

The Big Brother Database: Evaluating Face Recognition in Smart Home Environments

Annalisa Franco, Dario Maio, and Davide Maltoni

C.d.L. Scienze dell'Informazione, Università di Bologna, via Sacchi 3, Cesena, Italy
DEIS – Viale Risorgimento, 2 – Bologna, Italy
{annalisa.franco,dario.maio,davide.maltoni}@unibo.it

Abstract. In this paper a preliminary study on template updating techniques for face recognition in home environments is presented. In particular a new database has been created specifically for this application, where the face images acquired are characterized by a great variability in terms of pose and illumination but the number of subjects is quite limited and a large amount of images can be exploited for intensive incremental learning. The steps of database creation and the characteristics of the data collected are described in detail. We believe such a database could be very useful to develop and optimize face recognition approaches for smart home environments. Moreover some preliminary results on incremental learning are provided and analyzed to evaluate the effects of incremental template updating on the recognition performance.

1 Introduction

Face is one of the most studied biometric characteristics and a huge literature exists on the topic [9] [14]. In recent years the attention of the scientific community has been focused on face recognition in real applications where the operating conditions are not optimal due to several factors such as the low quality of the acquired images and variable conditions of pose and illumination. The results of recent evaluations [12] prove that state-of-the-art algorithms perform well in controlled environments but still poorly work in the presence of critical conditions.

The focus of this paper is on a particular kind of applications, referred to as *smart home environment* in the literature, where some intelligence mechanisms personalize the environment (home, office, etc.) based on the presence of one or more subjects in the room. This scenario is characterized by specific constraints: the resolution of the images is typically low because they are usually acquired by low cost surveillance cameras that can be far from the subjects and the pose and illumination conditions are extremely variable since the subjects are observed during their daily activities. On the other hand the smart home environment offers some advantages: the number of users is usually limited and a huge amount of data can be continuously collected and exploited for incremental learning: in fact, the availability of so many images allows to progressively update and refine the user's template, initially created on the basis of a few training images. This continuous learning may help to better deal with the intrinsic variability of the face, the changes of look, and with the aging affect that, as well known, can significantly affect recognition accuracy.

In order to develop, test and optimize face recognition and template updating approaches for this particular context, specific datasets reflecting the above mentioned characteristics are needed: a huge set of images of a limited number of subjects, acquired daily for a long period while the subjects normally perform their daily activities. Actually several face databases are publicly available [9] [14], and some of them include images captured under variable pose and illumination conditions or in different sessions that allow to evaluate the aging effects on face recognition performance. For example, the BANCA [1] or the Notre Dame Human ID [4] databases have been acquired over a period of about 3 months, but the acquisition conditions are rather controlled and/or the number of different acquisition sessions is limited. To the best of our knowledge, none of the available datasets meets all the specified requirements. These motivations induced us to create a new ad-hoc database of images that we intend to make available to the scientific community. Acquiring from scratch such a large dataset is objectively difficult and very time consuming, so we decided to exploit an existing data source: the full video of one edition of the “Big Brother” reality show. A software tool has been developed to analyze the videos and to quickly detect and label the face images. The rest of this paper describes the database structure and its creation procedure: in particular the steps carried out and the software tools used are described in order to provide useful indications to create similar databases. Moreover some experimental results on template updating are given to evaluate the effects of incremental learning on the recognition performance.

The work is organized as follows. In section 2 the procedure for data extraction and labeling is presented; a detailed description of the database is reported in section 3 together with an analysis of the image quality, in section 4 some preliminary results on template updating are reported and, finally, in section 5 some concluding remarks are given.

2 Database Construction

The database has been created starting from the 2 DVDs made commercially available at the end of the 2006 “Big Brother” reality show produced for the Italian TV [7]. The videos contained in the DVDs summarize the 99 days of permanence of 20 participants in the house. In particular they represent the main events occurred (participants entering or leaving the house, loves, quarrels, weekly trials, etc), some of the participants’ auditions and a brief presentation of all the participants, their work in the normal life, their family, etc. Most of the videos are labeled with the day of recording: this allowed us to organize the frames extracted from the videos in different subsets, each related to a specific day.

The steps carried out for the creation of the database are described in the following. **Conversion of the DVD video files to AVI files.** A simple free conversion tool [3] was used to this purpose. Each video track was converted to an AVI file with the following settings: output frame rate 29.97, output size 720×576, video codec Indeo® Video 5.10. The number of AVI files extracted is 22, about 15 GB in total.

Frames extraction. About 3 frames per second have been extracted from each AVI file using the free software Image Grabber II [8]. The frames extracted (more than

42,000, ~3.3GB in jpg format) have been initially classified according to the day of recording.

