Detection of Facial Landmarks from Neutral, Happy, and Disgust Facial Images

Ioulia Guizatdinova and Veikko Surakka
Research Group for Emotions, Sociality, and Computing
Tampere Unit for Computer-Human Interaction
Department of Computer Sciences
University of Tampere
FIN-33014 Tampere, Finland
ig74400@cs.uta.fi
Veikko.Surakka@uta.fi

ABSTRACT

Automated analysis of faces showing different expressions has been recently studied to improve the quality of human-computer interaction. In this framework, the expression-invariant face segmentation is a crucial step for any vision-based interaction scheme. A method for detecting facial landmarks from neutral and expressive facial images was proposed. In present study, a particular emphasis was given to handling expressions of happiness and disgust. The impact of these expressions on the developed method was tested using dataset including neutral, happiness and disgust images. The results demonstrated a high accuracy in detecting landmarks from neutral images. However, the expressions of happiness and disgust had a deteriorating effect on the landmark detection.

Keywords

Image processing, face segmentation, detection of local oriented edges, Gaussian, facial landmarks, human-computer interaction.

1. INTRODUCTION

In the past decades there has been a considerable interest in improving all aspects of human-computer interaction (HCI). One way to achieve intelligent HCI is making computers to interact with user in the same manner as it takes place in human-human interaction.

Humans naturally interact with each other through verbal (i.e. speech) and nonverbal (i.e. facial expressions, gesture, vocal tones, etc.) sign systems. It is argued that during human-human interaction only a small part of the conveyed messages is verbally communicated, and the greatest part is nonverbally coded. Considering nonverbal communication, it is possible to say that facial expressions occupy about a half of the transmitted signals. In the context of user-friendly HCI, a face is an important source of information about the user to be analyzed by the computer.

WSCG 2005 conference proceedings, ISBN 80-903100-7-9
WSCG'2005, January 31-February 4, 2005
Plzen, Czech Republic.

Automated analysis of a computer user's face has recently become an active research field in the computer vision community. Different vision-based schemes for intelligent HCI are currently being developed. The ability of a computer to detect, analyse and, finally, recognize a user's face has many applications in the domain of HCI.

The analysis and recognition of facial expressions in the context of HCI are elements of interaction design called affective computing [Jen98]. The main idea of the affective computing is that the computer detects the user's affective state and takes an appropriate action, for example, offers assistance for the user or adapts to the user's needs. Proper detection of the changes in the user's facial cues is a precondition for the computer to take any emotionally or otherwise intelligent socially interactive actions towards the user.

The Facial Action Coding System (FACS) [Ekm78] is widely used to analyse visually observable facial expressions. FACS has been developed for objective analysis of any changes in the facial appearances.

According to the FACS, a muscular activity producing changes in facial appearance is coded in the terms of action units (AU). Certain specific combinations of AUs have been frequently suggested to represent seven prototypical facial displays: neutral, happiness, sadness, fear, anger, surprise, and disgust.

It is known that reliable person identification and verification are important cornerstones for improving security in various contexts of information society. A natural means of identifying person that gives a close resemblance to the way how humans recognize persons is analysing a person's face.

Face identification has two important advantages. First, it requires a minimal interaction with a person, for example, compared with such biometrics as prompted speech or fingerprints. Second, it is impossible to lose or forget a face as it might happen with passwords or key-cards.

In this framework, automated detection of a face and its features is considered to be an essential requirement for any vision-based HCI scheme [Don99, Wis97]. However, due to such factors as illumination, head pose, expression and scale, facial features vary greatly in their appearance. It is shown that facial expressions are particularly important factors affecting the automated detection of facial features [Yac95]. Nowadays the problem of effective and expression-invariant face detection and segmentation still remains unsolved.

In our previous study we have proposed a method for detecting facial landmarks from neutral and expressive facial images [GuiS]. The developed approach has combined a feature-based method for face segmentation [Sha02] and a profound knowledge on how different facial muscle activations modify the appearance of a face during emotional and social reactions [Par04, Sur98].

