Jurnal
Teknologi

Full Paper

A STATE OF THE ART COMPARISON OF DATABASES FOR FACIAL OCCLUSION

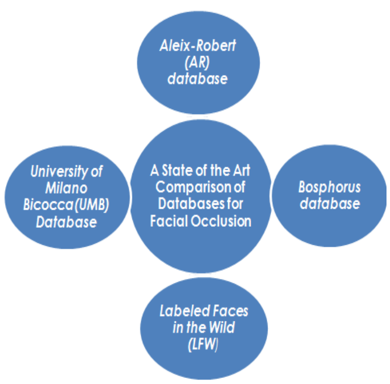
Abdulganiyu Abdu Yusuf[a,b*], Fatma Susilawati Mohamad[a], Zahraddeen Sufyanu[a]

[a]Faculty of Informatics and Computing, 21300 Gong Badak Campus, Universiti Sultan Zainal Abidin (UniSZA), Terengganu, Malaysia
[b]National Biotechnology Development Agency (NABDA), Abuja, Nigeria

Article history
Received
15 April 2015
Received in revised form
29 September 2015
Accepted
12 November 2015

*Corresponding author
abdulg720@gmail.com

Graphical abstract



Abstract

Face recognition continues to be one of the most popular research areas of image processing and computer vision. There are various face databases available to researchers for face detection and recognition. These databases are customized for a particular need of one algorithm. They are range in size, scope, and purpose. Few of these databases from the literature contain face occlusions in several positions of the faces to enable real world applications. In this paper, we present four different occlusion face databases. These are Aleix-Robert (AR), Bosphorus, Labeled Faces in the Wild (LFW), and University of Milano Bicocca Database (UMB) face databases. At each section, the key features of the database are presented with the recording conditions, though not all of them are discussed at the same level of details. Detailed comparisons of the databases were made based on controlled and uncontrolled databases, 2D and 3D databases and also their uniqueness. Comparison was also made with other databases out of the categorization mentioned. The databases are useful for performing a rigorous benchmarking of face detection and recognition algorithms.

Keywords: State-of-the-art, occlusion databases, face detection, face recognition

1.0 INTRODUCTION

The performance of face recognition system decreases drastically due to the presence of occlusion on the face. Consequently, the low performance prevents system from being widely deployed in real world applications (although many systems have been proposed, their accuracy is very low). It is therefore important to emphasize on the problem of occlusion since it is associated with several safety and severe security issues. Facial occlusions may occur for several (intentional or unintentional) reasons. For instance, some people wear veils for religious convictions or cultural habits. Also, facial accessories like sunglasses, scarf, facial make-up and cap are quite common in daily life. Handling all these occlusion instances in face recognition is essential to intelligence, security and law enforcement purposes. Face images are frequently part of many documents such as driving licenses, ID cards, and international passports. Collecting the required data in facial image is quite easier than in the case of other biometric modalities (e.g. fingerprints or iris images).

In order to train, build and reliably test face detection and recognition algorithms, considerable databases of face images are required. These databases were specifically acquired by teams of researchers for the purpose of face detection and recognition, and are therefore varied in sizes, scopes, and purposes. Several databases need to be recorded under various conditions and with different applications in mind, to enable comparison of different systems. A face database should contain enough variations in order to test the robustness of face detection and recognition algorithms. Facial changes such as occlusion, illumination, expression, and pose are generally considered in the face databases [1].

The accuracy of face detection and recognition algorithm largely depends on the data set used for training. Each algorithm uses different database to train and evaluate performance, so it is difficult to compare algorithms. Thus, there is need for globally accepted and error free dataset to compare the robustness of the algorithms.

This paper deals with most recent face databases on occlusion used for the face detection and recognition algorithms. These databases are notable and contain difficult types of the occlusion. They have been made available free of charge on the internet for researchers in the field. The main reason for carriying out this survey is to combine together details of some prominent facial occlusion databases into one document. To the authors' knowledge, very few studies of this kind (on occlusion databases) are conducted in the literature.

The rest of the paper is organized as follows: In section 2 we briefly introduces the concept of occlusion; key features of the databases are presented in section 3 focusing mainly on occlusion. Section 4 compares the performance of the databases and also with other related databases; concluding remark is drawn in section 5.

