Face Recognition under Occlusion for User authentication and Invigilation in Remotely distributed on-line assessments

Niloofar Tvakolian Department of Computer Science & Engineering and Information Technology, Shiraz University, Iran, E-mail: niloofar-ta@shirazu.ac.ir
Azadeh Nazemi School of Electrical Engineering, Computing and Mathematical Sciences, Curtin University, WesternAustralia, E-mail: azadeh.nazemi@curtin.edu.au
ZohrehAzimifar Department of Computer Science & Engineering and Information Technology, Shiraz University, Iran, E-mail: azimifar@cse.shirazu.ac.ir
Iain Murray School of Electrical Engineering, Computing and Mathematical Sciences, Curtin University, WesternAustralia, E-mail: i.murray@curtin.edu.au

Abstract: Generally, running assessment requires many resources. Remote assessments solve local assessments issues and eliminate some limitations. However, running the remote assessment service strictly security dependent. Multi-Factor Electronic Authentication (MFA) by inserting a protection layer on top of username and password addresses the monitoring and security issues of remote assessment. On the result, users require passing two security levels to access online assessment services. They confirm what they know using username/password and verify what they have using biometric factors. This research proposes a method to manage aspects of distance digital learning in the assessment area of the learning organisation. This study has focused on the face as a second biometric factor. Since Face recognition systems have been affected by variations of conditions such as expression, pose, occlusion, and illumination. The most highlighted challenge of this research was finding a solution to address uncontrolled conditions in face recognition. Obtained results of this project indicate reasonable accuracy to address the issue of occlusion using AR, MUCT and UMB-DB Datasets.

Keywords: Face Recognition, Machine Learning, Occlusion, Illumination Multi-Factor Authentication, Remote Assessment, Identification, Verification, Expression, Face Detection, Deep Learning

1. Introduction

Undertaking assessments can be a daunting process for users. Being in an assessment location in exact time for users who live a long distance from the assessment location or have a mobility condition can be frustrating. Running online assessments lead to some subsidy benefits such as the reduction of air pollution, public transport usage, and parking vacancies demands at the assessment location. Meanwhile, maintenance costs, repairs, and depreciation of equipment must be noted. In addition, supervisors, observer/assessors must be hired to conduct assessments and prevent fraud and cheating (Grant and Stewart, 2012).

On the other hand, the assessment service requires time allocation and severely depends on supervisor availability. It causes limitations when there is the time difference between assessment supervisors and users.

Finally, to access the assessment service, the remote candidate needs to authenticate. The username identifies the user, while the password authenticates that the user is whom he/she. Logging on username and password does not prove which logged on user and the remote candidate is same. In the online assessment, the system must identify users through something they know such as username and password and something that they are such as bio-metrical. A user must present a proof of presence to avoid cheating and fraud otherwise online remote assessments need restricted supervision for this purpose. An accurate multi-factor biometric authentication system provides the opportunity to run secure remote assessments without hiring observers. This system increases the layer of security to electronic authentication. It means users can only start remote assessments after they are verified through electronic authentication to decrease the possibility of cheating. The used Biometric factor in this project is the face. The main drawback of using face biometric for identification is that it can never be 100% accurate. Two statistics are used to measure system accuracy.

False Non-Match Rate (FNMR) that means how often a biometric is not matched to the template when it should be. False Match Rate (FMR) that means how often a false biometric is matched to the template when it should not be. (Jain et al., 2004) (Alyuz et al., 2008)

The rest of the paper was organized to respectively present many concepts in this area, prior requirements and related work in the next section then section three includes two implemented methods for online face recognition based on feature extracted and deep descriptors. Also experimental results and evaluation of each method brought in this section. Experimental result and evaluation is available in section four and finally section five contains further developments and conclusions.

2. Related Work

2.1 Authentication:

Authentication for remote assessments includes four steps:

2.1.1 Registration/enrolment

In the biometric system, Pre-authentication in client side refers to user enrolment by creating an account and providing some essential personal information and a standard facial photo. This photo is used to make training dataset and must be current, valid, and authentic.

