A Feedback Estimation Approach for Therapeutic Facial Training

Cornelia Dittmar[1], Joachim Denzler[2] and Horst-Michael Gross[1]
[1] Neuroinformatics and Cognitive Robotics Lab, Ilmenau University of Technology, Germany
[2] Computer Vision Group, Friedrich Schiller University Jena, Germany

***Abstract*— Neuromuscular retraining is an important part of facial paralysis rehabilitation. To date, few publications have addressed the development of automated systems that support facial training. Current approaches require external devices attached to the patient's face, lack quantitative feedback, and are constrained to one or two facial training exercises. We propose an automated camera-based training system that provides global and local feedback for 12 different facial training exercises. Based on extracted 3D facial features, the patient's performance is evaluated and quantitative feedback is derived. The description of the feedback estimation is supplemented by a detailed experimental evaluation of the 3D feature extraction.**

I. INTRODUCTION

Neuromuscular retraining is an important part of rehabilitation for patients with hemifacial paralysis. To date, biofeedback is provided via electromyography (EMG) or mirror (see Fig. 1).

We developed an automated camera-based training system that provides feedback for training exercises, with the goal of improving the process of facial paralysis rehabilitation. In the context of this paper, feedback refers to displaying an automatic estimate of the similarity between the patient's facial movement and target movements conducted by healthy persons. Fig. 2 presents an overview of our approach. In our scenario, RGB and 2.5D images are captured while the patient is performing facial training exercises specified by the training system. Input images are pre-processed to prepare them for feature extraction. The extracted 3D facial features are the basis for the feedback estimation, which relies on Random Forest classification and proximities [1]. The key contributions of this paper with respect to the described procedure are:

- Review and selection of 3D facial feature extraction methods. Detailed evaluation of features from emotion/face recognition systems for our therapeutic scenario.
- Introduction of a new algorithm for feedback estimation, which is based on the extracted 3D features. In comparison to state-of-the-art approaches, our feedback estimation: 1) provides *global and local feedback*, which allows well-directed adaption of facial movements by the patients, 2) does not require *external devices* attached to the patients face, 3) is based on generic feature extraction algorithms, which are not adapted to specific facial movements. This allows easy integration of further exercises. Thus, we employ *twelve facial exercises* compared to one or two exercises in similar state-of-the-art approaches. 4) is evaluated using a data set of *real patients*.

Fig. 1: Present and target state of facial paralysis therapy (*bottom image*: © TU ILMENAU / Michael Reichel).

The remainder of this paper is structured as follows: In Sec. II, related work in the areas of automated facial paralysis rehabilitation and 3D facial feature extraction is presented. Our method is introduced in Sec. III. Subsequently, a detailed experimental evaluation of the extracted 3D facial features is given, and examples of the estimated feedback are presented. Conclusions are drawn in Sec. V.

II. RELATED WORK

In this section we give an overview of related work with respect to the application scenario (Sec. II-A) as well as more general works addressing 3D facial feature extraction (Sec. II-B).

A. Facial Paralysis Rehabilitation

Recent works in image processing and machine learning aim to improve facial paralysis rehabilitation and its key elements, namely *diagnosis* and *therapy*.

Most of the work is focused on the development of methods for automated facial paralysis grading (i.a. [2], [3], [4], [5], [6], [7], [8]). The underlying objective is to enhance reliability and objectivity of processes and outcomes related to diagnosis. Compared to that, only few publications focus on the therapeutic part of the rehabilitation process due to higher demands for real-time capability. Diagnoses can be done offline using recorded images. An interactive therapy system requires attendance of the patient.

978-1-5090-4023-0/17 $31.00 © 2017 IEEE
DOI 10.1109/FG.2017.26

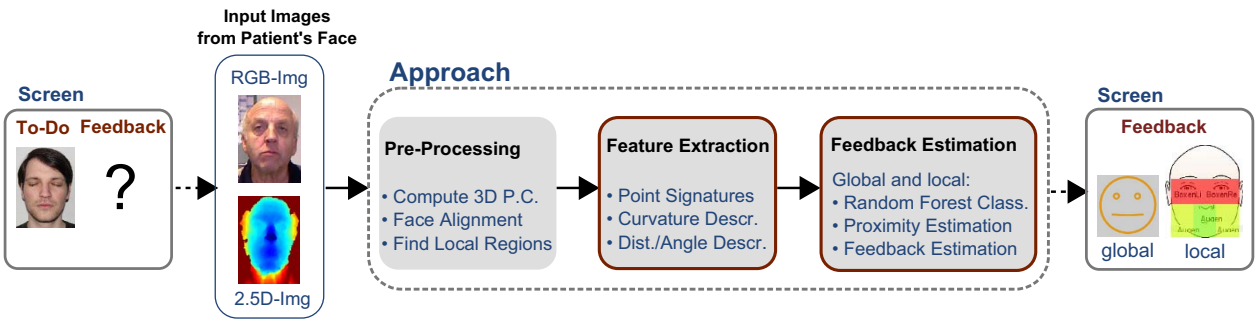


Fig. 2: Schematic overview of our approach. The red frames mark the key contributions of this publication.

