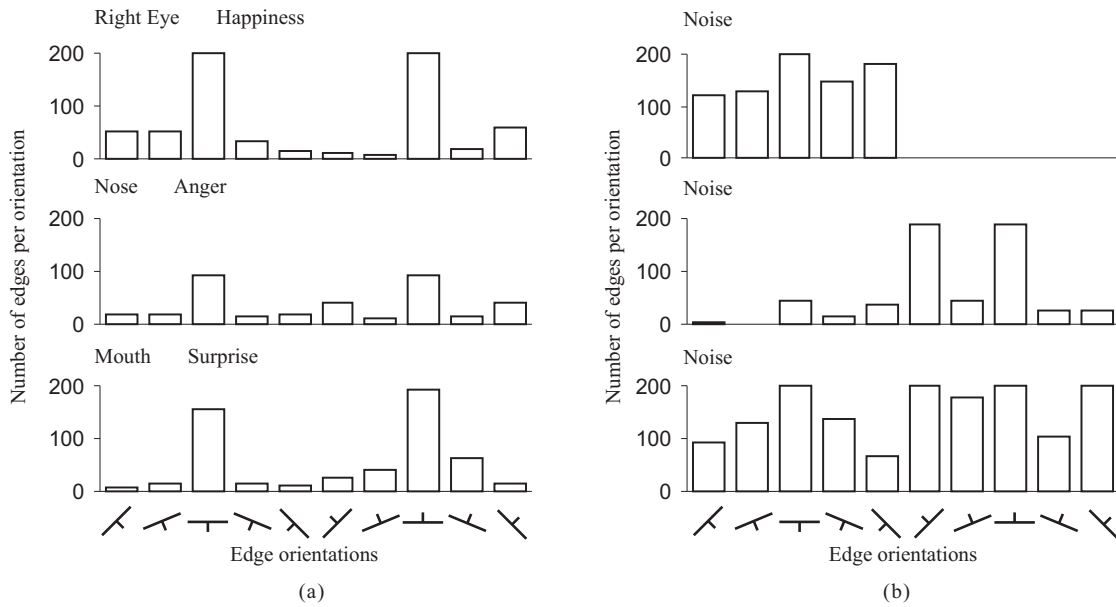
Fig. 3. Individual orientation portraits of (a) facial landmarks with specific distribution of the oriented edges, and (b) noise regions with arbitrary distribution of the oriented edges.
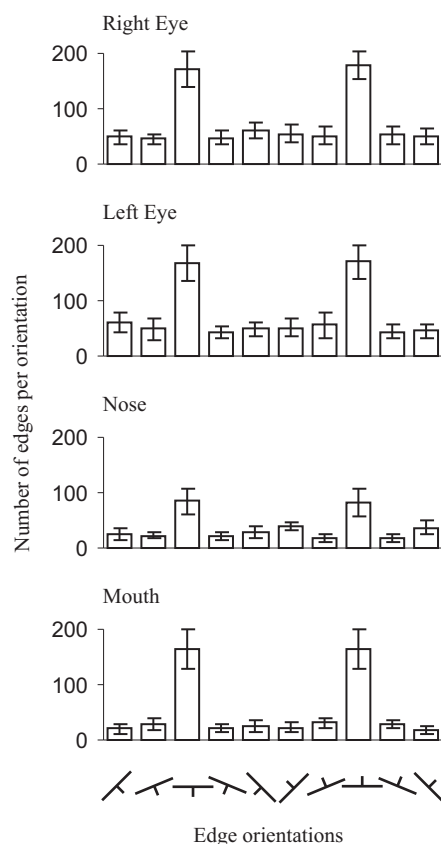
Fig. 4. Average orientation portraits of landmarks with specific distribution of the oriented edges. The error bars show plus/minus one standard deviation from the mean values.

1) Orientation Model

Such findings allowed us to design the characteristic orientation model for all four facial landmarks. The following rules define the structure of the orientation model: 1) horizontal orientations are represented by the greatest number of extracted edges; 2) a number of edges corresponding to each of horizontal orientations is more than 50% greater than a number of edges corresponding to other orientations taken separately; and 3) orientations cannot be presented by zero number of edges.