Face detection and labeling. A software tool (see Fig. 1) has been developed to analyze the sequences of frames obtained in the previous steps. The tool provides the following functionalities:

- *Face detection.* This operation is fully automatic and consists of checking each image to verify the presence of one or more faces; this is performed by means of the face detector implemented in the OpenCv library [11], based on the AdaBoost approach proposed by Viola and Jones in [13]. A default classifier provided with the OpenCV library (haarcascade_frontalface_alt.xml) and trained for frontal images, has been used in this work.
- *Image labeling.* This procedure is semi-automatic and allows to assign an ID to the detected faces. The tool partially automates the operation, based on the idea that, within the same scene, the identity of the faces detected in similar positions does not change. A simple scene change detection method, based on the difference between subsequent images, has been implemented. At the beginning of a new scene the user is required to assign (by pointing and mouse clicking) the correct ID to the detected faces. In the subsequent frames of the same scene, if a face is detected approximately in the same position of a previously labeled face and the size of the two face windows is similar, the same ID is proposed. In case a new face is detected within an ongoing scene, the human intervention is required to assign the correct ID. When a scene change occurs, the labeling system is re-initialized. The false positives of the face location module can be simply ignored (i.e., no ID is assigned). The use of this tool allows to process/label a large number of frames in an acceptable amount of time.
- *Image storing.* The detected faces, independently of the original dimensions, are resized to 70×70 pixels and stored. The result is a set of very heterogeneous images in terms of quality since some faces, taken when the subject is close to the camera, have a higher resolution with respect to those normally acquired in the house. It is worth noting that only the false positives (images that do not correspond to a face) have been eliminated (actually a low percentage with respect to the faces correctly detected) and that no selection of the face images has been performed. For this reason some images are very hard to recognize, also for a human, due for example to partial occlusions, strong blurring effects or strange disguising.

The original Big Brother images, due to copyright protection, cannot be printed in this paper, but some example images, acquired in analogous conditions in a laboratory environment, are reported in Fig. 2. and Fig. 3. The two figures clearly demonstrate the high degree of variability observable in a home environment. In particular Fig. 2. shows that, even during a single acquisition session, the face aspect can drastically change due to pose and expression variations, occlusions or variable distance from the camera. The variations are even more evident when images of different acquisition sessions are compared (Fig. 3.).

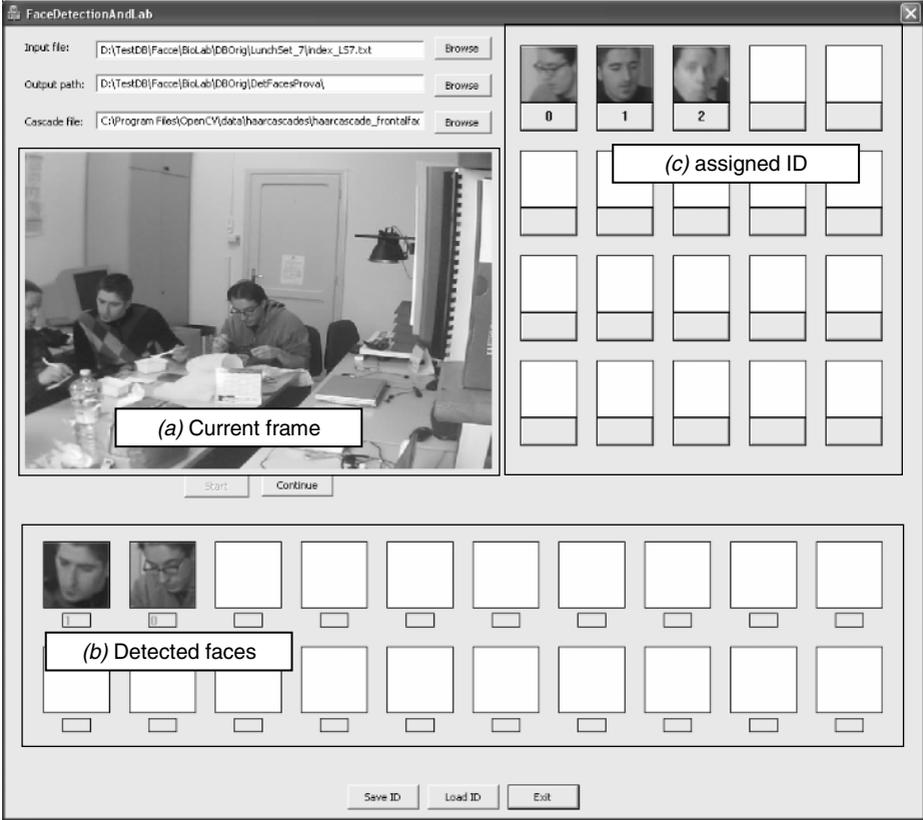


Fig. 1. A snapshot of the software tool developed for face detection and labeling. The three areas highlighted show the current frame analyzed (a), the faces detected in the current frame (b) and the ID assigned to each individual (c).