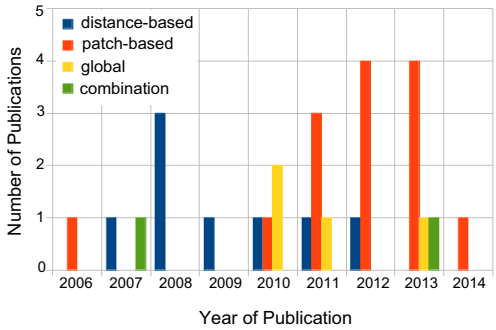


Fig. 3: Overview of the reviewed feature extraction approaches.

However, the improvement of automated landmark localization methods, also known as face alignment, directed more attention towards this subject.

In [9], a wearable robot mask is presented to assist physiotherapy of a hemifacial paralyzed patient. Physical support of the paralyzed hemiface is provided by pulling the facial skin through flexible wires attached to the face. The actuators are controlled by bioelectrical EMG-signals obtained from the healthy side of the face.

In [10], an interactive game for the training of the eyelid movement is proposed. The patient's face is captured by a camera, and a virtual seesaw is displayed on a screen. The degree of eyelid closure, extracted as the vertical length between the upper and the lower eyelid, is used to control the slope of the seesaw.

In [11], a Kinect-based interactive game for the rehabilitation of tongue and lip movements is presented. The patient has to collect virtual food displayed on a screen by performing licking and biting movements. Similar to [10], a knowledge-based feature extraction method is used. As a result, the integration of further facial exercises would require the implementation of additional feature extraction methods.

B. 3D Facial Feature Extraction

Automated face analysis is a popular research topic with various applications such as facial expression recognition, Action Unit detection and face recognition. As a result, there are many different feature extraction methods comprising a variety of properties ([12], [13]). We restricted our research to *3D feature-based static* extraction algorithms due to the following reasons:

- *2D vs. 3D:* Performing facial training exercises results in a continuous change of the facial surface. Especially changes in homogeneous regions, e.g., in the areas of the cheeks, are better represented by 3D than 2D intensity data as preliminary experiments showed. For the sake of brevity, in this paper we focus on 3D features only. However, in future work the combined strengths of both color-based and depth-based features should be exploited by means of a late or early fusion strategy.
- *Static vs. Dynamic:* Each of the facial training exercises that we selected in cooperation with speech therapists has to be retained for two to three seconds. Compared to maintaining each exercise, the pace of the movement is less important. Therefore, we extract static features from single images. Nevertheless, dynamic extraction from video data is a possible future extension.
- *Feature-based vs. Model-based:* An interactive therapy scenario requires real-time capability. However, the fitting of a 3D model can be computationally intensive compared to feature-based extraction algorithms.

We reviewed 28 papers, published between 2006 and 2014, that matched the listed properties. A majority of these publications belongs to the field of facial expression recognition (22 out of 28). Feature-based approaches can be divided into distance-based, patch-based and global approaches, thus we classified the 28 approaches according to these categories (see Fig. 3). Early approaches are dominated by distance-based feature extraction. Since 2010 the number of patch-based approaches is growing. For the sake of brevity, the following overview will be constrained to a subset of the 28 publications.

1) Extraction of Euclidean distances: The extraction of distance features is based on the underlying assumption that facial movements cause distinct changes in spatial relations between landmarks. In total, eight publications employ distance-based feature extraction ([14], [15], [16], [17], [18], [19], [20], [21]). Additionally, in some of the listed approaches angles between pairs of

distance vectors are extracted.

2) Patch-based feature extraction: Facial landmarks serve as reference points for the extraction of distances. However, some facial areas, e.g., the cheeks, have a lower spatial density of landmarks and therefore require alternative feature descriptors. Patch-based approaches extract features from the facial surface. Six of the reviewed patch-based approaches rely on curvature analysis ([22], [23], [24], [25], [26], [27]), five on 3D Local Binary Patterns ([28], [29], [30], [31], [27]).

III. APPROACH

Fig. 2 illustrates our system that consists of three main parts. Each part will be described in this section.

A. Pre-Processing

During each training session, the patient is seated in front of a screen while performing facial training exercises provided by the system (see Fig. 4). RGB and 2.5D facial images are continuously captured using a Kinect sensor. In each RGB image, 49 facial landmarks, shown in Fig. 5a, are detected using discriminative deformable models [32][1]. In the next step, a colored 3D point cloud is generated from the RGB and the 2.5D images, based on the intrinsic and extrinsic camera parameters previously estimated via calibration [33]. Finally, twelve local regions for patch-based feature extraction are determined by employing the position of the facial landmarks and a depth-based foreground-background segmentation. The positions of the local regions are visualized in Fig. 5b.

B. Feature Extraction

This subsection describes the 3D facial feature extraction methods of our approach. With regard to the findings of the literature review, we selected one exemplary distance-based approach and implemented patch-based curvature extraction. We compare both to extracted point signatures. The three feature vectors, resulting from the different extraction algorithms, are later concatenated and employed for feedback estimation.

