Robust landmark localization for facial therapy applications

Cornelia Lanz, Ilmenau University of Technology, Ilmenau, Germany

Joachim Denzler, Friedrich Schiller University, Jena, Germany

Horst-Michael Gross, Ilmenau University of Technology, Ilmenau, Germany

Abstract

In this paper, we present a robust method for the search of facial landmark positions in rigid and non-rigid regions of the face. The landmark localization is a necessary basis for the development of an automated training system for facial muscle dysfunction rehabilitation. We use 3D depth data captured with the Kinect camera and combine surface curvature analysis and anatomical knowledge constraints. Our approach has been successfully evaluated on a dataset with nine different facial expressions and generalizes well to unknown faces.

1 Introduction

A stroke or the Parkinson's disease may lead to facial muscle dysfunctions. The resulting problems are manifold: eating and swallowing difficulties, impaired face appearance, and possible long-term damages of the cornea because of weak eye closure. Regular facial training under supervision of a speech-language therapist is an important part of rehabilitation. Required by the high practicing frequency, patients need to conduct unsupervised facial exercises at home – accompanying to regular therapy. However, incorrect execution of exercises can impede the training success or even lead to further impairments [1]. Therefore, we aim at developing a therapy-accompanying technical assistance system for the home environment to overcome this problem. In our scenario, the patient sits in front of an intelligent assistive system equipped with a Kinect camera and performs facial exercises while the system gives feedback on the computer screen (Figure 1, left image). More details about the application scenario can be found in [2].

In cooperation with speech-language therapists, we selected nine exercises. Examples of these are presented in the right image of Figure 1. The execution of the therapeutic exercises strongly influences the surface of the face, e.g., with respect to convexity and concavity. These changes can be exploited in order to recognize and evaluate the performed exercises. For an automated system, this task includes several steps that are shown in the left image of Figure 1: the localization of the face position, the localization of landmarks or regions with distinctive surface changes, and the extraction of facial features from these regions.

In this paper, we focus on the localization of distinctive landmarks in 3D point clouds of the face. We are able to retrieve landmark positions – invariant to the facial expression –even in non-rigid, dynamic regions of the face by combining surface curvature analysis and anatomical knowledge constraints. In the following section, we give an overview of related work and our contribution. After that our method is introduced and validated by experiments.

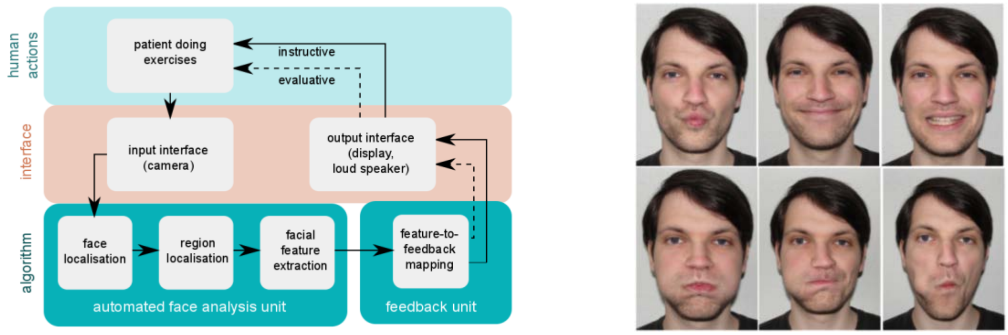

Figure 1 *Left image:* Overview of the assistive system currently developed. *Right image:* Examples for the therapeutic facial exercises selected in cooperation with therapists.

2 Related Work and Contribution

In literature, there are different approaches for landmark detection. Whereas some approaches are based on color data (e.g., [3], [4], [5]) others use depth information (e.g., [6], [7]). In our approach, we also employ depth information because of the exercises large influence on the face surface and the independence of lighting conditions. Many approaches for depth information make use of the characteristic surface properties in rigid areas of the face, like the nose tip, the eye corners, and the nose corners. The method presented in [6] is based on curvature analysis in combination with relief curves in order to obtain the position of these landmarks. In [7], a graph matching approach is employed for the localization of the inner eye corners and the nose tip. The approach from [8] combines 3D range and 2D intensity data for the detection of eye, nose, and mouth landmarks. However, they evaluate their method on a database of images with a limited number of facial expressions (only neutral and smile).