The detected candidates for facial landmarks are manually classified into one of the following groups: noise or facial landmark like eye, nose, and mouth. Fig. *2e* reveals the final feature map consisting of candidates whose orientation portraits match with the orientation model.

IV. RESULTS

Fig. *5* illustrates examples of the landmark detection from neutral and expressive facial images.

On the stage of edge map constructing, an average number of candidates per image was 8.35
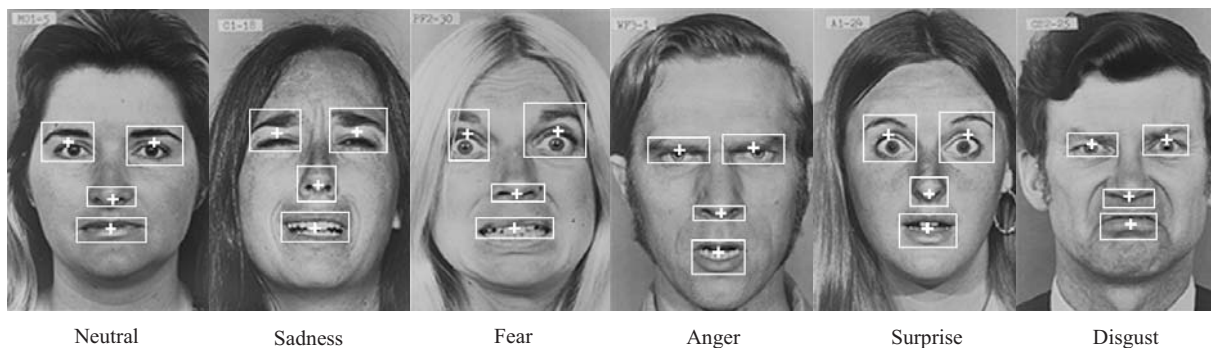
Fig. 5. Detected facial landmarks and their centres of mass. Images from Pictures of Facial Affect. Copyright ©1976 by Paul Ekman. Reprinted with permission of Paul Ekman.

and did not vary significantly by changes in facial expression and image size. After the orientation matching, the average number of candidates per image was reduced to almost a half and amounted to 4.52. Fig. *6* illustrates the decrease in the number of candidates per image averaged over different facial expressions.
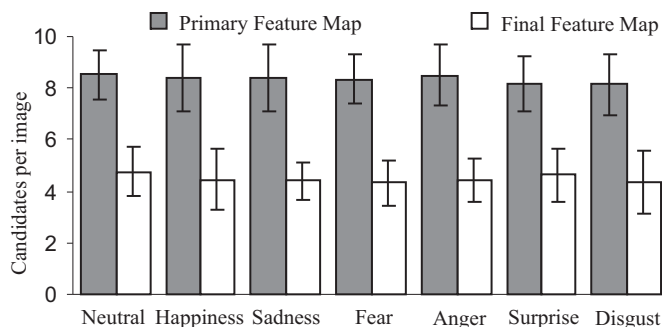
Fig. 6. Average number of candidates per image before and after the procedure of orientation matching. The error bars show plus/minus one standard deviation from the mean values.

Table I shows that the developed method revealed an average detection rate of 90% in detecting all 4 facial landmarks from both neutral and expressive images. The average detection rates were 94% and 90% for neutral and expressive images, respectively. The detection of nose

and mouth was more affected by facial expressions than the detection of eyes.

TABLE I
RATE (%) OF THE LANDMARK DETECTION AVERAGED OVER EXPRESSION AND
IMAGE SIZE

Image type		Facial landmark				Average
		Left eye	Right eye	Mouth	Nose	
Neutral		100	100	91	86	94
Expressive	Happiness	100	100	82	74	89
	Sadness	96	96	98	88	95
	Fear	100	100	93	82	94
	Anger	100	100	92	84	94
	Surprise	100	100	98	78	94
	Disgust	96	96	49	51	73
Average		99	99	86	78	90

The Expressive average is 90.