1) Extraction of Euclidean distances and angles: We rely on the work of Rabiu et al. [18], because it is the most recent distance extraction approach from the review in Sec. II. It comprises the extraction of 16 Euclidean distances δ_i between distinct facial landmarks and 27 angles θ_j between pairs of these 3-dimensional distance vectors. The resulting feature vector consists of 43 dimensions.

2) Extraction of surface curvature: We estimate mean and Gaussian curvature values for small facial surface patches and use HK-classification to classify them into eight discrete surface types visualized in Fig. 5d-5e [34]. We refer to [35] and [36] for a detailed description of our method. In the next step, the estimated surface types for each local region shown in Fig. 5b are combined in an 8-bin histogram vector. The vectors of all twelve local regions are concatenated to a final 96-dimensional feature vector.

[1]We use a pre-trained model provided by Asthana and Zafeiriou under `https://sites.google.com/site/chehrahome/`

3) Extraction of point signatures: Point signatures describe paths on a surface that run radially around a distinctive point (see Fig. 5f). We implemented a modified version of the algorithm originally introduced for object and face recognition by Chua et al. [37]. For a detailed description of our implementation, we refer to [35] and [36]. We selected the nose tip as center point and determined eight different radii r for our point signatures, with $r \in \{4\text{cm}, 4.5\text{cm}, ..., 7.5\text{cm}\}$ (see Fig. 6e). The angle for point signature sampling is $\alpha = 5.7°$. This results in 64-dimensional vectors for each of the eight point signatures, which are then concatenated to a vector of 512 dimensions. In the later evaluation, we additionally test a 32-dimensional point signature vector ($\alpha = 11.6°$).

C. Feedback Estimation

Our feedback estimation approach is based on Random Forests and proximities derived from these forests [1]. Proximites are an estimate for the similarity between two observations. Our feedback estimation consists of two main parts, namely preliminary offline preparations and online feedback estimation.

1) Offline steps: First, we train a Random Forest (RF) based on all training observations of the $K = 12$ facial training exercise classes. Training observations are represented by the concatenated feature vector of all three feature types. Each training observation, attached to class k, with $k \in \{1, ..., K\}$, should be a valid representative of a correct performance and therefore only comprise data of healthy persons. The next step focuses on the N_k training observations of the k-th exercise and is repeated for all K facial training exercise classes: For each training observation i, with $i \in [1, 2, ..., N_k]$, proximities to all j-th training observations, with $j \in [1, 2, ..., N_k]$ and $i \neq j$, are determined and saved in a $(N_k - 1)$ dimensional vector $\mathbf{v}_{k,i}$.

2) Online steps: The system provides a facial training class k, that the patient should ideally perform. Static images of the patient are captured in fixed time intervalls, features are extracted, and subjected to the trained RF classifier as test observation t_k. Now, the following feedback estimation steps are conducted for each observation. First, the proximities $a_{t_k,i}$ between the test observation t_k and the N_k training observations i are determined. Second, each $a_{t_k,i}$ is compared to the elements of the corresponding training vector $\mathbf{v}_{k,i}$. In the course of this comparison, we determine the percentage $p_{t_k,i}$ of elements of $\mathbf{v}_{k,i}$ that are smaller than $a_{t_k,i}$. The values are concatenated in a vector:

$$\mathbf{p}_{t_k} = \begin{bmatrix} p_{t_k,1} & \cdots & p_{t_k,i} & \cdots & p_{t_k,N_k} \end{bmatrix}. \quad (1)$$

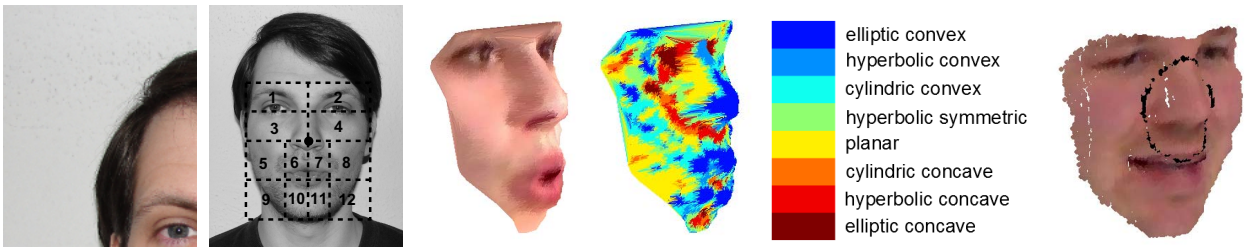
The estimated continuous feedback value $\tilde{f}_{t_k}$, with $\tilde{f}_{t_k} \in [0, 1]$, results from the median of all elements of vector $\mathbf{p}_{t_k}$:

$$\tilde{f}_{t_k} = \text{median}(\mathbf{p}_{t_k}). \quad (2)$$

We additionally compute the median $\tilde{a}_{t_k}$ of all N_k values of $a_{t_k,i}$. Based on $\tilde{f}_{t_k}$ and $\tilde{a}_{t_k}$, six discrete feedback levels are obtained according to Tab. I.