2.0 OCCLUSION

Various algorithms and constraints have been proposed to detect and recognize faces. Partial occlusions considerably change the original appearance of a face image, for this reason it significantly degrades the performances of face detection and recognition system [2] since the face representations are largely altered accordingly. Controlling partial occlusion is a critical issue to achieve robust face recognition. Most of the works in the literature focus on finding corruption-tolerant features or classifiers to reduce the effect of partial occlusions in face representation. There are Mainly two types of occlusions, The first one is self-occlusions during acquisition, where a part of the facial surface hinders acquisition of another region. This type of occlusion is caused by pose variations and appears as missing data on the facial surface. The second type of occlusion is caused by external objects such as hair, clothing material, hair, eyeglasses and other objects. This type of occlusion is more complex to handle since occluding objects alter the facial geometry. The problem of facial occlusion includes (i) transitions among occluding objects, (ii) image characteristics and level of description, (iii) reliability and validity of training and test data, (iii) individual differences among subjects in facial features, and (iv) head orientation and scene complexity.

A research conducted by [3] contains a list and comparisons of databases suited for the face detection and recognition, but only XM2VTS and AR had occlusions.

The main reason for chosen the aforementioned databases are, the databases are prominent with large number of subjects and contain difficult types of occlusion. Also, most of the recent studies on facial occlusion analysis from the literature employ at least one of these databases. For example, 3D Face Recognition under Occlusion using Masked Projection is proposed by [4]. Their main focus is to handle occlusions and missing information for surface enrolment. The recognition rate obtained using Bosphorus and UMB databases were 83.99 % and 65.25 % respectively. The author in [4] proposed a similar study in [5], where the proposed method is capable of detecting occluded face area to obtain occlusion-free surfaces, restoring or ignoring the missing parts using of statistical regional classifiers. Iterative Closest Point (ICP) and Gappy Principal Component Analysis (Gappy PCA) were the techniques used. Recognition of 94.23 % was obtained using Bosphorus database.

A statistical shape models technique for face recognition under occlusion variations is presented by [6]. The technique is employed to remove the occluded parts and to recover only missing data on the 3D scan. Databases used are FRGC v2 and Bosphorus, and Recognition rate of 97 % and 87 % were reported.

Moreover, an approach for detecting facial landmarks under various occlusions and expressions is proposed by Yusuf in [7]. Point Distribution Model (PDM) **is** techniques employed with total number of 66 points to fully describe the shape variations in accordance with five divided regions. The model worked robustly, even if more than half of the face is entirely covered. The proposed technique was evaluated using UMB database and obtained detection within shortest possible time.

Park *et al*. [8] presented a partially occluded facial image retrieval method based on a similarity measurement for forensic applications. The novelty of their method compared with other occluded face recognition algorithms is measuring the similarity based on Scale Invariant Feature Transform (SIFT) matching between normal gallery images and occluded probe images. The proposed method achieved retrieval rates of up to 94.07% using AR database. Similarly, Face recognition using Ensemble String Matching is proposed by [9]. Recognition rate of 96.58% is achieved using AR database. It demonstrates the predominance of the proposed method in perceiving part of the occluded faces and the practicality of utilizing a high level syntactic technique, utilized for face recognition for face coding and matching.

Shin *et al*. [10] proposed a Hybrid Approach for Facial Feature Detection and Tracking under Occlusion. Databases used are AR, LFPW, multi-PIE. Recognition rate of 96.4 % was obtained using AR database. The result demonstrates the smallest normal normalized error and standard deviation under strong occlusion.

Other recent related studies on the databases are [11], [12], [13] and [14] for AR database, [15] and [16] for Bosphorus database, [17] and [18] for LFW database and, [16] and [19] for UMB database.

The process of building the occlusion databases entails at least one of the following; (i) gathering raw images, (ii) applying a face detector and manually eliminating false positives, (iii) removing duplicate images, (iv) naming the detected people, (v) cropping and rescaling the images to the desired resolution. A short description of a few selected databases is described as below.