The contribution of this paper is that we complement curvature analysis, which is a common method in landmark localization, with anatomical knowledge. This enables the localization of landmarks in regions with less distinctive and constant surface curvature, for example on the mandible of the face. Furthermore, we obtain a consistent feedback system, because surface curvature can also be successfully employed for the recognition of therapeutic exercises as it was shown in [9].

3 Method and Experiments

The selection of the positions that are localized in this paper was based on the landmarks and regions that were used in [9]. The left image of Figure 2 shows their distribution in the face. In order to simplify the basic idea of our approach, we focus on one facial half. Of course, the approach can be transferred to the search for the remaining landmarks in the other half.

In the following subsections, we introduce our method and necessary preprocessing steps. Furthermore, detailed information about the employed image database and the results of evaluation are given.

3.1 Preprocessing and image database

The image database for our experiments was captured with the Kinect from Microsoft. Affordable hardware is an important aspect in order to facilitate the integration of the system in the therapy process. Additionally, the Kinect provides 2D color and 2.5D depth images. As a result, more information about the scenery can be obtained, and the advantages of both data types can be exploited. We transformed the 2.5D depth images to 3D point clouds in order to enable a simple calculation of intra-facial distances ([10]), which is the fundament for the anatomical knowledge constraints. Each facial point in a 3D point cloud is represented by (x, y, z)-coordinates in a metric system.

Prior to the retrieval of the landmark positions, we use the framework from [11] on the 2D images for robust face detection. Then – using the 3D information – mean and Gaussian curvatures are computed for all pixels in the bounding box [12]. This curvature computation is done once at the beginning of the procedure, and the results can be exploited in all following steps.

For the evaluation of the localization results, we employed a set of 696 images, which contain eleven different subjects. The subjects perform nine exercises that were selected in cooperation with speech-language therapists.In order to generate a reference for the evaluation of the automatically retrieved landmark positions, in all images landmark positions were labeled manually as well.

3.2 Steps of the Landmark Localization

We use a cumulative and iterative method for the localization of landmarks. The word cumulative implies that the approach starts with the search of landmark positions in rigid areas of the face and then transfers the knowledge about their position to improve the results for landmarks in less distinctive and more dynamic facial areas. The process, which is visualized in Figure 3, starts with the localization of the nose tip. It has a distinctive elliptical convex curvature and, therefore, can robustly be detected by using thresholds for the mean and the Gaussian curvature ([9], [12]). However, localization by thresholds is restricted to the nose tip because the other landmarks lie in areas with higher surface variation. Furthermore, their curvature is less consistent between different subjects.

The remaining landmarks are now localized subsequently according to the numeration in Figure 2. In each stage, the search space is constrained by using the minimum and maximum distance to the previously localized landmarks (Figure 2, center image). This constraint reduces false positive candidates. The distance values were determined on a small subset of 99 randomly selected images. Their value varies depending on the subject, the selected landmarks, and the performed exercises. In the next steps various features are extracted from the candidate pixels. These include: the Gaussian and the mean curvature values of the 3x3 neighborhood, the absolute value of the distance to the nose tip, and surface path profiles between the candidate pixels and the nose tip.

The surface path profiles contain the depth differences between ten equally spaced points on the shortest connection of the two landmarks. In contrast to the previously mentioned distances between the landmarks, which measure the Euclidean distances, the profiles run on the surface of the face. Now, for each candidate pixel, a support Vector Machine (SVM) with a Radial Basis Function is applied ([13]). All images of the person that is present in the test image are excluded from the training set. The result of the classification is a set of few candidate pixels.

In the post-processing step, we calculate the mean position of these landmarks in order to obtain an estimation for the landmark position. Now the process is iterated until all seven landmarks are detected. For landmarks number 6 and 7, feature extraction and SVM classification is omitted because curvature values in these regions are not distinctive enough and similar to a large neighborhood. However, there is enough distance information because of the previously detected five or six landmarks in order to constrain the search space to few pixels and calculate the mean position using these pixels. For landmark 6, we additionally compute an edge image using a Sobel operator in order to find areas with larger depth difference. These are located on the transition from the chin to the throat.