(a) (b) (c) (d) (e) (f) (g) (h) (i) (j)

Fig. 4: Facial training exercises that we selected in cooperation with speech therapists. The exercises (b) and (c) are also performed in a vertically mirrored manner, which results in 12 different exercise classes. **(a)** Cheek **(b)** CheekL (CheekR) **(c)** TongueL (TongueR) **(d)** Taut lips **(e)** Eyes closed **(f)** A-shape **(g)** I-shape **(h)** O-shape **(i)** U-shape **(j)** Pursed lips.



(a) (b) (c) (d) (e) (f)

Fig. 5: **(a)** Detected facial landmarks. **(b)** Local regions for patch-based feature extraction. **(c)-(e)** Examples of the eight estimated discrete surface types. **(f)** Radial point signature centered around the nose tip.

TABLE I: Discretization of the feedback based on the fulfillment of three different constraints ($B1 \wedge B2 \wedge B3$).

	Level 1	**Level 2**	**Level 3**	**Level 4**	**Level 5**	**Level 6**
B1	$\tilde{a}_{t_k} = 0$	$\tilde{a}_{t_k} > 0$	$\tilde{a}_{t_k} > 0$	$\tilde{a}_{t_k} > 0$	$\tilde{a}_{t_k} > 0$	$\tilde{a}_{t_k} > 0$
B2		$\tilde{f}_{t_k} = 0$	$\tilde{f}_{t_k} > 0$	$\tilde{f}_{t_k} > 0,05$	$\tilde{f}_{t_k} > 0,3$	$\tilde{f}_{t_k} > 0,5$
B3			$\tilde{f}_{t_k} \leq 0,05$	$\tilde{f}_{t_k} \leq 0,3$	$\tilde{f}_{t_k} \leq 0,5$	

IV. EXPERIMENTS AND RESULTS

This section presents the results of our work. First, we introduce the experimental data sets, followed by a detailed evaluation of the feature extraction methods. In the end, real-world feedback results are shown and discussed.

A. Experimental Data

For the evaluation, two different data sets are employed. Both were pre-processed according to the description in Sec. III-A.

The evaluation of the facial feature extraction is based on a collection of 931 colored 3D point clouds. The point clouds comprise facial data of 11 healthy persons, aged between 25 and 28 years, who perform 12 facial training exercises shown in Fig. 4. In order to exclude bias due to the automated landmark localization, feature extraction (and determination of local regions) is based on 58 manually labeled landmarks.

The second data set contains 117 colored 3D point clouds of facial paralysis patients during their rehabilitation session. It is solely used as test set and is evaluated using the RF that was trained with the first data set. Since the second data set was recorded in an other environment and contains additional persons (of different ages), it allows us to assess the generalization power of our model. Furthermore it represents a real therapy situation, e.g., a slight in-plane head rotation due to physical impairments of the patients. Experimental results, which were determined using the second data set, are given in Sec. IV-C.

B. Evaluation of Feature Extraction

In this section, we evaluate how well the 12 facial training exercise classes can be separated based on the extracted 3D facial features. For classification, we employ a Random Forest of 150 decision trees, which is also part of our feedback estimation algorithm. Results of the RF are compared to the results of a multi-class linear SVM. We use the RF-implementation provided by the TreeBagger-class of Matlab, and the SVM-implementation provided by the LIBSVM library [38]. The evaluation is based on a n-fold crossvalidation, where $n = 11$ corresponds to the number of subjects in the data set. This satisfies the constraints of a real-world scenario in which the images of the patient will not be part of the training data. The measure of our evaluation

is the mean accuracy, which is the arithmetic mean of all accuracies over the 11 test persons (subsequently middled over the 12 classes). Prior to training and classification, the vectors of the three feature types are concatenated to a final, 651-dimensional feature vector. RF-based classification of the concatenated feature vectors results in a mean accuray of 80.41%. When multi-class linear SVMs are used, 84.73% of all test observations are classified correctly (see Fig. 6a). The findings for the single feature types, the local classification and the automated landmark localization are as follows:

1) Evaluation of Euclidean distances and angles: Feature extraction according to [18] results in mean accuracies of 61.06% (RF) and 60.62% (lin. SVM), when used for the classification of the 12 facial training exercises[2]. We additionally evaluated distance and angle features separately using RF classification (distance: 49.78%, angle: 62.95%). Interestingly, classification solely based on extracted angles results in a higher mean accuracy than the classification based on the concatenated distance-angle vector. In the next step, we estimated the mutual information (MI) between each feature dimension and the target class variable [39]. Fig. 6c visualizes the features that resulted in a high MI value, Fig. 6d the features with low MI values. The horizontal stretching of the mouth δ_{14}, for example, shows a high relation to the target class. The mean vertical width of both eyes δ_8, however, exhibits a weak relation the target class.

2) Evaluation of point signatures: Point signature extraction according to Sec. III-B.3 results in mean accuracies of 65.94% (RF) and 75.4% (lin. SVM) for $\alpha = 5.7°$ ($\alpha = 11.6°$: 63.98% (RF), 73.89% (lin.SVM)). Compared to the outcomes of the classification based on distance-angle features, an improvement of 4.88 (RF) and 14.78 (lin. SVM) percentage points is yielded (for $\alpha = 5.7°$). However, the number of feature dimensions increases from 43 to 512. Once again, we estimated the MI between the single feature variables and the target class variable. The results are shown in Fig. 6e-6f. Features extracted in the lower cheek areas exhibit relatively large relation to the target variable, compared to features extracted in the areas of the eyes.