3.0 DESCRIPTION AND DATA ACQUISITION SYSTEM OF THE DATABASES

3.1 Aleix-Robert (AR) face Database

a. Database content

AR face database [20] was recorded at the Computer Vision Centre in Barcelona, Spain in the year 1998. The database contains image of 126 people (70 men and 56 women). Every person is recorded in two sessions which are recorded two weeks apart. Each subject has 13 different images, including 3 illumination variations, 4 facial expressions, 2 occlusions (wearing scarf and sunglasses) and 4 occlusions on top of illumination changes (wearing sunglasses and leftward illumination), resulting in a total of 4000 images. All images of the AR database were colored with a resolution of 768 × 576 pixels. Landmark annotations based on a 22-landmark scheme are available for some of the AR database images. The database is available at http://rvl1.ecn.purdue.edu/_aleix/aleix face DB.html

b. Data acquisition

All images were acquired using the same system. The illumination conditions, camera parameters, and the distance from the subject to the camera were carefully controlled during the whole acquisition process. The system is calibrated twice a day (in the morning and afternoon) to ensure that settings are identical across the subjects. During each session 13 conditions with varying occlusion, facial expressions, illumination and were captured. Figure 1 shows an example of each condition. Male images are stored as M-xx-yy.raw and Females as F-xx-yy.raw. 'xx' is a unique person identifier ranging from "00" to "70" for males and from "00" to "56" for females. 'yy' specifies the features of each image. Description of the system used is given below;

i. Pentium 133 MHz, 64 MB RAM and 2Gb HD
ii. Color Camera: SONY CCDs
iii. A 12 mm optics

Frame grabber: Matrox Meteor RGB

Figure 1 Sample of images obtained from AR database. The conditions are neutral, occlusion by scarf and sunglass

3.2 Bosphorus Database

a. Database content

Bosphorus 2D/3D face database is collected during the Enterface '07 workshop held in Bogazici University, Istanbul, Turkey [21]. The database consists of facial images of 81 subjects (51 men and 30 women) with 3396 facial scans. Some facial scans include occlusions like beard & mustache, hair, a hand or eyeglasses, and up to 35 facial expressions per subject. The subjects were taken with various extensive poses, expressions, and occlusion conditions. One-third of the subjects are professional actors/actresses and majority of the subjects are Caucasian aged between 25 and 35. Each scan has been manually labeled with 24 facial landmark points (such as the nose tip and inner eye corners) provided that they are visible in the given scan. The images are colored of high resolution (1600x1200 pixels) with "PNG" as the file format. This database can be used as a benchmark for 3D landmarking, facial expression analysis as well as for comparative study of 2D and 3D landmarking. Information on how to obtain the database can be found at http://bosphorus.ee.boun.edu.tr.

The database has two versions:

i. Bosphorus v.1: This version consists of 34 subjects with 4 occlusions, 13 poses, 10 expressions, and 4 neutral faces, thus resulting in a total of 31 scans per subject.

ii. Bosphorus v.2: This version has more expression variations than Bosphorus v.1. It contains 47 subjects with 34 expressions, 13 poses, 4 occlusions and 1 or 2 neutral/frontal faces, thus resulting in a total of 53 different face scans per subject. Each scan is intended to cover one pose and/or one expression type, and most of the subjects have only one neutral face, while few have two. It also incorporates 30 professional actors/actresses out of 47, which hopefully provide more realistic expression understanding.

Both Bosphorus v.1 and Bosphorus v.2 versions contain realistic occlusions of faces. For each subject, 4 occluded scans were gathered. These occlusions are i) mouth occlusion by hand, ii) eyeglasses, iii) occlusion of face with hair, and iv) occlusion of left eye and forehead regions by hands. Figure 2 shows sample images from the Bosphorus 3D database illustrating expression variations and typical occlusions.