2.1.2 Confirmation

Confirmation in server side means acceptance of the user logged on information comparing to registered information. Confirmation is applied if the logged on student information is matched to entered username and password and registration information saved on server sides as Figure 1 illustrates.

Figure 1: Normal Login

2.1.3 Identification

Identification in server-side means the image of the present user who is already logged in is matched with the image of the claimed person. Assessment will be stared as server identify student. The server sends an acknowledgment to the client machine.

2.1.4. Verification

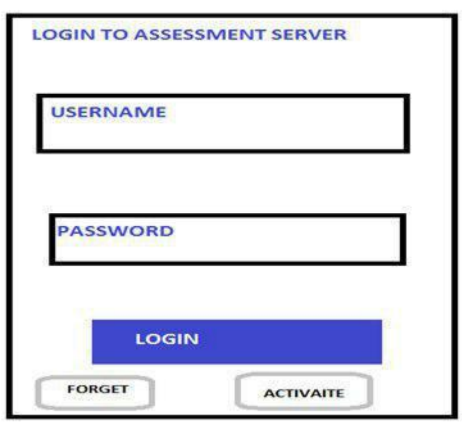
Post authentication means that users can access a remote online secured system as soon as is identified through their username and password (E-authentication). The access to the system will be continued as long as successful verification results obtained by face recognition indicates that the logged in user is current, enrolled and constant during the access. Since making restriction for user during verification reduce system functionality in remote assessment, making obligation for them to submit more than one normal frontal facial photo during registration or enrolment in most cases is not practical and ethics, then designing an accurate face recognition system under uncontrolled conditions such as illumination, occlusion, position, and expressions with a small size dataset is required. In small size face recognition system, the dataset includes just one sample of each class.

2.2. Requirements for two sides of Remote Assessment Service

Remote Assessment Service generally includes two sides:

2.2.1. Server side

Assessment side is equipped with a powerful face recognition application, the database of user's information and registration profiles. MFA addresses the dependency issues of remote assessments to the presence of the supervisor. However, the assessment side still requires a booking system to avoid conflict. This booking systems consider priorities and making appropriate time slots which leads to making some restrictions for users located in the different



2.2.2. Client side

The client side has an internet connection, activated webcam and installed the automatic developed application by this research. It is responsible for checking the availability of a webcam on the users' machine, capturing a frontal photo, detecting landmark of the user's face, cropping the face boundary box and finally uploading the aligned facial image to the web server and communicate with the assessment center. Users can attend a remote online exam as soon as they are identified through their username and password (E-authentication) and verified through face recognition. The exam is running as long as obtained results by face recognition indicate that the logged in user is current, enrolled and constant during the assessment. If face recognition results do not approve successful verification, then the assessment will be suspended. Therefore, remote online assessment requires frequent verification. It is implemented by taking a facial photo automatically and sending to the assessment center remotely in a random time interval. Figure 2 illustrates the essential requirements for both sides (client and assessment center) to communicate. Figure 3 is an overview block diagram of this authentication and invigilation system.

