Unconstrained Facial Images: Database for Face Recognition under Real-world Conditions*

Ladislav Lenc[1,2] and Pavel Král[1,2]

[1] Dept. of Computer Science & Engineering
University of West Bohemia
Plzeň, Czech Republic
[2] NTIS - New Technologies for the Information Society
University of West Bohemia
Plzeň, Czech Republic
{llenc,pkral}@kiv.zcu.cz

Abstract. The objective of this paper is to introduce a novel face database. It is composed of face images taken in real-world conditions and is freely available for research purposes at http://ufi.kiv.zcu.cz. We have created this dataset in order to facilitate to researchers a straightforward comparison and evaluation of their face recognition approaches under "very difficult" conditions. It is composed of two partitions. The first one, called *Cropped images*, contains automatically detected faces from photographs. The number of individuals is 605. These images are cropped and resized to have approximately the same face size. Images in the second partition, called *Large images*, contain not only faces, however some background objects are also present. Therefore, it is necessary to include the face detection task before the face recognition itself. This partition contains images of 530 individuals. Another contribution of this paper is to show the recognition accuracy of several state-of-the-art face recognition approaches on this dataset to provide a baseline score for further research.

Keywords: Unconstrained Facial Images, UFI, face database, face recognition, unconstrained conditions

1 Introduction

Face recognition has become a mature research field and the amount of approaches published every year is very high. We could state that the problem is already well solved but it is always not true. It holds only for the cases when the images are sufficiently well aligned and have limited amount of variations. Face recognition under general *unconstrained* conditions still remains a very challenging task.

Since the beginning of the era of computerized face recognition there have existed an important issue with a straightforward comparison and interpretation of the results.

* This work has been partly supported by the project LO1506 of the Czech Ministry of Education, Youth and Sports. We would like also to thank Czech New Agency (ČTK) for support and for providing the data.

Evaluation of the developed methods was often done on different databases. This issue was fortunately recognized very early. The FERET [1] database and a clearly defined testing protocol was designed in 1993. The FERET program motivated by the Defense Advanced Research Projects Agency (DARPA) brought a significant progress in the face recognition field.

It may seem straightforward to compare the methods on such a dataset but there may also emerge problems with the comparison of the results. The authors usually crop the faces according to the eye positions and the size of the resulting image can thus differ. This may cause differences in the recognition accuracy. An interesting comparison is available in [2] where three well known techniques (PCA, LDA and ICA) are compared on exactly the same data and the results are sometimes in contradiction with the previously reported ones of other authors.

There was much work done since the origin of the FERET database and some new datasets have been created since then. An important issue is that the majority of them was created in more or less controlled environment. There are only few datasets, such as Labeled Faces in the Wild (LFW) [3], Labeled Wikipedia Faces (LWF) [4], Surveillance Cameras Face Database (SCface) [5] and FaceScrub [6] that are acquired in real conditions. Therefore, we believe there is still room for another challenging face database.

Therefore, we would like to introduce a novel real-world database that contains images extracted from real photographs acquired by reporters of a news agency. It is further reported as Unconstrained Facial Images (UFI) database and is mainly intended to be used for benchmarking of the face identification methods, however it is possible to use this corpus in many related tasks (e.g. face detection, verification, etc.).

We prepared two different partitions. The first one contains the cropped faces that were automatically extracted from the photographs using the Viola-Jones algorithm [7]. The face size is thus almost uniform and the images contain just a small portion of background. The images in the second partition have more background, the face size also significantly differs and the faces are not localized. The purpose of this set is to evaluate and compare complete face recognition systems where the face detection and extraction is included.

Together with the dataset description we provide a set of experiments realized on this corpus. We use several state-of-the-art feature based methods that perform well on the other databases and that give particularly good accuracy on real-world data. The results should serve as a baseline and we would like to encourage researchers to surpass these results.

The structure of this paper is as follows. Section 2 describes the most important databases used for face recognition. The following section introduces the created database and the testing protocol. Section 4 shows the baseline recognition results on this dataset. Finally, Section 5 concludes the paper and proposes some further possible improvements of this dataset.