Both eyes were detected with a high detection rate from nearly all types of the images. In general, the correct detection of eyes did not require a strong contrast between the whites of eyes and iris. In such a way, eyes were found correctly regardless of whether the whites of eyes were visible or not (Fig. *5*). However, expressions of sadness and disgust reduced the average detection rate to 96%. The correct eye localization was only slightly affected by variations in image size.

Regions of both eyes had nearly the same number of the extracted oriented edges. About one third of a total number of edges were extracted from the regions of eyebrows. As a result, the mass centres of the eye regions slightly shifted up from the iris centres (Fig. *5*).

Detection of the mouth region was more affected by changes in facial expression and image size than detection of the eye regions. On average, the correct location of the mouth region was found in more than 90% of the expressive images with the exception of happiness (82%) and disgust images (49%). The smallest image size had a marked deteriorating effect on the mouth detection. However, the within-expression variations in the shape of the mouth had only a small influence on the ability of the method to mark the correct area. As a rule, the mouth region was

found regardless of whether mouth was open or closed and whether the teeth were visible or not.

The nose detection was even more affected by variations in facial expression and image size than the mouth detection. The expressions of happiness, surprise and disgust had the biggest deteriorating effect on the detection of the nose region. The average detection rate for nose region was 74% for happiness, 78% for surprise, and 51% for disgust images. It was more than 81% for other expressive images. In sum, the images expressing disgust was considered as the hardest to process.

There were three types of errors in detecting facial landmarks. Fig. *7* gives examples of such errors. The undetected facial landmarks were considered to be the errors of the first type. Such errors occurred when a region of interest including facial landmark was rejected as a noise region. In particular, the nose was the most undetectable facial landmark (Fig. *7a*). The incorrectly grouped landmarks were regarded as the errors of the second type. The most common error of the second type was grouping regions of nose and mouth in one region (Fig. *7b* and Fig. *7c*). There were only few cases of grouping together eye regions (Fig. *7c*). The errors of the third type were the misdetected landmarks that occurred when the method accepted noise regions as facial landmarks (Fig. *7a*).
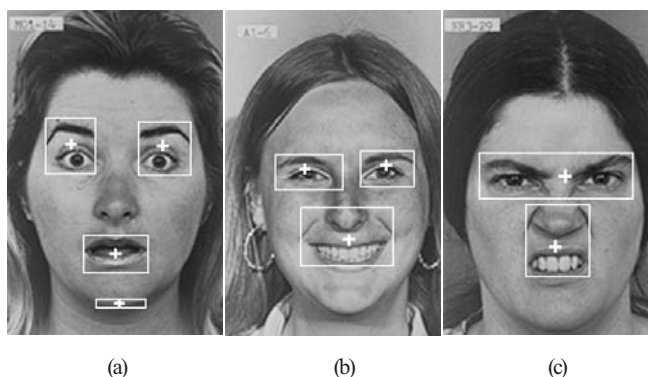

(a) (b) (c)

Fig. 7. Errors in detection of facial landmarks. Images from Pictures of Facial Affect. Copyright ©1976 by Paul Ekman. Reprinted with permission of Paul Ekman.

V. DISCUSSION

The feature-based method for detecting facial landmarks from neutral and expressive facial images was designed. The method achieved the average detection rate of 90% in extracting all four facial landmarks from both neutral and expressive images. The separate percentages were 94% for neutral images and 90% for expressive ones. The present results revealed that the choice of the oriented edges as the basic features for composing edge maps of the image ensured the invariance in a certain range for eye detection regardless of variations in facial expression and image size. The regions of left and right eyes were detected in 99% of the cases.

However, detecting landmarks of the lower face was affected by changes in expression and image size. The expressions of happiness and disgust had a marked deteriorating effect on detecting regions of nose and mouth. The decrease of image size also affected the detection of these landmarks. Variations in expression and decrease in image size attenuated the average detection rates of mouth and nose regions to 86% and of 78%, respectively.

The results showed that a majority of errors in detecting facial landmarks occurred at the stage of feature map construction. On the one hand, the results revealed that often the nose region remained undetected after the procedure of edge extraction. One possible reason for that was a low contrast of nose regions on the images. As a result, the number of edges extracted from the nose regions was smaller than those extracted from the regions of other landmarks. On the other hand, the threshold limiting number of edges was elaborated for detecting all four facial landmarks. Possibly for this reason, the nose region consisting of a small number of edges remained undetected.