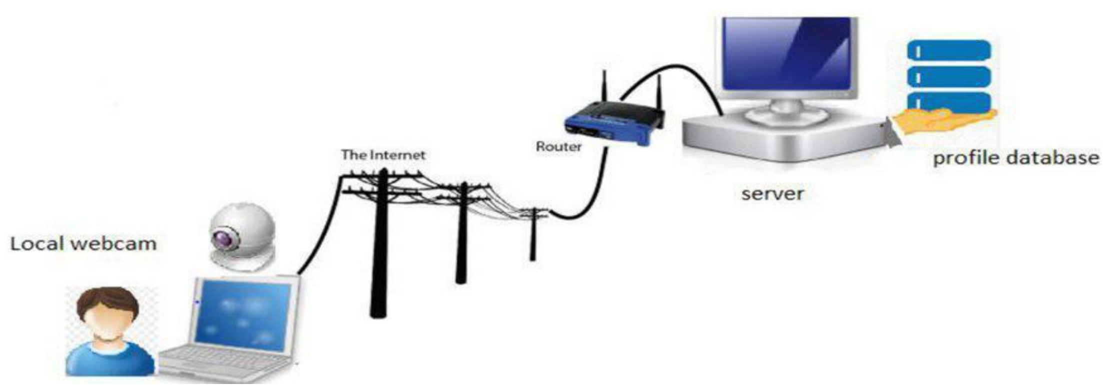


Figure 2: Server Client Connection

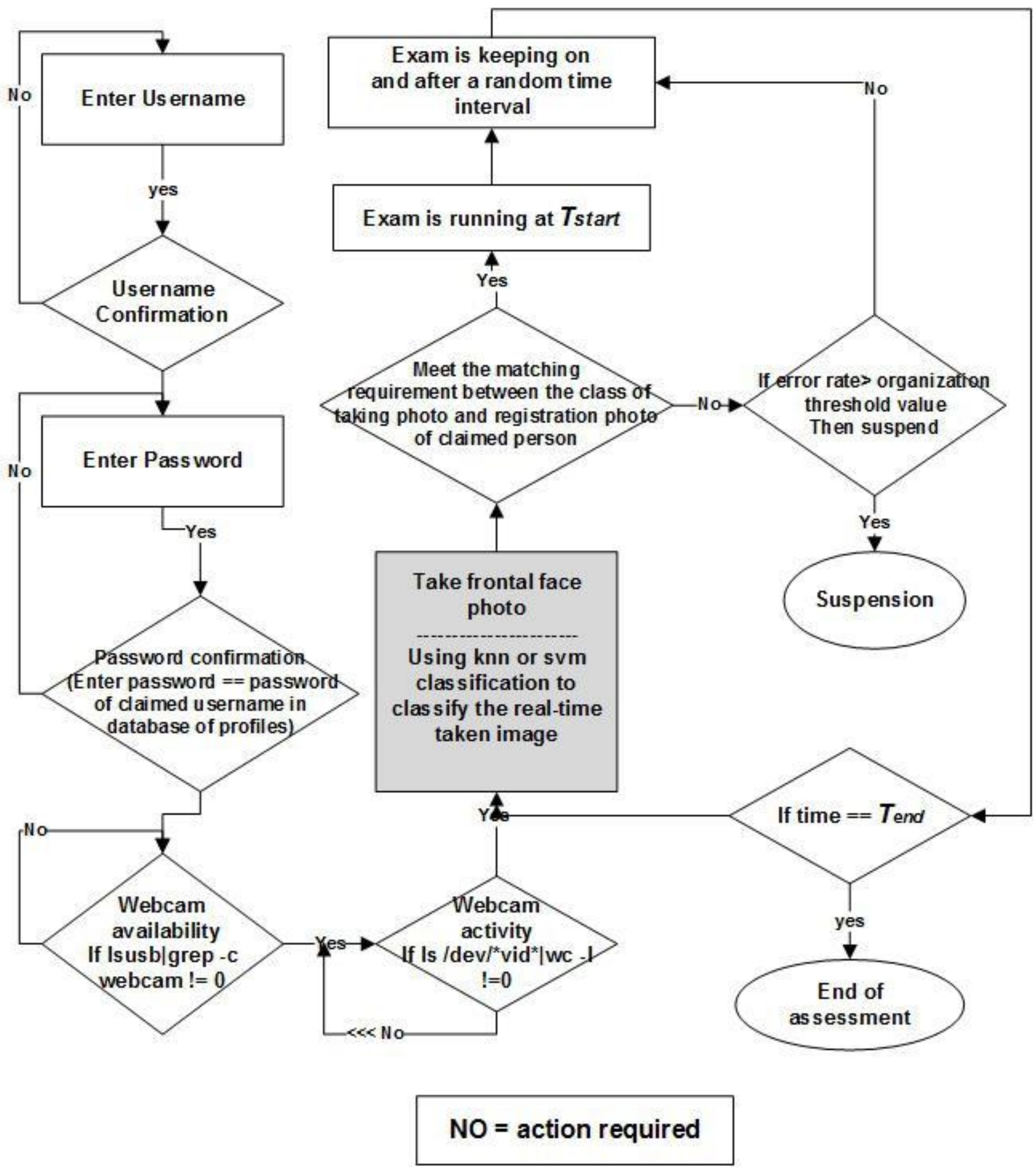


Figure 3. Online assessment Authentication and Invigilation

Face recognition module is the most challenging part and the main kernel of this project. In order to use the face as a biometrical factor in MFA, real-time face recognition is required. Due to the nature of real-time applications, it faces variations of conditions such as expression, pose, occlusion, and illumination. Traditional approaches to address these conditions are Model-based (Du and Ward, 2006) and pre-processing-based approaches. Model-based methods model all possible conditions. Pre-processing-based approaches using image processing technique try to remove unwanted conditions. The features such as Histogram of Oriented Gradient (HOG)(Sahoolizadeh and Ghassabeh, 2008), Scale-Invariant Feature Transform(SIFT)(Geng and Jiang, 2009), Eigen(Turk and Pentland, 1991), FISHER(Sahoolizadeh and Ghassabeh, 2008), Gabor(Abhishree et al., 2015), and Laplacian are the basis of the appearance-based approaches. However, using these purely appearance-based methods for face recognition under uncontrolled conditions will not be successful. According to (reference) the appropriate feature for face recognition under uncontrolled condition must be independent of face component/face texture. Based on this article the features can be categorized into two divisions: 1. dependent or occlusion sensitive features (Eigen/Fisher) 2. Independent or occlusion insensitive features (HoG/Gabor). Face recognition module is the most challenging part and the main kernel of this project Face recognition result strictly depends on face detection accuracy. On the other hand, Biometric recognition (face recognition) must be carried out in two modes: a) Identification and b) Verification. This research as a machine learning project contains several sub-modules such as data mining, training, and testing. Training and