b. Data acquisition

Facial data for each scan (the 2D texture image and the corresponding 3D point cloud) are acquired with Inspeck Mega Capturor II 3D, which is a commercial structured-light based 3D digitizer device. The sensor resolution in x, y and z dimensions are 0.3mm, 0.3mm and 0.4mm respectively. It is very fast, capable of capturing a face in less than a second. Subjects were positioned at a distance of about 1.5m away from the 3D digitizer. A 1000W halogen lamp is used in a dark room to obtain consistent lighting. Yet, specular reflections occur on the face due to the strong lighting of this lamp and the device's projector. This does not only affect the texture image of the face but can also cause noise in the 3D data. To prevent this, a special powder which does not change the skin color was applied to the subject's face. Also, several basic filtering operations like (Gaussian and Median filter) were applied. During acquisition, each subject wore a band to keep his/her hair above the forehead to prevent hair occlusion, and also to make face segmentation task simpler. Finally, each scan is down-sampled and saved in two separate files that store color photograph and 3D coordinates. Figure 2 shows sample of the images present in Bosphorus database with various occlusions.

Figure 2 Sample of images in Bosphorus face database with typical occlusion variations by hand and sunglass

3.3 University of Milano Bicocca (UMB) Database

a. Database content

UMB face database consists of a total of 1473 images with 143 subjects (98 male and 45 female), where a pair of male twins and a baby is included [19]. It is a collection of multimodal (3D + 2D color images) facial acquisitions. The database has been built to test algorithms and systems for 3D face analysis in uncontrolled scenarios, particularly in the cases where faces are occluded. Each subject has been acquired with different facial expressions (neutral, smiling, bored and hungry), and with the faces partially occluded by various objects (eyeglasses, hats, scarves, and hands). It also contains 883 non-occluded and 590 partially occluded images. During the acquisitions subjects were free to cover different parts of their faces; for this reason, occlusions do not hide the same parts of the face. This characteristic of the data makes it a reliable test of the robustness of face analysis algorithms. All images are in PPM file format with a resolution of 680 × 480 pixels. Masks representing the visual part of the face are available for the partially occluded pictures. The database can be obtained at http://www.ivl.disco.unimib.it/umbdb/request.html.

b. Data acquisition

The images were simultaneously captured using Minolta Vivid 900 depth laser scanner (with 25mm lens, in slow mode). They are annotated with seven feature points including corners of the eyes, of the tip of the nose and of the mouth, as shown in Figure 3. In the case of occluded acquisitions, only the corresponding non-occluded feature points are given. However, for each acquisition there is a label describing whether it is occluded or non-occluded, occluding object if present, facial expression and so forth. During the acquisitions, individual was asked to close their eyes for safety reasons. Faces have been acquired in various indoor locations with uncontrolled lighting conditions. The 2D/3D images are released in their original form samples without any further processing such as noise reduction or holes filling. Figure 4 show some sample of images obtained from UMB database.

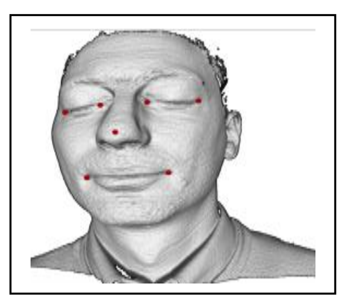


Figure 3 The seven annotated landmarks

Figure 4 Sample of images in Bosphorus face database with typical occlusion variations occlusion of left eye area with hand (top row), occlusion of mouth area with hand (second row), occlusion with eyeglasses (third row), and occlusion with hair (bottom row)

3.4 Labeled Faces in the Wild (LFW)

a. Database content

Labeled Face in the Wild (LFW) is a facial database consisting of 13,233 facial images with 5,749 distinct subjects downloaded from new articles on the web [22]. More than 29 % of the subjects contain more than one face image, but only faces that contain the central pixel of the image are considered as the defining face of the image. Faces other than the target face is ignored as background. Most images are in color, and few are in gray scale. The database contains labeled face photographs spanning the range of conditions typically encountered in everyday life. The database exhibits "natural" variability in factors such as occlusions, lighting, pose, race, accessories, and background. The sample of images obtained from LFW database is given in Figure 5. Manual landmark annotations are not provided for this database. We believe such a database will be a valuable tool in studying unconstrained face recognition.

b. Data acquisition

Initially, we used the raw images from the faces in the wild database. The OpenCV6 trained version of the Viola-Jones face detector [23] was run on each image. Subsequently, false positive face detections were manually removed, along with images for which the name of the individual could not be identified. The face region returned by the Viola-Jones detector contains only a subset of the whole head. So the region was automatically expanded by 2.2 in each dimension to capture the entire head. This expanded region was then cropped and rescaled to output a 250 x 250 JPEG image. The name of the person pictured in the center of the image is given. Each person (picture captured at the center) is given a unique name in such a way that no name corresponds to more than one person, and each individual appear once under one name (unless if there are unknown errors in the database).