3) Evaluation of surface curvature: The classification based on extracted curvature features results in mean accuracies of 69.71% (RF) and 71.81% (lin. SVM). Compared to point signature based classification, mean accuracies differ in $+3.77$ (RF) and -3.59 (lin. SVM) percentage points. However, the number of feature dimensions is smaller (96 vs. 512). Again, we estimated the MI between the target variable and the 96 feature variables. For better visualization, the eight MI values of each local region are averaged (see Fig. 5b for the positions of the 12 local regions in the facial area). The barplot in Fig. 6b shows the estimated MI values. Similar to the results obtained for the point signatures, feature variables that were extracted in the cheek regions exhibit a larger relation to the target class variable than features extracted in the areas of the eyes.

4) Evaluation of local feature extraction: Besides global feedback, we additionally want to provide local feedback for semantic areas of the face, namely eye, mouth, and cheek areas. For local feedback estimation, feature extraction is constrained to the local facial area. Point signature curves, for example, which consist of 64 sample values, are divided in five segments according to Fig. 6g. Extracted sample values of each segment are assigned to the corresponding region. The assignment of the 12 curvature vectors to five feedback regions is done according to Fig. 6h. Distance and angle features are attributed to one or more regions according to their position in the face. Compared to global classification (RF: 80.41%), local classification results in decreased recognition accuracies. Based on the concatenated features that were extracted from the regions of the eyes, only 42.31 (R1) and 38.17 (R2) percent of the 12 facial training exercises are correctly classified. Better results are achieved for the areas of the mouth (R4: 59.42%) and the cheeks (R3: 69.19%, R5: 67.10%).

5) Evaluation of automated landmark localization: When automated landmark detection according to [32] is used, a mean accuracy of 79.54% (RF) is obtained (manual: 80.41%).

C. Evaluation of the Feedback Approach

In this section, we provide outcomes of our feedback estimation based on automatically detected landmarks. Typical results for in therapy patients are presented in Fig. 7. Target performances by a healthy person were shown in Fig. 4. Different states of the target exercise *eyes closed*, for example, are illustrated by Fig. 7b and 7c. The insufficient closure of the eyelids in the first image is evaluated correctly. Fig. 7d shows a weaker performance of the exercise *cheek* compared to Fig. 7e, which also becomes apparent in the estimated feedback. The remaining subfigures show reasonable results for the estimated global feedback as well. However, the robustness of the local feedback needs further improvement.

We conclude the evaluation with these typical examples instead of a quantitative evaluation. The reason is, that currently, no ground truth feedback level annotations are available for our data set due to the enormous effort that would arise for speech therapists in order to label our data. However, the outcomes of our data driven feedback approach constitute an initial configuration of the final system. We plan to refine our results later on by collecting real-time evaluations of speech therapists. For this purpose, we want to integrate a feedback interface for speech therapists that allows optional evaluation of the estimated feedback levels. The feedback is then used to refine our system. We believe that it might pose less of a challenge for the therapist to correct decisions of the system rather than to start annotations from scratch.

V. CONCLUSIONS AND FUTURE WORKS

We introduced an automated training system for patients with hemifacial paralysis. The objective is to support practice

[2]For comparison: In [18] a mean accuracy of 92.2% was obtained for 7 basic emotion classes.