Fig. 2. Images of the same subject recorded during a single acquisition session



Fig. 3. Images of the same subject recorded during two different acquisition sessions

3 Database Description

The final database consists overall of 23842 images of size 70×70 pixels belonging to 18 subjects (one participant was immediately eliminated at the beginning of the reality show; no images are therefore available for him).

Since the aim of this work is to create a dataset that allows the training and evaluation of template updating approaches, the structure of the database is different from the traditional one, where the data are simply partitioned into *Training* and *Test sets*. Here an additional set of images, intended for template updating purposes, is provided so that the database is composed by the following subsets:

- *Training set*. Two sets of images are available for training: 651 images extracted from the video presenting the participants' biography (taken during the everyday life of each subject, at home or work, with friends, etc., low quality) and 5951 acquired during the participants' audition. The auditions video are not available for some of the participants.
- *Daily updates*. A set of 11898 images, hierarchically organized: the images are first partitioned according to the day they were recorded, and then according to the subject represented. Unfortunately only videos representing the main events have been included in the DVDs, so that images taken from 56 out of 99 days are available.
- *Testing set*. Some of the videos available do not report the day of recording. The images that cannot be included in the daily updates, since the temporal information is not available, can be used as testing set. These images refer to events occurred during the 99 days and are representative of the whole stay in the house. The number of images in this set is 5342.

It is worth noting the images for training can be gathered very quickly since about 200-300 images of a subject can be extracted from a video sequence of 1-2 minutes, if the acquisition conditions and the position of the individual are semi-controlled. On

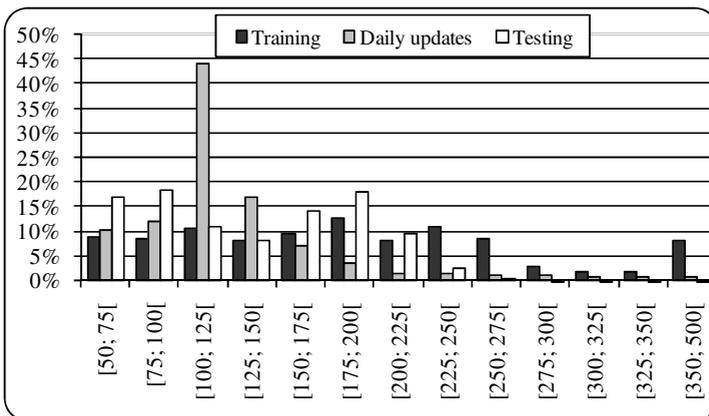


Fig. 4. Face size distribution in the training set, daily updates set and testing set. All the detected face windows are square and the size is simply the window side in pixel. Each bar represents the percentage of images of the related set having a size comprised in the specific interval.

the contrary, acquiring the images for updating is very hard since the acquisition conditions cannot be controlled and unfortunately the face detection module often fails to localize the individuals in the scene because they are too far from the camera or in particular poses.

To better understand the content of the database some statistics about the images are given. In particular the face size (before normalization) and the image quality distributions are reported in Fig. 4. and Fig. 5. respectively. In Fig. 4. the distribution of the face image size, as automatically detected, is reported for the training set, the daily updates and the testing set. It is worth noting that all the detected face windows are finally resized to a 70×70 pixels, that for most of the faces means a sub-sampling process. The graph shows that the original image size is quite variable: from 50 to more than 300 pixels in very few cases. However most of the detected face images have a size lower than 150 pixels.

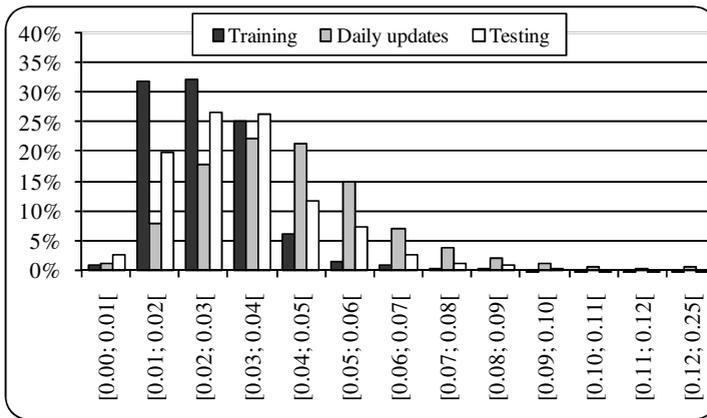


Fig. 5. Image quality distribution, evaluated according to the IQM measure, for the training set, daily updates set and testing set. Each bar represents the percentage of images of the related set having a quality score comprised in the specific interval.



