UNIVERSITY OF AMSTERDAM

Enabling Dynamics in Face Analysis
H. Dìbeklìoğlu

Summary

In this thesis, we have focused on the temporal changes in the human face and analyzed dynamic patterns to improve facial analysis methods. We have proposed an accurate facial landmarking method to enable accurate analysis of facial movements. Dynamic descriptors have been introduced to reveal the temporal patterns of facial expressions. Different frameworks have been proposed to enable temporal information and facial expression dynamics in spontaneity analysis, age estimation, and kinship verification. Additionally, an affect-responsive system has been designed to create an adaptive application empowered by face-to-face human-computer interaction. In what follows, each of the contributions is separately summarized.

Chapter 2: Facial Landmark Localization.
A robust and accurate automatic facial landmarking is vital in face analysis to obtain reliable results. To this end, in Chapter 2, we propose a statistical method for automatic facial landmark localization. Our landmarking approach relies on a parsimonious mixture model of Gabor wavelet features, computed in coarse-to-fine fashion and complemented with a shape prior. The flexible statistical models we use increase the accuracy of landmarking.

The accuracy and robustness of the proposed approach are assessed in the most extensive cross-database experiments conducted in the literature, performed on four face datasets (FRGC, Cohn-Kanade, Bosphorus, and BioID). For 17 landmarks, our method achieves 99.33% accuracy on the Bosphorus database and 97.62% accuracy on the BioID database on the average, which improves the state of the art. We show that the method is not significantly affected by low resolution images, small rotations, facial expressions and natural occlusions such as beard and mustache. We further test the goodness of the landmarks in a facial expression recognition application, and report landmarking induced improvement over baseline on two separate databases for video-based expression recognition (Cohn-Kanade and BU-4DFE).

Chapter 3: Recognition of Genuine Smiles.
Automatic distinction between spontaneous and posed expressions is important for visual analysis of social signals. Methods for this task based on the eyebrow, lip corner, and shoulder movements, as well as on facial appearance exist in the literature. However, the reliability of such methods are not tested on a large scale database. Besides, none of these methods analyze

the dynamics of individual facial areas (such as eyelid and cheek movements).

In Chapter 3, we describe an informative set of features for the analysis of regional face dynamics, and propose a completely automatic system to distinguish between genuine and posed enjoyment smiles. To this end, we have collected and annotated the large scale UvA-NEMO smile database and investigated the eyelid, cheek and lip corner movements during spontaneous and posed enjoyment smiles. This database and our research on it was widely publicized, and was described in the ACM Tech News Bulletin.

For the smile analysis application, several fusion strategies over region-based classifiers and over temporal phase-based classifiers are evaluated. An online version of the proposed method is implemented to understand how partial information would fare, and to demonstrate the success of estimation during the progression of the smile. Moreover, effects of age and gender on smile dynamics are analyzed in detail. Our findings reveal the importance of eyelid dynamics for spontaneity analysis, and show the differences in expression dynamics between young people and adults. The accuracy of our method is evaluated on BBC, SPOS, MMI and UvA-NEMO databases, and it is shown that it improves the state of the art.

Chapter 4: Enabling Facial Dynamics in Age Estimation.
Estimation of a person's age from the facial image has many applications, ranging from biometrics and access control to cosmetics and entertainment. In general, existing approaches to this problem use appearance features only. On the other hand, our findings in Chapter 3 show the relation between age and expression dynamics. Therefore, in Chapter 4, we analyze the surface deformation dynamics on different facial regions and assess their use in age estimation. We further propose a novel two-stage age estimation architecture, which adaptively selects the age range for each classifier in the first stage.

In our experiments, we use the UvA-NEMO smile database with an age range between 8 and 76, which allows us to extensively analyze the relevant aspects of the problem. Our results show that 1) the dynamics of facial surfaces can be used to improve image-based age estimation; 2) considered alone, static image-based features are more accurate than dynamic features. Four different appearance features are evaluated, and it is shown that improvement by facial dynamics is consistent across representations. We test our approach extensively, including an exploration of spontaneous versus posed smile dynamics, and gender-specific age estimation. We show that dynamic features using spontaneity information reduces the mean absolute error by up to 21%, advancing the state of the art for facial age estimation.

Chapter 5: Facial Expression Dynamics for Kinship Verification.
All the methods proposed so far to verify kinship rely solely on facial appearance and work with images. In contrast to all published material in the literature, Chapter 5 explores the possibility of employing facial expression dynamics for kinship verification. To this end, similarities between kin pairs in terms of movement dynamics and spatio-temporal appearance of eyebrows, eyes, cheeks and lip corners are analyzed. Since there is no kinship database that includes facial expression videos of relatives, we have obtained the kinship annotations of the UvA-NEMO Smile Database. By selecting the spontaneous and posed enjoyment smiles of the subjects who have kin relationships, we construct a kinship database.

The proposed method is assessed for different kin relationships such as sister-sister, brother-brother, sister-brother, mother-daughter, mother-son, father-daughter, and father-son. Our results show that it is possible to improve the state of the art in this problem, and verify that it is indeed possible to recognize kinship by resemblance of facial expressions. Although expression dynamics alone cannot perform as accurate as spatio-temporal features, output-level fusion of these two individual feature sets provide a statistically significant improvement on the verification accuracy. On the average, 72.89% kinship verification accuracy is achieved on spontaneous smiles.

Chapter 6: An Affect-Responsive Interactive Photo Frame.
In addition to our analysis on facial expression dynamics, in Chapter 6, we design an affect-responsive interactive photo-frame application that provides adaptive face-to-face interaction between an interactive photograph and its users. The proposed system relies on temporal analysis of facial expressions and activity levels of its users to select responses from a database of short video segments. This ever-growing database is automatically prepared by an offline analysis of user-uploaded videos. The resulting system matches its user's affect along dimensions of valence and arousal, and gradually adapts its response to each specific user. In an extended mode, two such systems are coupled and feed each other with visual content.

The strengths and weaknesses of the system are evaluated through a usability study employing ten users, where a Wizard-of-Oz response logic is contrasted with the fully automatic system that uses affective and activity-based features, either alone, or in tandem. Although automatic interaction cannot perform as well as Wizard-of-Oz response logic, users find the proposed affect-responsive system interesting and engaging. Besides, the experiments show that the dynamic user adaptation works efficiently and shortens the latency of affective system responses.