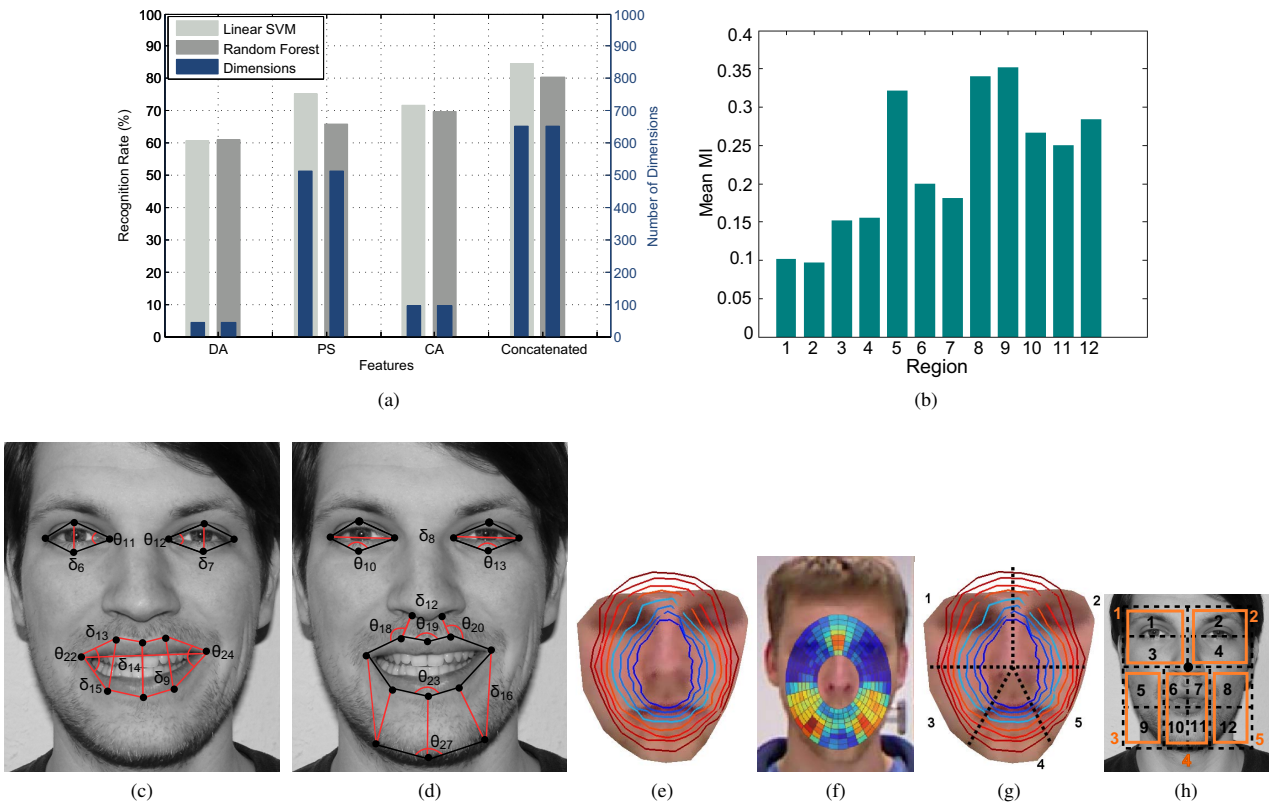


Fig. 6: **(a)** Mean accuracies and number of feature dimensions for the classification of 12 facial training exercises (extraction of DA: distance-angle, PS: point signature, CA: curvature). **(b)** Mean mutual information estimated for the 12 local extraction regions of the surface curvature extraction. **(c)-(d)** Distances and angles with (c) high MI and (d) low MI to target class variable. **(e)** Eight point signatures with different radii. Each point signature is sampled and results in a 32- or 64-dimensional feature vector (based on the radial sampling angle). **(f)** MI for the single feature variables of eight different 32-dimensional point signatures (blue: low MI, red: high MI). **(g)** Local segments of point signatures. **(h)** Five local feedback regions.

by giving global and local feedback with respect to the facial movements. We conducted a detailed evaluation of the feature extraction algorithms and presented typical results of our feedback estimation. In contrast to other state-of-the-art approaches, no wearable devices are necessary. Additionally, twelve instead of one ore two exercises can be evaluted.

Future work includes the improvement of the local feature extraction, i.a. by extracting local binary patterns. Additionally, we want to integrate a feedback interface that enables speech therapists to evaluate and reinforce our initial data driven feedback estimation results.

VI. ACKNOWLEDGMENTS

We would like to thank the patients and the staff of the “m&i-Fachklinik” Bad Liebenstein and the Logopaedische Praxis Irina Stangenberger, who supported our work by giving valuable insights into facial paralysis rehabilitation.

REFERENCES

[1] A. Cutler, D. R. Cutler, and J. R. Stevens. “Random forests”. In: *Ensemble Machine Learning*. Springer, 2012, pp. 157–175.

[2] A. Gebhard et al. “A system for diagnosis support of patients with facialis paresis”. In: *Kuenstliche Intelligenz (Kl)* 3/2000 (2000).

[3] S. He et al. “Quantitative analysis of facial paralysis using local binary patterns in biomedical videos”. In: *Transactions on Biomedical Engineering* 56.7 (2009), pp. 1864–1870.

[4] A. Gaber, M. F. Taher, and M. A. Wahed. “Quantifying facial paralysis using the kinect v2”. In: *Engineering in Medicine and Biology Society (EMBC)*. 2015, pp. 2497–2501.

[5] T. Wang et al. “Automatic evaluation of the degree of facial nerve paralysis”. In: *Multimedia Tools and Applications* (2015), pp. 1–16.

[6] H. S. Kim et al. “A smartphone-based automatic diagnosis system for facial nerve palsy”. In: *Sensors* 15.10 (2015), pp. 26756–26768.

[7] D. Haase et al. “Automated and objective action coding of facial expressions in patients with acute facial palsy”. In: *European Archives of Oto-Rhino-Laryngology* 272.5 (2015), pp. 1259–1267.