Fig. 6. IQM score of some example images taken an office environment

To characterize the image quality (see Fig. 5.), the Image Quality Measure (IQM) has been adopted. IQM is based on the analysis of the power spectrum and is well known in the literature [10]. It takes into account several factors that proved to be strictly related to the human visual quality assessment. The graph shows that the image quality is generally quite low, particularly for the images of the testing set. As a

reference, some images taken in our laboratory with different IQM scores are shown in Fig. 6. Moreover it is worth noting that acceptable IQM scores do not necessarily guarantee that the image is easy to be recognized, since recognition can be seriously affected by other factors not measurable by IQM (e.g. particular poses or illumination, partial face occlusions, etc.).

4 Preliminary Learning Results

The final aim of this research is to design unsupervised learning techniques for face template updating; in this work some preliminary results obtained with a supervised updating technique are reported and analyzed to evaluate the effects of template updating on the recognition performance. Unsupervised leaning procedures will be studied in a future work.

In this experiment each individual is represented by a MKL space [2], i.e. a mixture of KL subspaces [6], created from the images in the *training set*. Recognition is base on a simple nearest neighbor criterion: the face image to be identified is associated to the individual represented by the closest MKL space according to the “distance from space” metric [2]. The initial templates are incrementally updated with the images in the *daily updates* set. An efficient procedure for MKL spaces merging [5] is used to avoid a complete recalculation of the templates which would be computationally very expensive. Updating is here done in a supervised manner: we incrementally update the templates by using images of corresponding identity. This constitute a simplification of a real scenario where the system itself has to decide whether to use a particular (unlabelled) image to update a specific template.

In this work we are focusing on a typical closed-set identification scenario (i.e., all the subjects are known to the system) and the performance are therefore measured in terms of recognition rate defined as the percentage of images in the testing set correctly recognized.

The results obtained are reported in Fig. 7 and Fig. 8. In particular in Fig. 7 the recognition rate is measured for each individual at different stages:

- *Initial RR*: represents the initial recognition rate, i.e. the baseline performance obtained with the templates calculated from the images in the *training set*.
- *RR last day*: refers to the recognition rate measured the last day of presence of each individual in the house.
- *RR max*: represents the maximum recognition rate measured for each subject during his/her whole presence in the house.

The results clearly show that template updating produces a very positive effect: the graph in Fig. 8 shows that the overall daily recognition rate rises from 33.60% (G0, initial template) to 56.27% (G99, templates updated with all the images in the *daily updates*). In most cases a considerable improvement is measured, in particular when a significant set of images is available. Moreover it is interesting to note the noticeable improvement measured for some individuals whose initial recognition rate was very low (e.g., subject 14), thus confirming the importance of template updating.

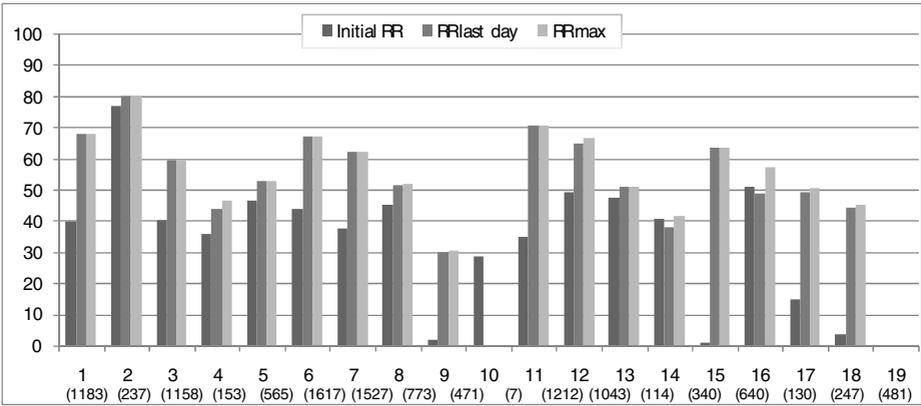
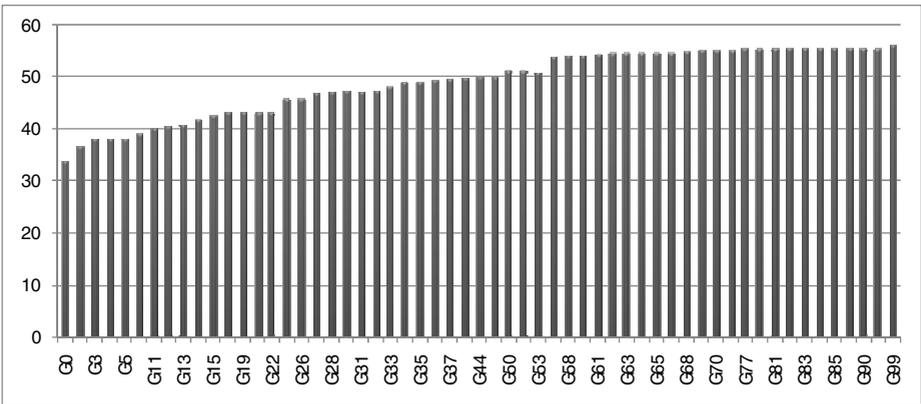