Another reason for errors in detection of nose as well as mouth was the decrease in image size. The decrease in image size did not affect the contrast around the eyes, but it reduced the contrast around the nose and mouth. Therefore, the number of edges extracted from these regions was reduced and they became less than the threshold and, finally, nose and mouth regions remained

undetected.

On the other hand, the procedure of grouping edges into candidates produced incorrect grouping of several landmarks into the one region. Many errors in constructing regions of nose and mouth were caused by the use of a fixed neighborhood distance for edge grouping. Utilizing fixed threshold produced a good landmark separation for almost all expressive images (i.e. the error rate in landmark grouping was less than 1%). However, the images of happiness and disgust produced a lot of errors in landmark grouping (i.e. the error rates were about 2% and 5%, respectively). This means that such a fixed neighborhood distance cannot be applied for separating regions of nose and mouth from the happiness and disgust images.

Why the expressions of happiness and disgust were especially difficult to process by the developed algorithms? Probably, the reasons for that were the specific changes of facial appearance while displaying these expressions. There are different AUs and their combinations that are activated during happiness and disgust. In particular, when a face is modified by the expression of happiness, the AU12 is activated. This AU pulled the lips back and obliquely upward.

Further, many of the prototypical disgust expressions suggested by Ekman and Friesen [6] include the activation of AU10. The AU10 lifts the centre of the upper lip upwards making the shape of the mouth resemble an upside down curve. Both AU10 and AU12 result in deepening nasolabial furrow and pulling it laterally upwards.

Although, there are marked differences in the shape of the nasolabial deepening and mouth shaping for these two AUs, it can be summed up that both expressions of happiness and disgust make the gap between nose and mouth smaller. Such modifications in facial appearance had a marked deteriorating effect on detecting landmarks from the lower part of a face. The neighborhood distances between edges extracted from the regions of nose and mouth became

smaller than a threshold. For this reason edges were grouped together resulting in incorrect grouping of nose and mouth regions. The expressions of disgust and sadness (i.e. the combination AU1 and AU4) caused the regions of eyebrows to draw up together, resulting in incorrect grouping regions of both eyes.

One possible way to eliminate errors in landmark separation could be a precise analysis of the density of the edges inside the detected edge regions. The areas with poor point density might contain different areas of edge concentration and could be processed further with some more effective methods like, for example, neighborhood method.

On the stage of orientation matching, there were some errors in classification between landmark and noise regions. Although the orientation model revealed a high classification rate for both eyes, it produced errors in classifying nose region. Such errors were caused by mismatching orientation portraits of the detected candidates and the orientation model. For example, in some cases the nose region did not have well defined horizontal dominants in edge orientations – all edge orientations were presented in nearly equal number. Therefore, such a region was rejected as a candidate for facial landmark. On the other hand, errors were caused by the fact that orientation portraits of some noise regions matched the orientation model. In this case, the noise regions were detected. However, most of errors in landmark detection were brought about by errors in the previous stage of feature map construction.

Based on the findings described above we can conclude that more accurate nose and mouth detection could be achieved by finding some adaptive thresholds for constructing landmark candidates. The overall detection performance of the algorithms could be improved significantly by analysing spatial configuration of the detected facial landmarks. The use of spatial constraints might be also utilized to predict the location of the undetected facial landmarks [18].

In summary, the method localized facial landmarks with acceptably high detection rate without

a combinatorial increase of complexity of the image processing algorithms. The detection rate of the method was comparable with the detection rate of the known feature-based [15], [17] and color-based [19] methods that have detection rates from 85% to 95%, but lower than neural network-based methods [20] with a detection rate of about 96-99.5%. Emphasizing simplicity of the algorithms developed for landmark detection, we conclude they might be implemented as a part of the systems for face and/or facial expression recognition. The discovered errors provided several guidelines for further improvement of the developed method. In our future work, we will focus on finding expression-invariant and robust representations for facial landmarks. Careful attention will be paid to the development of algorithms that are able to cope with images displaying happiness and disgust as the most demanding to process.