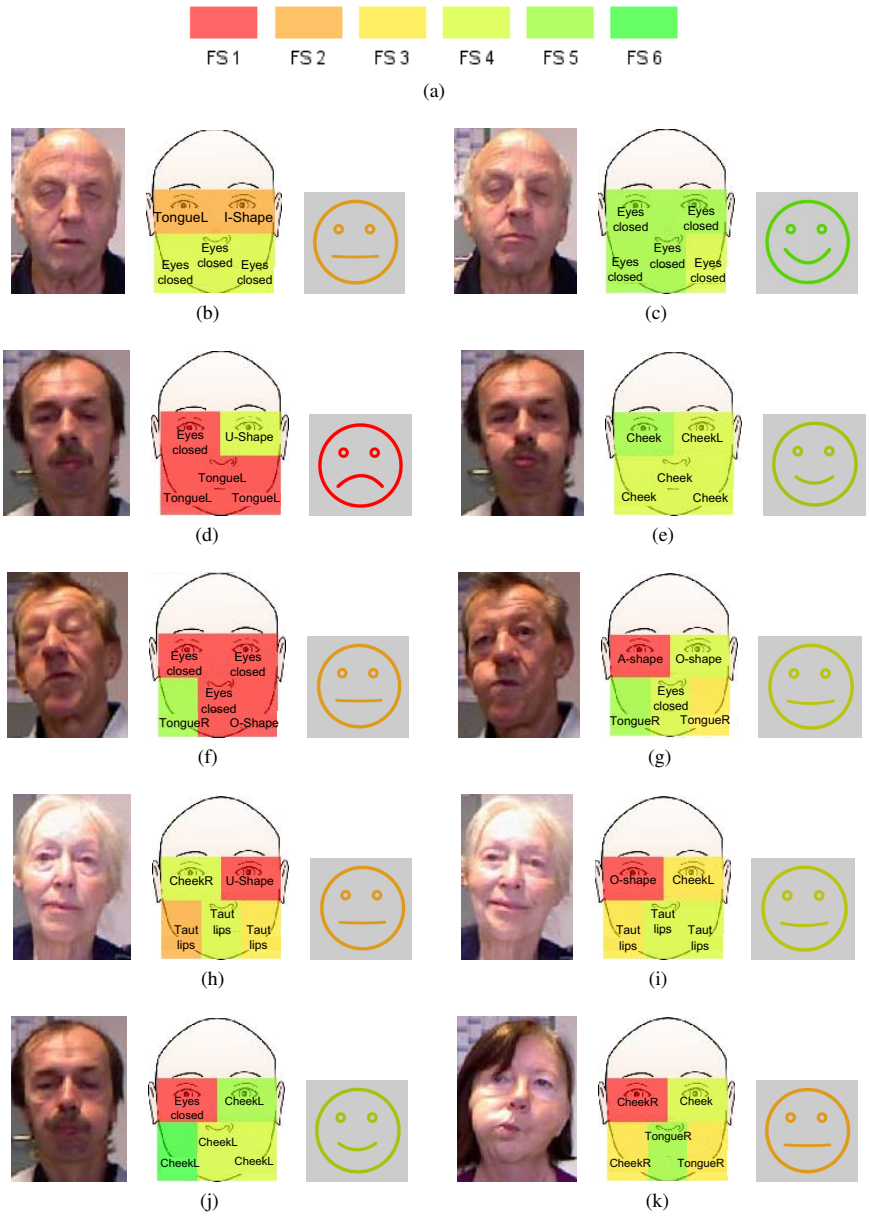


Fig. 7: Typical results of our feedback estimation. The smiley provides the estimated global feedback level (see Tab. I). Additionally, local feedback levels and classification results are given. **(a)** Color legend for the local feedback level. **(b)**-**(c)** Target exercise (TE): *eyes closed* **(d)**-**(e)** TE: *Cheek* **(f)**-**(g)** TE: *TongueR* **(h)**-**(i)** TE: *Taut lips* **(j)** TE: *CheekL* **(k)** TE: *TongueR*.

[8] L. Modersohn and J. Denzler. "Facial paresis index prediction by exploiting active appearance models for compact discriminative features". In: *Int'l Conf. on Computer Vision Theory and Applications (VISAPP)*. 2016.

[9] D. Jayatilake et al. "A wearable robot mask to support rehabilitation of facial paralysis". In: *Int'l Conf. on Biomedical Robotics and Biomechatronics (BioRob)*. 2012, pp. 1549–1554.

[10] T. Tasneem, A. Shome, and S. Alamgir Hossain. "A gaming approach in physical therapy for facial nerve paralysis patient". In: *Int'l Conf. on Computer and Information Technology (ICCIT)*. 2014, pp. 345–349.

[11] Y.-X. Wang, L.-Y. Lo, and M.-C. Hu. "Eat as much as you can: A kinect-based facial rehabilitation game based on mouth and tongue movements". In: *Int'l Conf. on Multimedia*. 2014, pp. 743–744.

[12] T. Fang et al. "3d facial expression recognition: A perspective on promises and challenges". In: *Automatic*

Face and Gesture Recognition and Workshops. 2011, pp. 603–610.

[13] G. Sandbach et al. "Static and dynamic 3D facial expression recognition: A comprehensive survey". In: *Image and Vision Computing* 30.10 (2012), pp. 683–697.

[14] H. Soyel and H. Demirel. "Facial expression recognition using 3D facial feature distances". In: *Image Analysis and Recognition*. Springer, 2007, pp. 831–838.

[15] H. Soyel and H. Demirel. "3D facial expression recognition with geometrically localized facial features". In: *Int'l Symposium on Computer and Information Sciences (ISCIS)*. 2008, pp. 1–4.