data mining (registration and enrolment) are the same process in both modes (identification and verification) however testing is different. In both modes, registration or enrolment means collecting training samples from each candidate and storing in a reference database for training. In the identification mode during the testing process, a candidate sample will be compared against all enrolled samples to find the best match to identity this sample. In the verification mode, the claim of the candidate is only compared to enrolment samples of the claimed identity to decide whether it meets minimum requirements for matching or not. Another difference between identification and verification is about restriction for conditions of capturing the image. Upon identification, due to user awareness, they may be asked to follow some instructions such as sitting in the front position, utilizing a uniform lighting condition with uncovered main face components and normal expression. Since verification performs at random time intervals during assessments providing predefined conditions may disturb the user.

2.3. Face Recognition Modules

Face recognition regardless of being offline or online includes some modules such as data mining, data augmentation or increasing the size of the dataset by applying the random block to make synthetic occlusion in order to address the lack of a sufficient number of samples, data preparations by re-sizing or scaling, Face detection and alignment by cropping face boundary box (it was supposed that all samples in this research were detected, cropped and aligned properly, therefore, this research does not describe the methods challenges and approaches behind the face detection),eliminate illumination impact applying Equalization/ Normalization, extraction to making a dictionary, key point extraction such as SURF, SIFT, ORB to make Descriptor and finally classification using k Nearest Neighbor(kNN)/Support Vector Machine(SVM)(Jia and Martinez, 2008)