REFERENCES

[1] A. Pentland, B. Moghaddam, and T. Starner, "View-based and modular Eigenspaces for face recognition," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Seattle, WA, June 1994, pp. 84-91.

[2] L. Wiskott, J-M. Fellous, N. Kruger, and C. von der Malsburg, "Face recognition by elastic bunch graph matching," *IEEE Trans. PAMI.*, vol. 19, pp. 775-779, July 1997.

[3] G. Donato, M. Bartlett, J. Hager, P. Ekman, and T. Sejnowski, "Classifying facial actions," *IEEE Trans. PAMI*, vol. 21, pp. 974–989, Oct. 1999.

[4] I. Essa and A. Pentland, "Coding, analysis, interpretation, and recognition of facial expressions," *IEEE Trans. PAMI*, vol. 19, pp. 757-763, July 1997.

[5] Y. Yacoob, H-M. Lam, and L. Davis, "Recognizing faces showing expressions," in *Proc. Int. Workshop Automatic Face- and Gesture-Recognition*, Zurich, Switzerland, June 1995, pp. 278-283.

[6] P. Ekman and W. Friesen, *Facial action coding system (FACS): A technique for the measurement of facial action*. Palo Alto, Calif.: Consulting Psychologists Press, 1978.

[7] P. Ekman, "The argument and evidence about universals in facial expressions of emotion." *Handbook of Social Psychophysiology*, H. Wagner and A. Manstead, eds. Lawrence Erlbaum, 1989, pp. 143-164.

[8] P. Ekman, W. Friesen, and J. Hager, *Facial action coding system (FACS)*. Salt Lake City, UTAH: A Human Face, 2002.

[9] A. Fridlund, "Evolution and facial action in reflex, social motive, and paralanguage," *J. Biological Psychology*, vol. 32, pp. 3-100, Feb. 1991.

[10] V. Surakka and J. Hietanen, "Facial and emotional reactions to Duchenne and non-Duchenne smiles," *Int. J. Psychophysiology*, vol. 29, pp. 23-33, June 1998.

[11] E. Hjelmas and B. Low, "Face detection: A survey," *J. Computer Vision and Image Understandin*g, vol. 83, pp. 235–274, Sep. 2001.

[12] M. Yang, D. Kriegman, and N. Ahuaja, "Detecting face in images: A survey," *IEEE Trans. PAMI*, vol. 24, pp. 34-58, Jun. 2002.

[13] Y. Liu, K. Schmidt, J. Cohn, and S. Mitra, "Facial asymmetry quantification for expression invariant human identification," *J. Computer Vision and Image Understanding*, vol. 91, pp. 138-159, Aug. 2003.

[14] Y. Tian, T. Kanade, and J. Cohn, "Recognizing action units for facial expression analysis," *IEEE Trans. PAMI*, vol. 23, pp. 97-115, Feb. 2001.

[15] A. Golovan, "Neurobionic algorithms of low-level image processing," in *Proc. Second All-Russia Scientific Conf. Neuroinformatics*, Moscow, Russia, May 2000, vol. 1, pp. 166-173.

[16] P. Ekman and W. Friesen, *Pictures of facial affect*. Palo Alto, Calif.: Consulting Psychologists Press, 1976.

[17] D. Shaposhnikov, A. Golovan, L. Podladchikova, N. Shevtsova, X. Gao, V. Gusakova, and I. Guizatdinova, "Application of the behavioral model of vision for invariant recognition of facial and traffic sign images," *J. Neurocomputers: Design and Application*, vol. 7, no. 8, pp. 21-33, 2002.

[18] G. Yang, and T. Huang, "Human face detection in a complex background," *J. Pattern Recognition*, vol. 27, pp. 53–63, Jan. 1994.

[19] K. Sobottka and I. Pitas, "Extraction of facial regions and features using color and shape information," in *Proc. Int. Conf. Pattern Recognition*, Vienna, Austria, Aug. 1996, vol. 3, pp. 421-425.

[20] H. Schneiderman and T. Kanade, "Probabilistic modeling of local appearance and spatial relationships for object recognition," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Santa Barbara, CA, June 1998, pp. 45-51.