Experimented findings have revealed that detection of landmarks from the lower part of a face was especially affected by expressions of happiness and disgust. In particular, detection of the nose and mouth produced the greatest number of detection errors. We assumed that these expressions modify the lower face so that it becomes difficult to differentiate lower face landmarks like nose and mouth. For this reason the present aim was to analyse an accuracy of landmark detection from images of happiness and disgust to corroborate the previous findings.

2. FACIAL LANDMARK DETECTION

The method for detection of facial landmarks consisted of three stages: image preprocessing, image map constructing and orientation matching [GuiS]. These stages are described below.

2.1. Image preprocessing

First, an image was transformed into the 256-grey-level-scale format. Then, a recursive Gaussian transformation was used to smooth the grey-level image [Gol00]. Image smoothing reduced a search space for detecting facial features (i.e. eliminated noise edges and removed small details) [Can86].

In the following stages of the landmark detection, the smoothed grey-level images were used to detect candidates for facial landmarks. The non-smoothed grey-level images allowed us to analyse the detected candidates in details. In that way, the amount of information to be processed was significantly reduced.

2.2. Image map constructing

The local high-contrast oriented edges were used as basic features for constructing edge maps of the image [Ryb98]. Apart from previous studies [Sha02], we decreased a number of edge orientations to construct edge maps of the image. In particular, we used $2 \div 6$ and $10 \div 14$ edge orientations (see Fig.1). Decreasing a number of edge orientations allowed us to reduce sufficiently the computational complexity of the method.

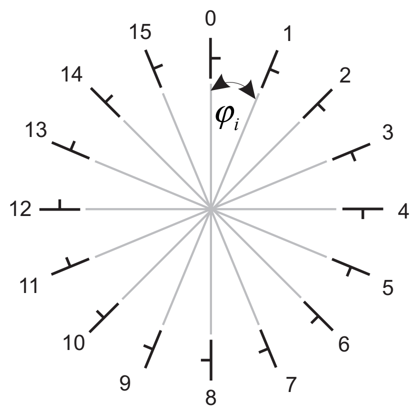


Figure 1. Orientation template, $\varphi_i = i \cdot 22.5°$, $i = 0 \div 15$.

The oriented edges were extracted by convolving the smoothed image with a set of ten convolution kernels. Each kernel was sensitive to one out of ten chosen edge orientations. For each pixel, the contrast magnitude of a local edge was estimated with maximum response of ten kernels at this pixel location. The orientation of a local edge was estimated with orientation of a kernel that gave the maximum response. The whole set of ten kernels resulted from differences between two oriented Gaussians with shifted kernels.

After the local oriented edges had been extracted, they were filtered by a contrast. The threshold for contrast filtering was determined as an average contrast of the whole smoothed image.

Then, the extracted oriented edges were grouped into edge regions presumed to contain facial landmarks. Edge grouping was based on neighbourhood distances between edges and was limited by a number of possible neighbours for each oriented edge. The optimal thresholds for edge grouping were determined using a small set of expressive images of the same person. The optimal thresholds represented landmark candidates as regions of connected edges that were well separated from the rest of edges.

Once the limits of edge regions had been detected, these regions were analysed more precisely. The procedures of edge extracting, contrast thresholding and edge grouping were applied to the non-smoothed image within the limits of the extracted edge regions. The threshold for contrast filtering was determined as a double average contrast of the non-smoothed image.

In the end, the primary image map consisted of edge regions representing candidates for facial landmarks. The centres of mass determined the locations of the landmark candidates. In the next stage, the landmark candidates were analysed according to their orientation description and matched with an orientation model.

2.3. Orientation matching

The orientation portraits of the landmark candidates were constructed on the basis of their local orientation description. The analysis of the orientation portraits revealed four important findings.

First, local oriented edges extracted within regions of eyebrows, eyes, nose and mouth had a characteristic density distribution. Thus, the orientation portraits of these landmarks had two dominant horizontal orientations. The results of the present study corroborated our previous findings [Sha02].