As stated before, this research supposed that in all samples, faces are detected and aligned with specific size depends on the used method for face recognition (feature-based or deep learning). Deep learning methods firmly require much more samples compare to feature-based ones. Data augmentation is the way to generate the different varieties of each exist sample by rotation, shifting and other image processing techniques. Besides data modelling is a method to generate training data under specific conditions such as illumination or occlusion. In this method using some image processing technique normal frontal face sample is converted to an affected sample by a specific condition. The random block is a method to replace the intensity of monkey face block in the random location of the original face sample. This method assists to implement the real-time condition of the occluded face on the standard face dataset as YALE-B. Since the final target of this project is MFA based on face recognition and online systems suffers from illumination and occlusion, then addressing probable illumination and occlusion are the most dominant part of the project. According to the latest research, there are several approaches to solve the issue regarding illumination. The following methods were proposed to decrease illumination effects by preprocessing: Histogram Equalization (HQ)(Fierrez et al., 2009), Histogram Truncation and Stretching (HT), Extreme Value Distribution(EV), Exponential Distribution (EN), Normal Distribution (ND) Gaussian (Uniform), Lognormal Distribution LN), Single Scale Retinex (SSR)(Choi et al., 2007), Mutli Scale Retinex (MSR), Multi Scale Self Quotient Image (MSQI)(Chen et al., 2013, Homomorphic Filtering (HOMO), Single Scale Self Quotient Image(SSQ), daptive Single Scale Retinex (ASR), Wavelet-Based Normalization WAV), DCT-Based Normalization (DCT)(Azam et al., 2010), Illumination depends on two main concepts. Supposing the illumination changes slowly across different locations of the image and the local reflectance tolerates quickly across different location. Therefore, illumination can be drastically reduced using the high-pass filtering while reflectance still is very close to the original reflectance. Histogram equalization can be used as high pass filters. However, the reconstructing efficiency of normalization is more than equalization as Figure 4 illustrates.

RGB
Grayscale similarity=0.51
Histogram Equalization similarity=0.53
Normalization similarity=0.57

Figure 4. The similarity comparison of Ggrayscale, Histogram Equalization and Normalization.

2.4. The used Datasets

2.4.1. AR Dataset

AR database was created by Alex Martinez and Robert Benoventi (AR) (Martinez, 1998). Alex Martinez and Robert Benoventi database have created AR dataset. It is free for academic use and includes over 4,000 color images of 70 men and 56 women in different facial expressions, illumination conditions and occlusions such as sunglasses and scarf. This frontal view facial in two sessions and they represent as the followings: 1: Neutral expression 2: Smile 3: Anger 4: Scream 5: left light on 6: right light on 7: all side lights on 8: wearing sun glasses 9: wearing sun glasses and left light on 10: wearing sun glasses and right light on 11: wearing scarf 12: wearing scarf and left light on 13: wearing scarf and right light on| 14 to 26: second session (same conditions as 1 to 13)

2.4.2. UMB-DB Dataset

UMB-DB dataset (Colombo et al., 2011) contains 98 male and 45 female classes in 883 normal and 590 occluded frontal photos. Each class may contain occluded, black skin, hair, twin, female, smiling, misc_expression, angry, bored, hat, open mouth, free, hand, phone, neutral, misc_object, white skin, bottle, disgusted, male, scarf, and glasses images focusing on the following regions: Right_Eye_Outside_Corner, Right_Eye_Inside_Corner, Left_Eye_Inside_Corner, Left_Eye_Outside_Corner, Nose_Tip, Mouth_Right_Corner, Mouth_Left_Corner,

2.4.3. Muct Dataset

The MUCT database consists of 3755 facial photos of 625 classes with 76 manual landmarks. (Milborrow et al., 2010) It demonstrates the diversity of lighting, age, and ethnicity.

2.4.4. Shiraz Dataset

Shiraz data set includes 93 classes of staff and student Shiraz University. 750 low resolution facial images were taken to represent the impact of positioning, illumination, uncontrolled conditions in capturing procedure

3. Methodology Online Face Recognition (Identification)

This section proposes two methods to address issues regarding uncontrolled conditions in face recognition. The first method extracts affecting factor from the test sample utilizing mask projection. Current methods remove the occlusion from the test sample and reconstruct it. Unlike these methods, the first proposed method tries to add extracted occlusion to all normal training samples and compares the test sample with all synthetic affected training samples. The method has been applied for multi-factor authentication/verification based on face biometric. Obtained results indicate high accuracy, comparable to the best sparse method, in the lake of sufficient training samples for each class (single sample classes).

The second method tries to improve occluded face recognition based on a deep neural network (ResNet Based Face Identification). For extracting most accurate and discriminative descriptors uses modified ResNet 109.

