DYNAMIC FACIAL FEATURES FOR INHERENTLY SAFER FACE RECOGNITION

Davide Iengo¹, Michele Nappi¹, Stefano Ricciardi², Davide Vanore¹

¹ Department of Informatics, University of Salerno – Italy
² Department of Biosciences and Territory, University of Molise - Italy

The idea inspiring this work is that dynamic facial features (DFF) extracted from facial expressions while a sentence is pronounced, could possibly represent a salient and inherently safer biometric identifier, due to the greater difficulty in forging a time variable descriptor instead than a static one.

We investigated on how a set of geometrical features, defined as distances between landmarks located in the lower half of face, changes across time while a sentence is pronounced, to find the most effective yet compact representation. The features vectors built upon these time-series were used to train a deep feed-forward neural network. Testing in identification modality resulted in state-of-art recognition accuracy, and a remarkable robustness to how the sentence is pronounced.

Abstract

The experiments were performed using the OuluVS visual speech database. It includes the acquisition of 20 subjects each one repeating 10 sentences 5 times. The sentences are: "Hello", "Excuse me", "I am sorry", "Thank you", "Good bye", "See you", "Nice to meet you", "You are welcome", "How are you", "Have a good time". The whole gallery comprising 1000 videos was divided into two parts: one dedicated to the training phase (80%) and one to the test phase (20%). Two main experiments were conducted to assess the performance of the FFD descriptor for person identification, (see Table 1), while two additional experiments aimed to stress-test the robustness of the network (see Table 2) to challenging input (unknown sentences) and possible attacks (a fake video with no face motion at all).

Proposed approach

Acquisition is performed by recording a short video clip of a subject uttering a short sentence whose duration is typically around one second. Each frame in the sequence is then analyzed by a face detector that is further processed for face landmarks detection and localization. The facial landmarks found are key-points characterizing the face's shape. These points are connected to form a lattice, denser in the lower half of the face. Each segment connecting two landmarks represents one out of fourteen geometrical features. The Dynamic Facial Features descriptor is obtained as a sequence of time series $T_{sj}$, each related to the variation of a particular feature throughout the $N$ captured frames, smoothed by a Kalman filter. A fully-connected deep feed-forward neural network was trained on the previously extracted DFF vectors. In this network architecture, the optimal number of hidden levels has been experimentally determined in three hidden levels.

Experiments and results

The experiments were performed using the OuluVS visual speech database. It includes the acquisition of 20 subjects each one repeating 10 sentences 5 times. The sentences are: "Hello", "Excuse me", "I am sorry", "Thank you", "Good bye", "See you", "Nice to meet you", "You are welcome", "How are you", "Have a good time". The whole gallery comprising 1000 videos was divided into two parts: one dedicated to the training phase (80%) and one to the test phase (20%). Two main experiments were conducted to assess the performance of the FFD descriptor for person identification, (see Table 1), while two additional experiments aimed to stress-test the robustness of the network (see Table 2) to challenging input (unknown sentences) and possible attacks (a fake video with no face motion at all).

Conclusions

Dynamic Facial Features represent a highly accurate and inherently safe face descriptor, thanks to the dynamic characteristics which result hard to forge. An extension of this work could exploit the audio signal to further improve both accuracy and reliability of the method through either feature-level or score-level fusion strategy.