Second, we found that prototypical facial expressions did not affect the distribution of the oriented edges in the regions of facial landmarks [GuiS]. The orientation portraits of facial landmarks still had the same structure including two dominants corresponding to horizontal orientations (see Appendix 1*a*).

Moreover, for the regions of eyes and mouth the number of edges corresponding to horizontal orientations was more than 50% larger when compared to a number of edges corresponding to other orientations. All edge orientations were represented by non-zero number of the edges.

Third, the average orientation portraits of facial landmarks revealed the same structure including two horizontal dominants (see Fig.2, Appendix 2) [GuiS].

Fourth, noise regions extracted from the expressive images had an arbitrary distribution of the oriented

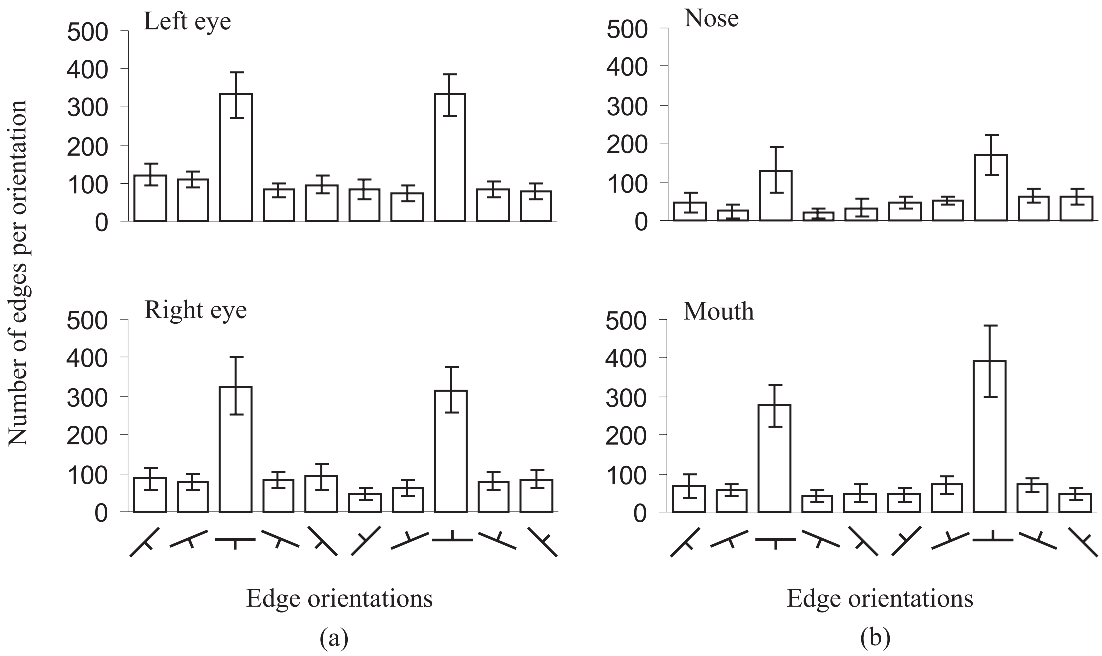


Figure 2. Orientation portraits of facial landmarks averaged over prototypical facial displays.

edges and often had orientations represented by zero number of edges (see Appendix 1*b*).

The knowledge on clear-cut distinction between orientation portraits of facial landmarks and noise regions allowed us to verify the existence of a landmark on the image. To do that, the orientation portraits of facial candidates were matched with an orientation model of facial landmarks.

2.3.1. Orientation model

The characteristic orientation model for detecting facial landmarks consisted of ten possible edge orientations, namely, edge orientations ranging from $45°$ to $135°$ and $225°$ to $315°$ in step of $22.5°$.

The following rules defined the structure of the orientation model: (a) horizontal orientations are represented by the biggest number of edge points; (b) a number of edges corresponding to each of the horizontal orientations is more than 50% bigger than a number of edges corresponding to other orientations taken separately; and (c) orientations can not be represented by zero-number of edge points.