Figure 5 Sample images from LFW database

4.0 COMPARISON OF THE DATABASES

From description of the databases in section 3, outstanding occlusion face databases categorized into either 2D or 3D faces databases and databases under controlled conditions or databases without any control condition.

a. 2 Dimensional or 3 Dimensional databases

In comparison with the large number of published 2D face databases, the number of available 3D face databases is relatively small. From the study, AR and LFW are 2D databases. Yale and PUT databases contain occluding images but, they are not as challenging as the corresponding 2D face databases considered in this paper. Others that were not explained or mentioned in the table include FERET, Yale B, MIT, ORL, IMM, and BioID.

The Bosphorus and UMB are 3D face databases. Others include FRGC, ND-2006, GavabDB, BJUT-3D, 3DTEC, UHDB11, and BU-4DFE. Almost all 3D face databases are also multi-modal face databases since 2D texture images are also provided. One of the commonly used variations for both 2D and 3D face databases is partial occlusion. Only a few 3D databases contain occluded faces. Modeling occlusion is fairly easier for 2D face databases than 3D face databases because images are at different poses in 3D which causes self-occlusion.

b. Controlled and Uncontrolled databases

Controlled databases are collected within the framework of a defined experimental setup using one or more of the four variations mentioned earlier. For example, CMU Multi-PIE database is suitable for evaluating different poses; MMI database for a variety of facial expressions. In this study, AR and Bosphorus are controlled databases. Other examples of controlled databases include XM2VTS, UHDB11, FERET, and FRGC.

On the other hand, uncontrolled databases, are collected without any directives given to the subjects. They are sometimes referred to as "faces in the wild".

Recently, databases collected from social network sites such as google.com, flickr.com, facebook.com have gained a lot of interest. They provide more realistic and challenging databases, and also due to the huge potential of web sources. Uncontrolled databases naturally contain complicated type of the variations seen on a daily basis. This includes change in camera quality, color saturation, pose, lighting, expression, age, gender, clothing, background, race, ethnicity hairstyles, focus, and other parameters [22]. UMB and LFW are uncontrolled databases considered in this study. Others include BioID, AFLW, AFW, 300-W, COFW databases.

After studying details of the databases, it can be concluded that the UMB is more challenging, since it contains large number of occluded faces, but only few works were reported in the literature using this database. In both AR and UMB databases, no restriction on wear (clothes, glasses, etc.), makeup and hair style were imposed on the subjects. LFW have large number of subjects, but the images are captured with low and limited to number of views. It is designed for face detection problem only as its image contained a high percentage of label errors and duplicated images, hence making it unsuitable for recognition.

The state of the art databases mentioned in this paper contain significant characteristics that are useful in this field but still challenging in terms of reflecting real world applications. Although UMB constitutes large number of subject in terms of occlusion, it may be seen as great pluses over the other databases and achieved great results [2], [24]. But, it contain limited amount of images with expressions unlike in Bosphorus database which contains occluded faces too and has large number of expressions and poses. Bosphorus database contain images with high resolution and can be used as a benchmark for 3D landmarking, facial expression analysis as well as for comparative study of 2D and 3D landmarking [25], [26]. UMB in particular, is suitable to designing and testing algorithms dealing with partially occluded faces. The main features of the aforementioned data sets are summarized in Table 1.

Table 1 An overview of the selected databases for occlusion analysis

Name	Number of		Variability	Resolution	Modality
	images	subjects			
AR [20]	4,000	126	ex, il, oc	768 × 576	Color image
Bosphorus [21]	4,666	105	ex, oc, po	1600x1200	Color, 3D data
UMB [19]	1,473	143	ex, oc, po	640 × 480	Color, 3D data
LFW [22]	13,233	5,749	ex, il, oc, po	250 × 250	Color image

Acronyms on variability column are as follows: ex: facial expressions, il: illumination changes, oc: occlusion, po: pose variations.