[16] X. Li, Q. Ruan, and Y. Ming. "3D facial expression recognition based on basic geometric features". In: *Int'l Conf. on Signal Processing (ICSP)*. 2010, pp. 1366–1369.

[17] C. Li and A. Soares. "Automatic facial expression recognition using 3D faces". In: *Int'l Journal of Engineering Research & Innovation* 3.1 (2011).

[18] H. Rabiu et al. "3D facial expression recognition using maximum relevance minimum redundancy geometrical features". In: *EURASIP Journal on Advances in Signal Processing* 2012.1 (2012), pp. 1–8.

[19] H. Tang and T. S. Huang. "3D facial expression recognition based on automatically selected features". In: *Computer Vision and Pattern Recognition Workshops (CVPRW)*. 2008, pp. 1–8.

[20] H. Tang and T. S. Huang. "3D facial expression recognition based on properties of line segments connecting facial feature points". In: *Automatic Face and Gesture Recognition (FG)*. 2008, pp. 1–6.

[21] R. Srivastava and S. Roy. "3D facial expression recognition using residues". In: *TENCON 2009-2009 IEEE Region 10 Conf.* 2009, pp. 1–5.

[22] J. Wang et al. "3D facial expression recognition based on primitive surface feature distribution". In: *Computer Vision and Pattern Recognition*. Vol. 2. 2006, pp. 1399–1406.

[23] P. Wang et al. "Quantifying facial expression abnormality in schizophrenia by combining 2D and 3D features". In: *Computer Vision and Pattern Recognition (CVPR)*. 2007, pp. 1–8.

[24] L. Broadbent et al. "2.5 d facial expression recognition using photometric stereo and the area weighted histogram of shape index". In: *Int'l Symposium on Robot and Human Interactive Communication (RO-MAN)*. 2012, pp. 490–495.

[25] P. Lemaire et al. "Fully automatic 3D facial expression recognition using differential mean curvature maps and histograms of oriented gradients". In: *Int'l Conf. and Workshops on Automatic Face and Gesture Recognition (FG)*. 2013, pp. 1–7.

[26] A. Savran, R. Gur, and R. Verma. "Automatic detection of emotion valence on faces using consumer depth cameras". In: *Int'l Conf. on Computer Vision Workshops (ICCVW)*. 2013, pp. 75–82.

[27] Y. Wang, M. Meng, and Q. Zhen. "Learning encoded facial curvature information for 3D facial emotion recognition". In: *Int'l Conf. on Image and Graphics (ICIG)*. 2013, pp. 529–532.

[28] G. Sandbach, S. Zafeiriou, and M. Pantic. "Binary pattern analysis for 3D facial action unit detection". In: *British Machine Vision Conference (BMVC)*. 2012.

[29] G. Sandbach, S. Zafeiriou, and M. Pantic. "Local normal binary patterns for 3D facial action unit detection". In: *Int'l Conf. on Image Processing (ICIP)*. 2012, pp. 1813–1816.

[30] N. Bayramoglu, G. Zhao, and M. Pietikainen. "CS-3DLBP and geometry based person independent 3D facial action unit detection". In: *Int'l Conf. on Biometrics (ICB)*. 2013, pp. 1–6.

[31] T. Huynh, R. Min, and J.-L. Dugelay. "An efficient LBP-based descriptor for facial depth images applied to gender recognition using RGB-D face data". In: *Asian Conf. on Computer Vision - Workshops*. 2013, pp. 133–145.

[32] A. Asthana et al. "Incremental face alignment in the wild". In: *Conf. on Computer Vision and Pattern Recognition (CVPR)*. 2014, pp. 1859–1866.

[33] K. Khoshelham and S. O. Elberink. "Accuracy and resolution of kinect depth data for indoor mapping applications". In: *Sensors* 12.2 (2012), pp. 1437–1454.

[34] P. J. Besl and R. C. Jain. "Invariant surface characteristics for 3D object recognition in range images". In: *Computer Vision, Graphics, and Image Processing* 33.1 (1986), pp. 33–80.

[35] C. Lanz et al. "Automated classification of therapeutic face exercises using the kinect". In: *Int'l Conf. on Computer Vision Theory and Applications (VISAPP)*. Barcelona, Spain, 2013, pp. 556–565.

[36] C. Dittmar, J. Denzler, and H.-M. Gross. "Facial movement dysfunctions: Conceptual design of a therapy-accompanying training system". In: *Ambient Assisted Living*. Advanced Technologies and Societal Change. Springer Berlin Heidelberg, 2014, pp. 123–141.

[37] C.-S. Chua, F. Han, and Y.-K. Ho. "3D human face recognition using point signature". In: *Int'l Conf. on Automatic Face and Gesture Recognition*. 2000, pp. 233–238.

[38] C.-C. Chang and C.-J. Lin. "LIBSVM: A library for support vector machines". In: *ACM Transactions on Intelligent Systems and Technology* 2.3 (3 2011). Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`, 27:1–27:27.

[39] E. Schaffernicht et al. "On estimating mutual information for feature selection". In: *Int'l Conf. on Artificial Neural Networks*. Springer. 2010, pp. 362–367.