2 Summary of the Main Face Databases

FERET Creation of this dataset [1] is connected with the FERET program that started in 1993. It was designed to allow a straightforward comparison of newly developed face recognition techniques under the same conditions. This database was acquired during 15 sessions within 3 years and contains 14,051 face images belonging to 1,199 individuals. The images are divided into the following categories according to the face pose: frontal, quarter-left, quarter-right, half-left, half-right, full-left and full-right. The images are also grouped into several probe sets. The main probe sets of the frontal images are summarized in Table 1.

Table 1: Image numbers in the main frontal probe sets of the FERET dataset.

Type	Description	Images no.
fa	face gallery (for training)	1,196
fb	different facial expressions	1,195
fc	different illuminations	194
dup1	obtained over a three year period	722
dup2	sub-set of the *dup1*	234

CMU PIE CMU PIE database [8] was created at the Carnegie Mellon University (CMU). It contains images of 68 people and the total number of images is 41,368. All the images were recorded in a single session. There are variations in pose (13 poses) and lighting conditions (43 different illumination conditions). The differences in facial expression are limited and can be categorized to 4 expressions.

Multi-PIE This database [9] builds on the success of the CMU PIE database. Its goal is to remove the shortcomings that the PIE database has. The number of individuals is 337 and the total number of images is 755,370. The images were taken under 15 different viewpoints and 19 lighting conditions. There were 4 recording sessions compared to just one in the case of the CMU PIE.

Yale Face Databases The original Yale Face Database [10] contains images of only 15 subjects, 11 images are available for each person. They differ in lighting conditions and in the details as for instance wearing glasses or not. This dataset was extended to the Yale Face Database B [11] which contains 16,128 face images of 28 individuals under 9 poses and 64 lighting conditions.

AT&T "The Database of Faces" AT&T database [12] (formerly known as "The ORL Database of Faces") was created at the AT&T Laboratories[3]. It contains facial images

[3] http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html

of 40 people that were captured between the years 1992 and 1994. 10 pictures for each person are available. The images have a black homogeneous background. They may vary due to three following factors: 1) time of acquisition; 2) head size and pose; 3) lighting conditions.

AR Face Database The AR Face Database[4] was created at the Universitat Autonòma de Barcelona. This database contains more than 4,000 colour images of 126 individuals. The individuals are captured under significantly different lighting conditions and with varying expressions. Another characteristic is a possible presence of glasses or scarf.

CAS-PEAL The creation of CAS-PEAL face database [13] was sponsored by the National Hi-Tech Program and ISVISION. It contains the faces of 1,040 Chinese people which represents in total 99,594 face images. The images differ in pose, expression, accessories (glasses and caps) and lighting. One part of this database called CAS-PEAL-R1 containing 30,900 images is available for the researchers.

Banca This database [14] was designed for testing of multi-modal verification systems. It consists of image and audio data and contains the images of 208 people. The images were captured under three different conditions: controlled, degraded and adverse.

Labeled Faces in the Wild Labeled Faces in the Wild (LFW) [3] is a database collected from the web. It contains the images of 5,749 people and the total number of images is more than 13,000. 1,680 people has two or more images. Its purpose is to test the face verification scenario under unconstrained conditions. There are four available sets. The first one is the original and the others are aligned using three different methods.

PubFig PubFig [15] database comprises also the images collected from the Internet. Compared to LFW, it has lower number of individuals (200). The total number of images is 58,797 and thus the number of images per person is much higher. There are significant differences in lighting, pose, expression, camera quality and other factors. This dataset is also used for the face verification.

Labeled Wikipedia Faces Labeled Wikipedia Faces (LWF) [4] is a large collection of images from Wikipedia biographic entries. It contains 1,500 individuals which represents 8,500 images in total. There are available the original raw images as well as the aligned ones. Compared to LFW, it contains also historical images of a particular person and the time span is thus very large.

SCface Surveillance Cameras Face Database (SCface) [5] was captured in indoor environment using 5 surveillance cameras of different qualities. It contains 4,160 images of 130 individuals. Some of the images are in the infrared spectrum.