Deep learning approach in this study has been performed after running a preprocessing module based on the experimental result obtained from the first method. The final target of this project leads to design a face network based on the nature of mask projection, it supposes that all input samples to the system are occluded. So, in preprocessing module initially the input sample as S_t compare to the enrolled/registration (S_r)sample using function XOR and the result will be added to the S_r.

$$S_r xor\ S_t = Occlusion$$
$$Occlusion \,\&\, S_r = Occluded\ synthetic\ sample\ (OSS)$$

Then obtained descriptors of S_t and OSS will be compared.

$$\Delta = \overrightarrow{S_t} - \overrightarrow{OSS}$$

Figure 5, illustrates the above equations.

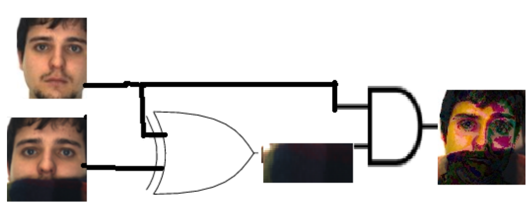


Figure 5 .Occulsuon extraction from claimed samplae, add it to enrolled sample

In order to predict the class of input sample the preprocessing module must be performed. Since the main target of online face recognition in this specific project is Identification, thus instead of comparing the test or claimed sample with all trained samples, the similarity of this test sample and its enrolled sample in training dataset must be calculated.

3.1. Mask Projection, HoG/Gabor Features, and kNN classifier

Sparse family classification methods claim high accurate and robust face recognition under occlusion. This family includes Sparse Representation-based Classification (SRC)(Wright et al., 2009) methods claim for highly accurate and robust face recognition under occlusion. Gabor feature-based sparse representation for face recognition (GSRC)(Yang and Zhang, 2010) , Extended SRC(ESRC)(Deng et al., 2012), Structured SRC (SSRC) (Ou et al., 2014) and Structured Occlusion Coding (SOC) (Wen et al., 2016). All stated sparse methods: Assume that d_{ij} is a sample of training dataset. $i = class\ label\ of\ training\ sample\ dataset$ and $j = sample\ label\ in\ each\ class$.

$$d_{ij} \in D_T$$

The SRC family utilizes D_T as a training sample dictionary generated by reshaping and concatenating all samples. Additionally SRC family uses another dictionary (occlusion dictionary) named O obtained from offline mode. Non-occluded testing sample define as

$$y_{input} = D_T\alpha + \varepsilon \qquad (2)$$

For occluded testing sample:

$$y_{input} = D_T\alpha + O\beta + \varepsilon \qquad (3)$$

error: $\varepsilon \approx 0$

Obtained results by SRC methods confirmed SOC is the most accurate one. Accuracy in SRC family methods is relevant to the features dimension.

Since in verification process type of occlusion is unknown so predicting occlusion dictionary is not possible. Therefore, this research assumes all input test samples are occluded and need to extract occlusion mask from them. In verification mode Red, Green and Blue (RGB) of the aligned detected face image is converted to Hue, Saturation, Value (HSV).

The procedure of occlusion mask extraction starts by extracting HSV mask from R (received sample) then obtaining an RGB mask from HSV mask and providing YCRCB. Skin mask can be obtained by finding the average of three applied mask includes HSV, RGB, and YCRCB on the captured facial photo.

Finally, the distance between histogram of oriented gradients(HoG) features of input test sample and each synthetic occluded training data is calculated and shortest distance represent the class of input sample this procedure is explained in detail here:

Let T_i Show trained front fascia normal image and $i = 1. \dots. l$ where l Is number of classes in training dataset. in offline mode mask extract or function is applied to all training samples and the results are represented as $M_1. M_2. \dots. M_l$ .For all training samples there are two distance parameters, basic distance(BD) and related distance(RD) that play important role to find actual distance between R and T_i . DB is distance between a sample and it's mask and RD is the distance between a sample and occlusion mask of R. $i = 1. \dots . l$; $BD_{T_i} = |T_i - M_i|$. Assuming M_R is occlusion mask of R then $RD_{T_{iR}} = |T_i - M_R|$. $BD_R = |R - M_R| and\ RD_{RT_i} = |R - M_i|$