The candidates that did not correspond to the orientation model were removed from the final image map. In such a way, the procedure of orientation matching filtered the regions containing landmarks from the noise.

The detected candidates for facial landmarks were further classified manually into one of the following groups: noise or facial landmark (i.e. eye-eyebrow, nose and mouth).

3. DATABASES

To evaluate the accuracy of the proposed method we used the Pictures of Facial Affect (PFA) database [Ekm76] and the Cohn-Kanade Face (CKF) database [Kan00].

The PFA database consisted of 110 frontal-view 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. On average, there were about sixteen pictures per expression. The size of the images was preset into 250 by 300 pixels.

The CKF images were originally coded using single AUs and their combinations. In according to translation rules defined in the Investigator's Guide to the FACS manual [Ekm00], the images were relabelled into the emotional prototypes. The images corresponding to the prototypes of happiness and disgust were selected. Thus, there were 172 images: 65 neutral images, 65 images of happiness and 42 images of disgust expression. All the images were normalized to contain only a facial part of the original image. Either of the datasets included faces with facial hair and glasses. All the images were resized into 250 by 480 pixel arrays.

The PFA database was used to select the optimal thresholds for edge grouping and to construct the landmark orientation model [GuiS]. In present study, the CKF database was used to test the accuracy of the method in detection of facial landmarks specifically from the images showing happiness and disgust.

(a) (b) (c) (d)

Figure 3. (a) original facial image; (b) extracted local oriented edges (black dots); (c) primal edge map represents candidates for facial landmarks (white regions) and their mass centers (crosses); (d) final edge map represents the detected facial landmarks.

4. RESULTS

Figure 3 gives an example of edge map composed of the local oriented edges extracted from the expressive facial images. Thus, local edges of $45° \div 135°$ and $225° \div 315°$ defined in step of $22.5°$ constituted the edge map of the happy image shown on Figure 3*b*. Figure 3*c* demonstrates the edge map after contrast thresholding and grouping extracted edge points into the candidates for facial landmarks. Figure 3*d* illustrates the final image map that included only the candidates having orientation portraits well matched with the orientation model.

The average number of the candidates per image of the primary edge map was 7.46. The results revealed that variations in facial expressions did not affect significantly the average number of the candidates per image. The average number of candidates per image was reduced to 3.71 for the final edge map. Such a fact allows us to claim that the procedure of orientation matching reduced the number of landmark candidates by 50%. Figure 4 illustrates the decrease in the number of candidates per image averaged over neutral, happy, and disgust images.

The accuracy of the proposed method was calculated as a ratio of the number of detected landmarks to the number of images used in testing. As it can be seen from Table 1, the developed method achieved a sufficiently high accuracy of 95% in detecting all four facial landmarks from the neutral images. As it can be seen from the table, both eyes are represented as a single column since these landmarks had equal detection accuracy.

However, the results showed that the expressions of happiness and disgust had a marked deteriorating effect on detecting facial landmarks. It is noteworthy that the detection of nose and mouth was more affected by facial expressions than the detection of eyes.

Three types of detection errors caused the decrease in detection accuracy. Figure 5 gives examples of such errors. The undetected facial landmarks were considered to be the errors of the first type. Such errors occurred when a facial landmark was rejected as a noise region after orientation matching. In particular, the nose was the most undetectable facial landmark (see Fig. 5*a*).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 into one region (see Fig. 5*b*). The errors of the third type were the misdetected landmarks that occurred when the noise regions were accepted as the facial landmarks (see Fig. 5*c*).

	Eye	**Nose**	**Mouth**	**Average**
Neutral	98	92	92	95
Happiness	100	50	50	75
Disgust	67	57	59	62
Neutral & Expressive	88	66	67	78

Table 1. Average accuracy (%) of the landmark detection

5. CONCLUSIONS

The method for detecting facial landmarks from both neutral and expressive facial images was presented and described. The method revealed an average accuracy of 95% in detecting four facial landmarks from neutral facial images.