[4] http://www2.ece.ohio-state.edu/ aleix/ARdatabase.html

FaceScrub FaceScrub [6] dataset was collected from the images available on the Internet. There is an automatic procedure that verifies that the image belongs to the right person. It contains the images of 530 people which is 107,818 in total. The images are provided together with the name and gender annotations.

The other thorough summaries of face databases can be found in [3] or in [16].

3 Unconstrained Facial Images Database

The Unconstrained Facial Images (UFI) database is composed of real photographs chosen from a large set of photos owned by the Czech News Agency(ČTK) [5]. Each photograph is annotated with the name of a person. However, some background objects and also other persons are often available. Due to a) financial/time constraints; b) necessity to be able to create quickly another face dataset on demand, we would like to create the database as automatic as possible (with minimal human efforts). We are inspired by [17] and we do thus a similar series of tasks in order to build the UFI database. As already mentioned, we created two different partitions.

3.1 Cropped Images: Creation & Dataset

The first step is face detection in the input images. We utilized the widely used Viola-Jones detector [7]. It is possible that the given photograph contains more than one person. In this case, we do not know which of the detected faces belongs to the correct person in annotation. In this step, we do not solve this problem and choose the first detected one. Another important issue is a presence of false detections (e.g. background objects instead of the faces) among the results. This issue will be addressed in the following steps.

Next, we detect the eyes in the detected faces. This step has two reasons: a) to remove a significant number of non-face images (false detections); b) to remove some face images that have significant out-of-plain rotation. The images with both eyes detected are then rotated to have the eyes on a horizontal line and resized to a specified size.

The resulting set of images is used as an input to the cleaning algorithm. The algorithm tries to chose the most similar images in the set of images for one person. Its aim is to remove the faces belonging to the other people and the possible non-face images that were not excluded in the previous step.

From the remaining images of each person we randomly choose one example for the test set. The remaining ones will be used for training. Finally, the database is manually checked to correct the possible errors.

The resulting set contains images of 605 people with an average of 7.1 images per person in the training set and one in the test set. The distribution of training examples per person is depicted in Figure 1. The images are cropped to a size of 128×128 pixels. Figure 2 shows the images of two individuals from the *Cropped images* partition.

[5] http://www.ctk.eu/

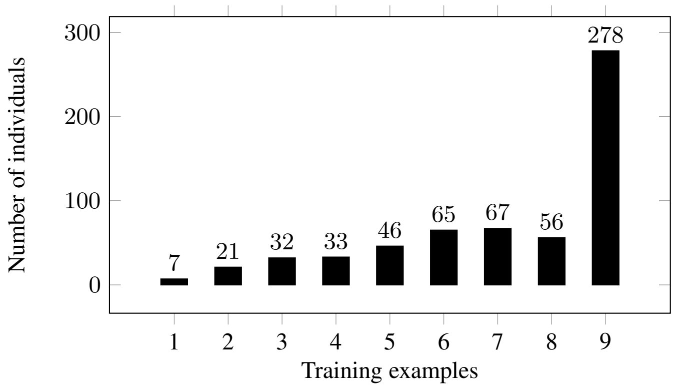

Fig. 1: Distribution of the training image numbers in the *Cropped images* partition.

Fig. 2: Example images of two individuals from the *Cropped images* partition.

3.2 Large Images: Creation & Dataset

First, we apply the three similar tasks as in the previous case (i.e. face detection, eye detection and rotation according to the eyes and cleaning algorithm).

Then, we randomly choose the image portion that the face should fill in the image and random shifts in both horizontal and vertical axis are made. The random values define the maximal size of the freely available images for research purposes (the size of all images in this partition is 384×384 pixels). After applying this step, the face can be located in any part of the resulting image. Moreover, the face size is not specified and can occupy the whole image as well as only a small part.

This procedure is followed by a manual checking and no additional alignment or rotation is performed. The total number of the subjects in this partition is 530 and an average number/person of training images is 8.2 . The distribution of the numbers of training examples is depicted in Figure 3. Figure 4 shows some example images from the *Large images* partition.

The main goal of this partition is to evaluate and compare complete face recognition systems. Therefore, additional steps before recognition itself are expected (face detection, background removal, etc.).