Fig. 7. Recognition rate of the 18 individuals, measured at different stages of the reality show. The *x*-axis reports the subject id and the total number of images available in the *daily updates* set for that individual.



acquiring from scratch such a huge set of data, characterized by a great variability in terms of pose and illumination conditions, image quality and, above all, acquired from the same set of subjects observed daily for a long period. The analysis of the data obtained, on the basis of image quality and face image size, shows that the dataset created is complex and represents an interesting challenge for face recognition algorithms. Moreover some preliminary results on template updating are provided; the experiments carried out show that incremental learning allows to considerably improve the initial recognition performance.

At this stage copyright issues do not allow us to distribute the database created to the scientific community, but we are exploring the possibility of releasing a batch procedure to recreate it starting from the original source (DVDs) that can be purchased online for a few USD.

The work carried out to create the Big Brother database is of course just a first step towards the main objective of our research, which is to design effective unsupervised template updating techniques exploiting the peculiar information that can be obtained from a continuous observation of subjects in smart home environments.

References

- [1] Bailly-Bailliere, E., Bengio, S., Bimbot, F., Hamouz, M., Kittler, J., Mariethoz, J., Matas, J., Messer, K., Popovici, V., Poree, F., Ruiz, B., Thiran, J.P.: The BANCA database and evaluation protocol. In: Kittler, J., Nixon, M.S. (eds.) AVBPA 2003. LNCS, vol. 2688, pp. 625–638. Springer, Heidelberg (2003)
- [2] Cappelli, R., Maio, D., Maltoni, D.: Multi-space KL for Pattern Representation and Classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23(9), 977–996 (2001)
- [3] DVDx, <http://www.labdv.com/dvdx/>
- [4] Flynn, P., Bowyer, K., Phillips, P.J.: Assessment of time dependency in face recognition: an initial study. In: Kittler, J., Nixon, M.S. (eds.) AVBPA 2003. LNCS, vol. 2688, pp. 44–51. Springer, Heidelberg (2003)
- [5] Franco, A., Lumini, A.: Mixture of KL Subspaces for relevance feedback. *Multimedia Tools and Applications* 37(2), 189–209 (2008)
- [6] Fukunaga, K.: *Statistical Pattern Recognition*. Academic Press, San Diego (1990)
- [7] Grande Fratello: Edizione 2006. Tutti i 92 giorni (2006), <http://www.internetbookshop.it/dvd/8032442208456/grande-fratello-edizione.html>
- [8] Image Grabber II, <http://www.videohelp.com/tools/ImageGrabber>
- [9] Li, S.Z., Jain, A.K.: *Handbook of Face Recognition*. Springer, Heidelberg (2005)
- [10] Nill, N.B., Bouzas, B.H.: Objective Image Quality Measure Derived from Digital Image Power Spectra. *Optical Engineering* 31(4), 813–825 (1992)
- [11] OpenCV Library, <http://www.intel.com/technology/computing/opencv/>
- [12] Phillips, P.J., Scruggs, W.T., O’Toole, A.J., Flynn, P.J., Bowyer, K.W., Schott, C.L., Sharpe, M.: FRVT 2006 and ICE 2006 Large-Scale Results (2006), <http://www.frvt.org/FRVT2006/docs/FRVT2006andICE2006LargeScaleReport.pdf2007>
- [13] Viola, P., Jones, M.J.: Rapid Object Detection using a Boosted Cascade of Simple Features. In: *Proc. ICPR*, vol. 1, pp. 511–518 (2001)
- [14] Wechsler, H.: *Reliable Face Recognition Methods*. Springer, Heidelberg (2007)