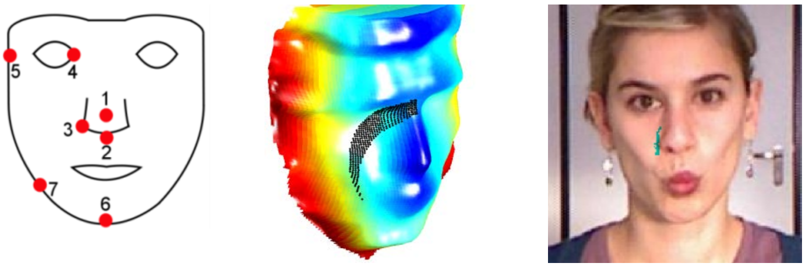

Figure 2*Left image*: The facial landmarks numbered according to the order of localization. *Center image*: Constrained region for the inner eye corner (landmark number 4). *Right image*: Candidate pixels for landmark number 3 after the SVM classification.

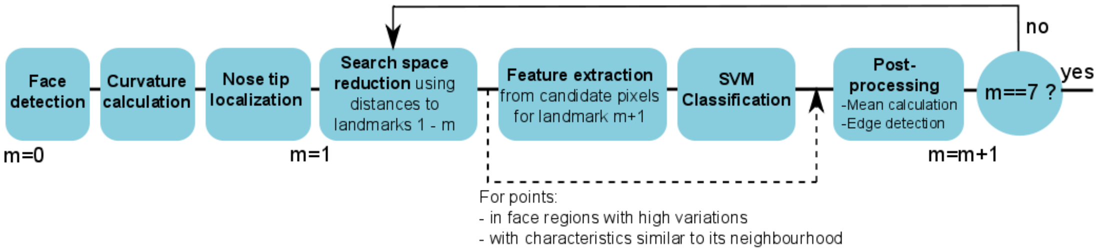
Face detection
Curvature calculation
Nose tip localization
Search space reduction using distances to landmarks 1 - m
Feature extraction from candidate pixels for landmark m+1
SVM Classification
Post-processing
-Mean calculation
-Edge detection
m==7 ?
no
yes
m=0
m=1
m=m+1
For points:
- in face regions with high variations
- with characteristics similar to its neighbourhood

Figure 3 Procedure of the landmark localization. The variable *m* represents the number of detected landmarks.

3.3 Experiments and Discussion

In this section, we evaluate our method by comparison of the manually labeled positions with the automatically localized positions. The mean and standard deviations are plotted in the right image of Figure 3. A deviation of one pixel corresponds to about 1.5 millimeters. It can be seen that the deviation for the landmarks 6 and 7 in the non-rigid regions of the face is higher than in rigid ones. The high deviation in the chin region for the y-axis is also caused by the manual label positions, which were not marked directly on the edge between chin and throat.