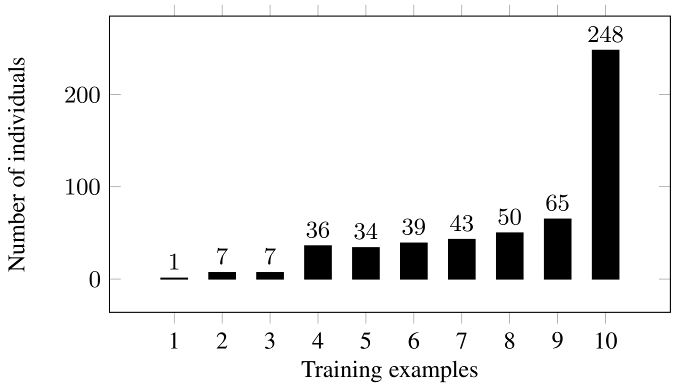

Fig. 3: Distribution of the training image numbers in the *Large images* partition.

Fig. 4: Example images of two individuals from the *Large images* partition.

3.3 Testing Protocol

We would like to keep the testing protocol as straightforward and simple as possible. Therefore, both partitions are divided into training and testing sets. All images from the training sets are available as a gallery for training. The test sets are used as test images.

The images in the *Cropped images* partition should be used in its original size. Additional cropping or resizing is undesirable because of the comparability of the results. The images may be preprocessed and the preprocessing procedure must be described together with the reported results.

On the other hand we allow any preprocessing or cropping in the case of *Large images* partition. However, the whole procedure must be reported and thoroughly described. The recognition results should be reported as an accuracy (i.e. ratio between correctly recognized faces and all the faces).

Database Structure The database is distributed in a directory structure. Each partition contains *train* and *test* directories which are composed of the sub-directories for each person named *sxxx* (*xxx* is the number of the subject).

4 Baseline Evaluation

This section provides a baseline evaluation of four selected methods on both partitions of the UFI database. As already stated in the introduction section, we concentrated on the state of the art feature based methods that perform better than the holistic ones under unconstrained conditions.

4.1 Face Recognition Algorithms

Histogram Sequence Histogram Sequence (HS) [18] is a method of creating the face descriptors from the local image operator values. This concept is common for most of the operators based on or similar to the Local Binary Pattern (LBP) and therefore it is briefly described in this section.

The image is first divided into rectangular regions according to a regular grid. In each of these regions a histogram of operator values is computed. The histograms are then concatenated into a vector called *histogram sequence* that is used as a descriptor. This method ensures that the corresponding image parts are correctly compared.

Although there are a lot of sophisticated classifiers that can be employed for classification of the face descriptors created using the HS, we chose the simple nearest neighbour algorithm for classification in this baseline evaluation. It is used in all following methods.

Local Binary Patterns The LBP operator [19] is based on a simple procedure that encodes a small neighbourhood of a pixel as follows: 8 neighbouring pixels are compared against the central one. The pixels with higher intensity are assigned to 1 and those with lower intensity are assigned to 0. The result is an eight bit binary number which corresponds to the decimal value in the interval $[0; 255]$.

The LBP operator was extended to use the points on a circle of given radius R that are compared to the central pixel. The number of the points is not fixed and is marked P. LBP operator in this form is referred to as $LBP_{P,R}$.

The LBP Histogram Sequences (LBPHS) were first used for face recognition by Ahonen in [20] and we use this method as the first baseline.

Local Derivative Patterns Local Derivative Patterns (LDP) operator was proposed in [21]. Its main difference against LBP is that it uses the features of higher order than the LBP operator. It thus should capture more information than LBP. We will refer next the face recognition method method as LDP Histogram Sequences (LDPHS).

Patterns of Oriented Edge Magnitudes (POEM) This operator [22] uses gradient magnitudes instead of the intensity values in LBP. The magnitudes of pixels within a *cell* (square region around the central pixel) are accumulated in a histogram of gradient orientations. The values for each orientation are then encoded using a circular LBP operator with a radius $L/2$. The circular neighbourhood of a pixel with a diameter L is called *block* in this method. The operator value is thus d-times longer (d is the number of discrete orientations). We will next refer to this method used for face recognition as POEM Histogram Sequences (POEMHS).