However, the detection of facial landmarks from happy and disgust facial images produced a large number of detection errors. Thus, the expressions of happiness and disgust attenuated the average (i.e. over all regions) detection accuracy to 75% and of 62%, respectively. Especially the detection of nose and mouth were affected by both expressions of disgust and happiness. These expressions deteriorated the detection of nose and mouth to 50% for happiness. For the disgust expression the detection of nose and mouth deteriorated to 57 and 59, respectively. The present results corroborated our earlier findings that facial expressions have a marked deteriorating effect on the landmark detection

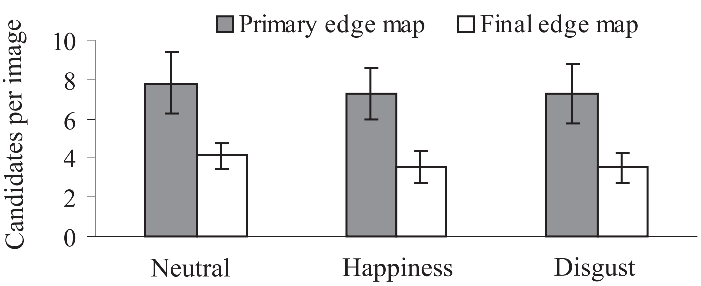


Figure 4. Average number of candidates per image before and after orientation matching.

(a) (b) (c)

Figure 5. Examples of the detection errors: (a) undetected nose; (b) incorrectly grouped nose and mouth; (c) detected noise region.

algorithms.

In summary, the accuracy of the landmark detection from neutral images was comparable with a detection accuracy of the known feature-based and colour-based methods though it is lower than neural network-based methods. The algorithms developed for landmark detection were simple and fast enough to be implemented as a part of systems for face and/or facial expression recognition.

The detection of facial landmarks from expressive images, especially from happy and disgust images needs to be improved. This is especially important in order to make a computer differentiate between positive expressions of emotions, for example, smiling and some negative expressions like disgust. To detect and differentiate between positive and negative user emotions, it is the very minimum prerequisite for affective HCI. This kind of an improvement of the method is also a precondition for recognizing facial identity of a user as well.

6. ACKNOWLEDGMENTS

This work was financially supported by the Finnish Academy (project number 177857), the Finnish Centre for International Mobility (CIMO), the University of Tampere and the University of Tampere Foundation.

7. REFERENCES

[Can86] Canny, J. A computational approach to edge detection. IEEE Trans. on Pattern Analysis and Machine Intelligent 8, No.6, pp.679–98, 1986.

[Don99] Donato, G., Bartlett, M., Hager, J., Ekman, P., and Sejnowski, T. Classifying facial actions. IEEE Trans. on Pattern Analysis and Machine Intelligent 21, No.10, pp.974–989, 1999.

[Ekm76] Ekman, P., and Friesen, W. Pictures of facial affect. Consulting Psychologists Press, Palo Alto, California, 1976.

[Ekm78] Ekman, P., and Friesen, W. V. Facial Action Coding System (FACS): A technique for the measurement of facial action. Consulting Psychologists Press, Palo Alto, California, 1978.

[Ekm00] Ekman, P., Friesen, W., and Hager, J. Facial Action Coding System (FACS). UTAH: A Human Face, Salt Lake City, 2002.

[Gol00] Golovan, A. Neurobionic algorithms of low-level image processing. in Second All-Russia Scientific Conference Neuroinformatics-2000 conf.proc., vol. 1, pp.166-173, 2000.

[GuiS] Guizatdinova, I., and Surakka, V. Detection of facial landmarks from emotionally expressive and neutral facial images. IEEE Trans. on Pattern Analysis and Machine Intelligent, submitted.

[Jen98] Jennifer, H., and Picard, J. Digital processing of affective signals. in IEEE ICASSP'98 conf.proc., Seattle, 1998.