$$BD_{T_i}||RD_{T_{iR}} = \sum_{n=0}^{[\log_2(x)]} \left[2^n \left[\left[\left(\left\lfloor \frac{BD_{T_i}}{2^n} \right\rfloor mod\ 2 + \left\lfloor \frac{RD_{T_{iR}}}{2^n} \right\rfloor mod\ 2 \right) + \left(\left\lfloor \frac{BD_{T_i}}{2^n} \right\rfloor mod\ 2 \right) + \left\lfloor \frac{RD_{T_{iR}}}{2^n} \right\rfloor mod\ 2 \ \right] mod\ 2 \right] \right] \quad (7)$$

$$BD_R||RD_{RT_i} = \sum_{n=0}^{[\log_2(x)]} \left[2^n \left[\left[\left(\left\lfloor \frac{BD_R}{2^n} \right\rfloor mod\ 2 + \left\lfloor \frac{RD_{RT_i}}{2^n} \right\rfloor mod\ 2 \right) + \left(\left\lfloor \frac{BD_R}{2^n} \right\rfloor mod\ 2 \right) + \left\lfloor \frac{RD_{RT_i}}{2^n} \right\rfloor mod\ 2 \ \right] mod\ 2 \right] \right] \quad (8)$$

Equation (7), (8) respectively generate synthetic occluded training samples and received testing sample affected by mask of training samples, which are namely called synthetic Occluded Training (SOT) and Testing Mask Affected (TMA) in this research. HOG features are extracted from two collections of images $\{SOT_1. \dots . SOT_l\} and\ \{TMA_1. \dots . TMA_l$. Eventually, Euclidean distance is calculated based on 1×64 HoG feature vector. $Di = EuclideanDistance(SOT_i;\ TMA_i)$;

If D_i is minimum value in $D_1. \dots . D_l\ and\ 1 <= i <= l$ then i indicates the class of testing sample. Figure 6 demonstrates the stated method visually. Figure 6-A form right to the left demonstrates a training sample, the mask of the sample and subtraction result. Figure 6-B from right to left shows the training sample, the mask of the testing sample and subtraction result. Figure 6-C from right to left illustrates an occluded testing sample, the mask of the testing sample and subtraction results. Figure 6.D represents the related distance between a received testing sample and the mask an individual training sample mask, Figure 6.E is bit-wise OR between A and B, and Figure 6.F is bit-wise OR between C and D. As it is observed from Figure 6 the results of 6-E and 6-F are quietly similar.

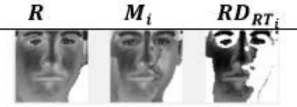
Order	Image	Formula
A	 T_i M_i BD_{Ti}	$BD_{T_i} = \|T_i - M_i\|$
B	 T_i M_R RD_{T_iR}	$RD_{T_{iR}} = \|T_i - M_R\|$
C	 R M_R BD_R	$BD_R = \|R - M_R\|$
D	 R M_i RD_{RT_i}	$RD_{T_R} = \|R - M_i\|$
E	 BD_{Ti} RD_{T_iR} SOT_i	Equation (5)
F	 BD_R RD_{RT_i} TMA_i	Equation (6)

Figure 6. A visual description of the steps in the proposed method

3.2. Experimental results of shallow method

The HOG recognition rate for illumination is 93%, for sunglasses, 90%, for a scarf, 85%. The HOG accuracy for the faces covered by scarf under illumination condition is 81%. The Gabor recognition rate for illumination is 98%, for sunglasses, 100%, for a scarf, 81%. The Gabor accuracy for the faces covered by scarf under illumination condition is 92% Obtained results prove Gabor is more accurate than HOG in recognition. as Figure 7 presents.