Face Specific LBP This method [23] differs from the previous ones in the way of the computing the image representation. First, the representative face points are detected automatically using Gabor wavelets (instead of the regularly defined grid). Then, the LBP histograms are created in the regions around these points in the same way as in the other three previous approaches. However, the face is not represented by a single descriptor but by a set of the features (histograms). No HS is used in this case because the features are compared individually. We will further refer to this method as Face Specific LBP (FS-LBP).

4.2 Results on the Cropped Images Partition

This section presents results of the four selected methods on the *Cropped images* partition. The images are used in their original form as defined in the testing protocol (see Sec. 3.3). The Histogram Intersection (HI) metric is used for descriptor comparison in all cases. The grid size is set to 13 for LBPHS, LDPHS and POEMHS. It means that the histograms are computed within the square regions of size the 13×13 pixels. The similar value is used also in the FS-LBP method where it cannot be referred as a grid but it has similar interpretation that the histograms are computed within 13×13 square region. We use the circular $LBP_{8,2}$ in the FS-LBP method. POEM descriptors are calculated using three gradient directions. The cell size is set to 7 and the block size to 10. The results reported for the LDP method use LDP of first order because it surprisingly reaches better accuracy than the higher ones.

Table 2 shows the results of the four selected baseline methods on this partition. This table shows that the best performing method is POEMHS. Surprisingly, LDPHS has the worst results on this partition.

Table 2: Recognition results of the baseline methods on the *Cropped images* partition.

Method	Accuracy in %
LBPHS	55.04
LDPHS	50.25
POEMHS	**67.11**
FS-LBP	63.31

Then, we have done some error analysis. Two incorrectly recognized face examples/method are depicted in Figure 5. This examples shows the complexity of this dataset where some examples are difficult to be correctly recognized even by humans.

4.3 Results on the Large Images Partition

As already stated, the recognition methods cannot be applied directly on the images in this partition. We therefore first applied the Viola-Jones algorithm to detect the faces. Additionally, we tried to detect the eyes and if both eyes were detected the faces were

Fig. 5: Examples of the incorrectly recognized face images from the *Cropped images* partition using LBPHS, LDPHS, POEMHS and FS-LBP methods (from top to bottom). Each triplet contains a probe image, corresponding gallery image, incorrectly recognized image (from left to right).

rotated and aligned according to the ayes. All resulting images were resized to the size 128×128 pixels. For the face recognition itself, we use the same configuration of the baseline methods as in the *Cropped images* case (see Section 4.2).

Table 3 summarizes the face recognition results on this partition. In this case, the best performing method is FS-LBP with score nearly 10% higher than the remaining methods. The other methods perform comparably.

Table 3: Recognition results of the baseline methods on the *Large images* partition.

Method	**Accuracy in %**
LBPHS	31.89
LDPHS	29.43
POEMHS	33.96
FS-LBP	**43.21**

Then, we have also realized some error analysis. Two incorrectly recognized face examples/method are depicted in Figure 6. This examples shows the complexity of this dataset even more clearly than the ones in the previous experiment.

Fig. 6: Examples of the incorrectly recognized face images from the *Large images* partition using LBPHS, LDPHS, POEMHS and FS-LBP methods (from top to bottom). Each triplet contains a probe image, corresponding gallery image, incorrectly recognized image (from left to right).

5 Conclusions

In this work, we presented a novel face database intended primarily for testing of the face recognition algorithms. It represents a challenging dataset that addresses the main issues of the current face recognition approaches, the performance on low quality real-world images. We provide a simple testing scenario that must be kept so that the results are directly comparable. Together with the dataset we provide a set of experiments that evaluate some state-of-the-art face recognition approaches on this dataset. The best obtained accuracy on *Cropped images* partition is 67.1% using the POEMHS method. The highest score on *Large images* partition is 43.2% obtained by the FS-LBP method. The database is freely available for research purposes[6].

One possible future work consists in adding the coordinates of the faces in the *Large images* partition and the coordinates of the important facial features. The dataset could then be used also for face detection and facial landmark detection algorithms.