[Kan00] Kanade, T., Cohn, J.F., and Tian, Y. Comprehensive database for facial expression analysis. in AFGR'00 conf.proc., Grenoble, p.46, 2000.

[Par04] Partala, T., and Surakka, V. The effects of affective interventions in human-computer interaction. Interacting with Computers, 16, pp.295-309, 2004.

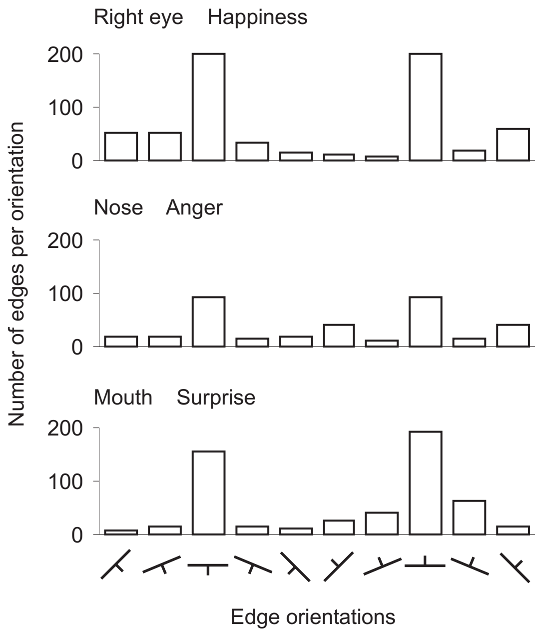
[Ryb98] Rybak, I. A model of attention-guided invariant visual recognition. Vision Research 38, No.15/16, pp.2387-2400, 1998.

[Sha02] Shaposhnikov, D., Golovan, A., Podladchikova, L., Shevtsova, N., Gao, X., Gusakova, V., and Gizatdinova, Y. Application of the behavioural model of vision for invariant recognition of facial and traffic sign images. Neurocomputers: Design and Application 7, No.8, pp.21-33, 2002.

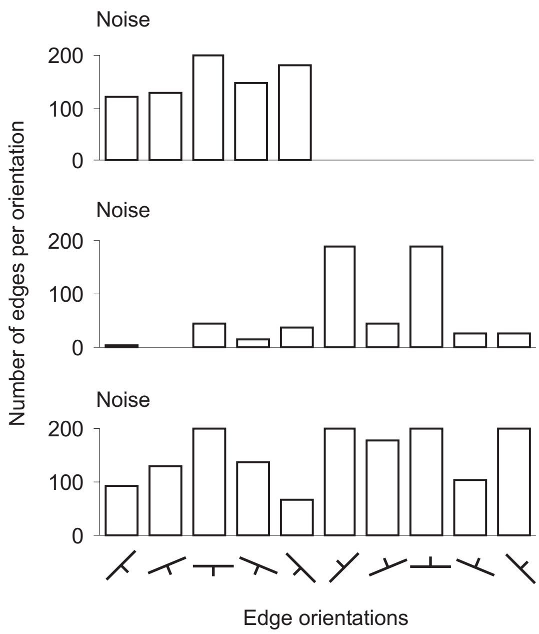
[Sur98] Surakka, V., and Hietanen, J. Facial and emotional reactions to Duchenne and non-Duchenne smiles. International Journal of Psychophysiology 29, pp.23-33, 1998.
[Wis97] Wiskott, L., Fellous, J-M., Kruger, N., and von der Malsburg, C. Face recognition by elastic bunch graph matching. IEEE Trans. on Pattern Analysis and Machine Intelligent 19, No.7, pp.775-779, 1997.
[Yac95] Yacoob, Y., Lam, H-M., and Davis, L. Recognizing faces showing expressions. in IWAFGR'95 conf.proc., Zurich, pp.278-283, 1995.

Appendix 1. Orientation portraits of (a) landmarks with characteristic edge distribution, and (b) noise regions with arbitrary edge distribution.

(a)



(b)



Appendix 2. Average orientation portraits for facial landmarks. The columns represent four facial landmarks and rows represent seven prototypical facial displays.