論文題目　Study on Facial Expression Analysis based on
3D Deformable Model
（3次元変形モデルに基づく顔表情解析に関する研究）

氏　　名　　イェニ　ラースロ　アッティラ

Our everyday communication is highly influenced by the emotional information available to us about our partner. Facial expression and body language are the main sources of this information. Thus, recognition of facial expression is highly relevant for human-computer interaction and may gain broad applications in video annotation, situation analysis of social interactions.

In the last decade many approaches have been proposed for automatic facial expression recognition. We are experiencing a breakthrough in this field due to two factors;
1. high quality databases that have been made available to everybody, like the marked Cohn-Kanade Extended Facial Expression Database (CK+), its enhanced version and the dynamic 3D facial expression database (BU-4DFE) as well as
2. the advance of learning algorithms, most notably the advance of constrained local models (CLM)

Recently, very good results have been achieved by means of textural information. On the other hand, shape of the face extracted by active appearance models (AAM) showed relatively poor performance.

Line drawings, however, can express facial expressions very well, so shape information could also be a good descriptor of emotions. Shape – as opposed to texture – is attractive for facial expression recognition since it should be robust against rotations and may be robust against light conditions.

The author studied facial expression recognition using all available information of the shape. A preliminary study was conducted: the full shape information was used provided through the 2D landmarks (i.e., no PCA compression was used) and applied Procrustes normalization. The result is close to 100% performance in frontal faces, indicating that the compression inherent in AAM was responsible for the relatively poor performance. This result was similar to human performance.

Motivated by the findings in the preliminary study a novel algorithm was developed to extract shape information. The approach adapted the CLM method and extended it to exclude occluded landmarks and maintain high precision shape information even for large degrees of head rotations. The main result is that shape information and CLM based automated marker generation gives rise to state-of-the-art performance and it is robust against pose changes.

A variety of evaluations were used to study performance of shape representations for facial expression recognition in order to have the ability of comparing results achieved with different methods and on different databases. In all studies, multi-class support vector machine (SVM) classification was applied, both on expert annotated frontal databases as well as on 3D dynamic datasets. One of the main achievements of the study is that the neutral shape can be recovered from temporal sequences quickly and the AU0 based normalization can be replaced with the personal mean shape. For the expert annotated CK+ database results changed only slightly, but for CLM estimations this method gave rise to considerable improvements. The difference is in the noise of CLM based AU0 estimation, which is larger than the discrepancy between AU0 values (as determined by the experts) and mean shape values (as determined by averaging over the shapes of the same person). Note that personal mean shape can be estimated in a number of ways. The personal mean shape is better if it is averaged over the neutral shape and the shapes of the different facial expressions. This intriguing result has great promises for practice, since it allows for online updates of the mean shape with different time windows allowing for the detection of slight changes in the mood and also for the estimation of the more-and-more precise personal mean shape.

Furthermore, personal mean shape normalization gave rise to very good results for the case when shape information was used exclusively. The results surpass performances of the best available AAM methods and CLM that utilize shape plus texture and temporal information, respectively. It may be worth noting that human Facial Action Coding (FACS) experts work from using local textural information and further improvements are expected by using this valuable additional piece of information. The author studied the robustness of the CLM method for yaw rotations. Rotated 3D faces were rendered using the BU-4DFE database and found that CLM based shape estimation and shape based emotion recognition are highly robust against such pose variations. The results compare favorably to other methods in the literature, although only shape information was applied for emotion estimation. Further performance gains can be expected if textural information and if temporal information are included. Smoothing over time, e.g., by Hidden Markov Models seems crucial for sensitive detection of emotions and Action Unit (AU) intensities. Practice of FACS coders point to textural information, whereas noise filtering is better if temporal information is exploited.

For situation analysis, recognition of facial expressions beyond the basic emotions is of high relevance and such information is encoded into the AU intensities. In turn, CLM's precision for AU estimation was studied.

The Enhanced CK dataset was used to tune binary linear SVM for deciding if an AU was active or not, and Least Square-SVR for the intensity estimation. The proposed method compares favorably with the other methods and it is able to handle high degree of head pose rotations.

In sum, shape information is very efficient for facial expression recognition provided that details of shape changes are spared in the shape representation. This can be of high value in situation analysis, since shape estimation is robust against pose variations. Further improvements are expected for methods that include textural and temporal information. Such improvements are necessary for situation analysis and human-computer interaction.