References

1. Phillips, P.J., Moon, H., Rizvi, S., Rauss, P.J., et al.: The feret evaluation methodology for face-recognition algorithms. Pattern Analysis and Machine Intelligence, IEEE Transactions on **22** (2000) 1090–1104

[6] http://ufi.kiv.zcu.cz

2. Delac, K., Grgic, M., Grgic, S.: Independent comparative study of pca, ica, and lda on the feret data set. International Journal of Imaging Systems and Technology **15** (2005) 252
3. Huang, G.B., Ramesh, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical report, Technical Report 07-49, University of Massachusetts, Amherst (2007)
4. Hasan, M.K., Pal, C.: Experiments on visual information extraction with the faces of wikipedia. In: Twenty-Eighth AAAI Conference on Artificial Intelligence. (2014)
5. Grgic, M., Delac, K., Grgic, S.: Scface–surveillance cameras face database. Multimedia tools and applications **51** (2011) 863–879
6. Ng, H.W., Winkler, S.: A data-driven approach to cleaning large face datasets. In: Image Processing (ICIP), 2014 IEEE International Conference on, IEEE (2014) 343–347
7. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on. Volume 1., IEEE (2001) I–511
8. Sim, T., Baker, S., Bsat, M.: The cmu pose, illumination, and expression (pie) database. In: Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on, IEEE (2002) 46–51
9. Gross, R., Matthews, I., Cohn, J., Kanade, T., Baker, S.: Multi-pie. Image and Vision Computing **28** (2010) 807–813
10. Georghiades, A., et al.: Yale face database. Center for computational Vision and Control at Yale University, http://cvc. yale. edu/projects/yalefaces/yalefa (1997)
11. Georghiades, A.S., Belhumeur, P.N., Kriegman, D.J.: From few to many: Illumination cone models for face recognition under variable lighting and pose. Pattern Analysis and Machine Intelligence, IEEE Transactions on **23** (2001) 643–660
12. Jain, A.K., Li, S.Z.: Handbook of face recognition. Volume 1. Springer (2005)
13. Gao, W., Cao, B., Shan, S., Chen, X., Zhou, D., Zhang, X., Zhao, D.: The cas-peal large-scale chinese face database and baseline evaluations. Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on **38** (2008) 149–161
14. Bailly-Bailliére, E., Bengio, S., Bimbot, F., Hamouz, M., Kittler, J., Mariéthoz, J., Matas, J., Messer, K., Popovici, V., Porée, F., et al.: The banca database and evaluation protocol. In: Audio-and Video-Based Biometric Person Authentication, Springer (2003) 625–638
15. Kumar, N., Berg, A.C., Belhumeur, P.N., Nayar, S.K.: Attribute and simile classifiers for face verification. In: Computer Vision, 2009 IEEE 12th International Conference on, IEEE (2009) 365–372
16. Gross, R.: Face databases. In: Handbook of Face Recognition. Springer (2005) 301–327
17. Lenc, L., Král, P.: Automatic face recognition system based on the SIFT features. Computers & Electrical Engineering (2015)
18. Ahonen, T., Hadid, A., Pietikäinen, M.: Face recognition with local binary patterns. In: Computer vision-eccv 2004. Springer (2004) 469–481
19. Ojala, T., Pietikäinen, M., Harwood, D.: A comparative study of texture measures with classification based on featured distributions. Pattern recognition **29** (1996) 51–59
20. Ahonen, T., Hadid, A., Pietikainen, M.: Face description with local binary patterns: Application to face recognition. Pattern Analysis and Machine Intelligence, IEEE Transactions on **28** (2006) 2037–2041
21. Zhang, B., Gao, Y., Zhao, S., Liu, J.: Local derivative pattern versus local binary pattern: face recognition with high-order local pattern descriptor. Image Processing, IEEE Transactions on **19** (2010) 533–544
22. Vu, N.S., Dee, H.M., Caplier, A.: Face recognition using the poem descriptor. Pattern Recognition **45** (2012) 2478–2488
23. Lenc, L., Král, P.: Automatically detected feature positions for LBP based face recognition. In: Artificial Intelligence Applications and Innovations. Springer (2014) 246–255