Università degli Studi di Salerno

Dottorato di Ricerca in Informatica e Ingegneria dell'Informazione
Ciclo 30 – a.a. 2016/2017

TESI DI DOTTORATO / PH.D. THESIS

Real-time face analysis for gender recognition on video sequences

ANTONIO GRECO

SUPERVISOR: **PROF. MARIO VENTO**

PHD PROGRAM DIRECTOR: **PROF. PASQUALE CHIACCHIO**

Dipartimento di Ingegneria dell'Informazione ed Elettrica
e Matematica Applicata
Dipartimento di Informatica

This research work has been produced with the aim of performing gender recognition in real-time on face images extracted from real video sequences. The task may appear easy for a human, but it is not so simple for a computer vision algorithm. Even on still images, the gender recognition classifiers have to deal with challenging problems mainly due to the possible face variations, in terms of age, ethnicity, pose, scale, occlusions and so on.
Additional challenges have to be taken into account when the face analysis is performed on images acquired in real scenarios with traditional surveillance cameras. Indeed, the people are unaware of the presence of the camera and their sudden movements, together with the low quality of the images, further stress the noise on the faces, which are affected by motion blur, different orientations and various scales. Moreover, the need of providing a single classification of a person (and not for each face image) in real-time imposes to design a fast gender recognition algorithm, able to track a person in different frames and to give the information about the gender quickly.
The real-time constraint acquires even more relevance considering that one of the goals of this research work is to design an algorithm suitable for an embedded vision architecture.
Finally, the task becomes even more challenging since there are not standard benchmarks and protocols for the evaluation of gender recognition algorithms.
In this thesis the attention has been firstly concentrated on the analysis of still images, in order to understand which are the most effective features for gender recognition. To this aim, a face alignment algorithm has been applied to the face images so as to normalize the pose and optimize the performance of the subsequent processing steps. Then two methods have been proposed for gender recognition on still images.
First, a multi-expert which combines the decisions of classifiers fed with handcrafted features has been evaluated. The pixel intensity values of face images, namely the raw features, the LBP histograms and the HOG features have been used to train three experts which takes their decision by taking into account, respectively, the information about color, texture and shape of a human face. The decisions of the single linear SVMs have been combined with a weighted voting rule, which demonstrated to be the most effective for the problem at hand.
Second, a SVM classifier with a chi-squared kernel based on trainable COSFIRE filters has been fused with an expert which rely on SURF features extracted in correspondence of certain facial landmarks. The complementarity of the two experts has been demonstrated and the decisions have been combined with a stacked classification scheme.
An experimental evaluation of all the methods has been carried out on the GENDER-FERET and the LFW datasets with a standard protocol, so allowing the possibility to perform a fair comparison of the results. Such evaluation proved that the couple COSFIRE-SURF is the one which achieves the best accuracy in all the cases (accuracy of 94.7% on GENDER-FERET and 99.4% on LFW), even compared with other state of the art methods. Anyway, the performance achieved by the multi-expert which rely on the fusion of RAW, LBP and HOG classifiers can also be considered very satisfying (accuracy of 93.0% on GENDER-FERET and 98.4% on LFW).
After the preliminary analysis carried out on still images, the research has been focused on video sequences. A new dataset, namely the UNISA-dataset, has been acquired in different real environments (university and supermarket) with classic surveillance cameras and a part of it has been made publicly available. In these video sequences people are unaware to be framed, so the face images are significantly more challenging than the ones available in the standard datasets.
Such benchmark has been used firstly for an extensive analysis of the processing time required by the above-mentioned gender recognition algorithms. The profiling activity demonstrates that the face alignment algorithm is very costly and it is not suitable for real-time elaboration, as well as the impossibility to use SURF and COSFIRE features.
Considering that the pixel intensity values are not reliable with face images not aligned, the analysis also allowed to choose the HOG expert for gender recognition on video sequences, since it is more efficient and effective than the LBP one.

Finally, the analysis shows that, although the HOG-based classifier is able to process images in real-time on classic server side architectures, it is not fast enough for embedded vision systems.
Starting from these observations, a multi-sensor architecture has been proposed. It consists of a smart camera dedicated to people counting, of a classic camera installed to capture the faces and of a low cost embedded device used for gender recognition. The idea is that the people counting camera sends a notification when a passage of at least one person is detected, indicating the position where the passage occurred. In this way, the gender recognition algorithm can be applied only on a subregion of the images where at least one face is present. Such architecture allows to process more high resolution images, so obtaining in parallel the possibility to maximize the accuracy and to recognize the gender of the persons in real-time on low cost devices. Moreover, the temporal coherency has been taken into account to associate the same identity to a person captured in different frames, obtaining an accuracy above 90%.
An extensive experimental evaluation on the UNISA-dataset proves the effectiveness of the proposed method and its suitability for gender recognition in real-time.