The table gives a clear and thorough description of the databases in one representation. The parameters include; the database name and year of publication, total number of images and subjects present in each database, recording conditions (such expression, illumination, pose and occlusion) on the face, the image sizes and the modality on which the images operates in that order.

5.0 CONCLUSION

We described the key features of the selected databases such as the total number of images and subjects, recording conditions and procedures, image resolution, availability, file format and image samples. The comparison of the selected and other related databases was discussed. Some notable related works conducted on the databases were referenced from the table. The scope of the paper is limited to occlusion databases though some other parameters were taken into consideration for the sake of comparison. The comparative study of these occluded databases will give a room to researchers to debate on some questions like; (i) Is 3D better than 2D when unknown objects partially occlude faces? (ii) Do occlusions introduce some degree of uncertainty in the localization of facial features? (iii) How is the uncertainty influences face recognition performances? (iv) Which types of databases are good for occlusion analysis? and (v) to discover if multimodal approaches are better than single modality approaches. Lastly, there is need for globally accepted and error free dataset to compare the robustness of algorithms.

Acknowledgement

This study is academically supported by Universiti Sultan Zainal Abidin (UniSZA), Kuala Terengganu, Malaysia, and financially supported by Kano State Government Nigeria and National Biotechnology Development Agency (NABDA), Abuja, Nigeria.

References

[1] P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. 2005. Overview of the Face Recognition Grand Challenge. *Computer Vision and Pattern Recognition IEEE Computer Society Conference.* San Diego, CA, USA. 20-25 June 2005. 947-954.

[2] Colombo, A., C. Cusano, and R. Schettini. *2011.* UMB-DB: A Database of Partially Occluded 3D Faces. IEEE *International Conference on Computer Vision Workshops*. Barcelona, Spain. 2 6-13 November 2011. 113-2119.

[3] R. Gross. 2005: Face Databases. *Handbook of Face Recognition.* New York: Springer. ISBN 0-387-40595-x.

[4] A.M. Martinez and R. Benavente. 1998. The AR Face Database. CVC Technical Report. 24.

[5] A. Savran, N. Aly¨uz, H. Dibeklio˘glu, O. C, eliktutan, B. G¨okberk, B. Sankur, and L. Akarun. 2008. *Biometrics and Identity Management: First European Workshop, BIOID.* Springer Berlin Heidelberg*.* 5372: 47-56.

[6] G. Huang, M Ramesh, T Berg, E Learned-Miller. 2007. Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments. Tech. rep. Technical report, University of Massachusetts, Amherst.

[7] Viola, P., and Jones, M. J. 2004. Robust real-time face detection. *International Journal of Computer Vision.* 57(2): 137-154.

[8] A. Colombo, C. Cusano, and R. Schettini. 2011. Three-Dimensional Occlusion Detection and Restoration of Partially Occluded Faces. *Journal of Mathematical Imaging and Vision*. 40(1): 105-119.

[9] Colombo, C., C. Cusano, and R. Schettini. 2007. A 2D+3D Robust Face Recognition System. *14th International Conference on Image Analysis and Processing.* Modena, Italy. 10-14 September 2007. 393-398.

[10] A. Savran, B. Sankur, M. T. Bilge. 2012. Comparative Evaluation of 3D versus 2D Modality for Automatic Detection of Facial Action Units**.** *Pattern Recognition.* 45(2): 767-782.

[11] Çeliktutan, O., H. Çınar, B. Sankur. 2008. Automatic Facial Feature Extraction Robust Against Facial Expressions and Pose Variations**.** *IEEE International Conference on Automatic Face and Gesture Recognition, Amsterdam, Holland. 17-19 September 2008.*

[12] Milborrow, S. and F Nicolls. 2008. Locating Facial Features with an Extended Active Shape Model*. European Conference on Computer Vision.* Marseille, France. 12-18 October 2008. 504–513.

[13] Urclar, M. and Franc, V. 2015. Open-source implementation of facial landmark detector [online]. From: http://cmp.felk.cvut.cz/~uricamic/flandmark/. [Accessed on 13 January 2015].