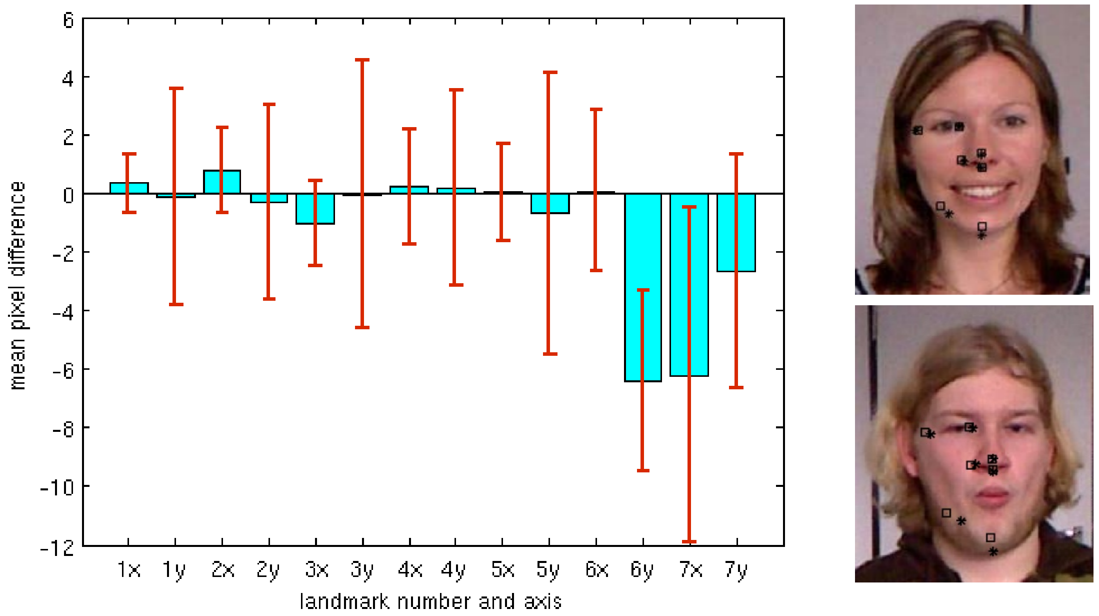

Figure 4 *Left image*: Mean and standard deviations for the comparison of manually and automatically localized landmarks. *Right image*: Examples for manual (square) and automated (star) localizations.

4 Conclusion and future work

In this paper, we have presented the robust localization of facial landmarks using surface curvature analysis and anatomical knowledge constraints. Future work will focus on the integration of angular relationships between the landmarks in order to complement the anatomical knowledge constraints. Furthermore, color information can be added in order to improve the results.

Literature

[1] Wolowski, A. (2005). Fehlregenerationen des Nervus facialis – ein vernachlässigtes Krankheitsbild. *Dissertation*. Universität Münster.

[2] Lanz, C., Denzler, J. and Gross, H.-M. (2013). Facial movement dysfunctions: Conceptual design of a therapy-accompanying training system. *Proceedings of the 6th German Ambient Assisted Living Congress*, pages 186-195.

[3] Dong, J., Ma, L., Li, Q., Wang, S., Liu, L., Lin, Y. and Jian, M. (2008). An approach for quantitative evaluation of the degree of facial paralysis based on salient point detection. *International Symposium on Intelligent Information Technology Application Workshops*, pages 483-486.

[4] Limbeck, P., Kropatsch, W.G. and Haxhimusa, Y. (2012). Semi-automatic tracking of markers in facial palsy. *International conference on Pattern Recognition*, pages 69-72.

[5] Zobel, M., Gebhard, A., Paulus, D., Denzler, J. and Niemann, H. (2000). Robust facial feature localization by coupled features. Proc. of the *Int. Conf. on Automatic Face and Gesture Recognition*, pages 2-7.

[6] Segundo, M.P., Silva, L., Bellon, O.R.P. and Queirolo, C.C. (2010). Automatic face segmentation and facial landmark detection in range images. *Systems, Man, and Cybernetics*, pages 1319-1330.

[7] Romero-Huertas, M. and Pears, N. (2008). 3d facial landmark localisation by matching simple descriptors. *Proc. of the Int. Conf. on Biometrics: Theory, Applications and Systems*, pages 1-6.

[8] Lu, X. and Jain, A.K. (2005). Multimodal facial feature extraction for automatic 3d face recognition. Department of Computer Science, Michigan State University, Tech. Rep..

[9] Lanz, C., Olgay, B., Denzler, J. and Gross, H.-M. (to appear). Automated classification of therapeutic face exercises using the kinect. *Proc. of the Int. Conf. on Computer Vision Theory and Applications*.

[10] Hartley, R. and Zisserman, A. (2000). *Multiple view geometry in computer vision*. Cambridge university press.

[11] Viola, P. and Jones, M. (2004). Robust real-time face detection. *Int. Journal of Computer Vision*, 57(2):137-154.

[12] Besl, P. and Jain, R. (1986). Invariant surface characteristics for 3d object recognition in range images. *Computer Vision, Graphics, and Image Processing*, 33(1);33-80.

[13] Chang, C.-C. and Lin, C.-J. (2011). LIBSVM: A library for support vector machines. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

Contact: Cornelia Lanz, cornelia.lanz@tu-ilmenau.de, +49 (0)3677 69-4172