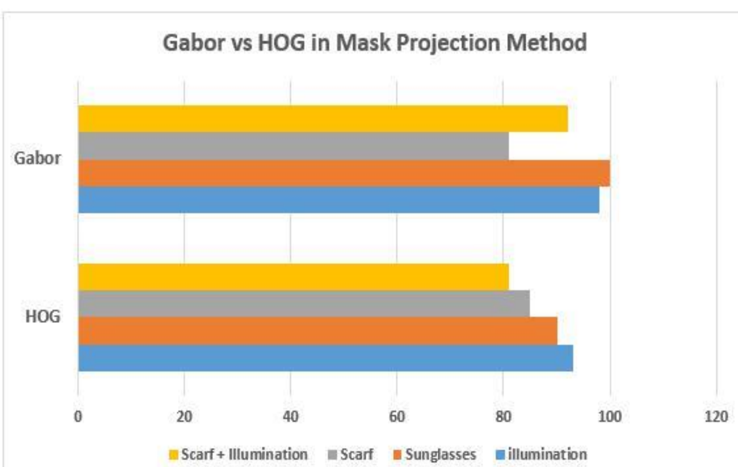


Figure 7: Comparison Recognition rate of the proposed method based on HOG and Gabor on the AR database. Top: HOG feature extraction, Bottom: Gabor feature extraction

There is an inverse relationship between accuracy and occlusion percentage in face recognition. The experimental results on AR dataset confirm, illumination approximately causes 10% occlusion. While sunglasses occlude about 25% of face and a scarf covers at least 40% of face.
As stated before, this study has considered all uncontrolled conditions as occlusion. Figure 8 compares the relation between occlusion percentage and accuracy in previous methods. Figure 8-top illustrates the obtained results of recognition accuracy calculation of Mask Projection(MP), HOG features extraction and Nearest Neighborhood(NN). (MP+HOG+NN). Figure 8-bottom shows the results of Mask Projection(MP), GABOR features extraction and Nearest ENeighborhood(NN) (MP+GABOR+NN) .

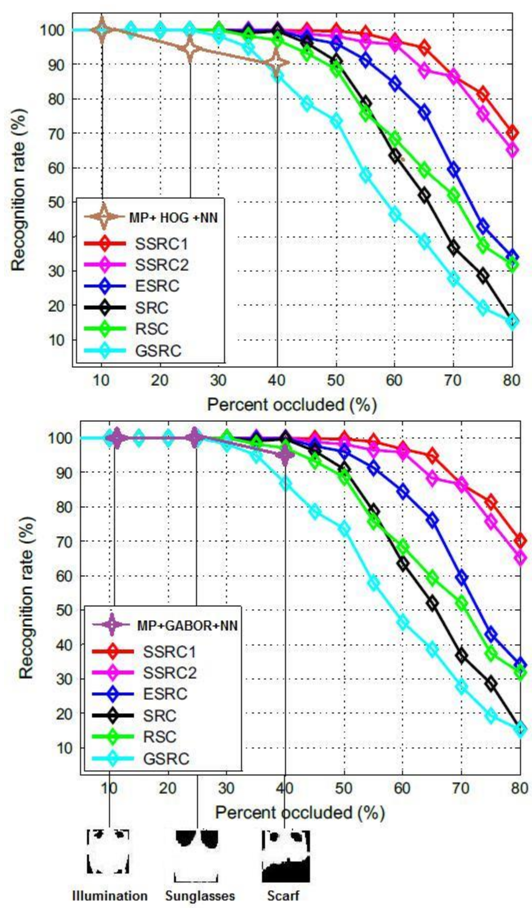


Figure 8: Comparison Recognition rate of the proposed method and six SRC family Methods on AR database. Top: HOG feature extraction, Bottom: Gabor feature extraction

3.3. ResNet Based Face Recognition

Nowadays, deep learning becomes a prevalent method for any applications such as DeepFace and Open Face for face recognition (Parkhi et al., 2015) (Amos et al., 2016). Deep Face accuracy is 97% on Labeled Faces in the Wild(LFW) and 91% on YouTube dataset. OpenFace accuracy for

LFW is approximately 100%. Deep neural networks have high accuracy in classification application and specifically face recognition. In training these networks need large scale datasets to learn the network's weights. For example, the FaceNet as a deep network includes 22 layers and 120 million weights that trained with 100-200 million face images from YouTube. This dataset is not accessible for public. Since data mining in the large-scale and labelling them is a cost consuming process and using a pre-training without weights modification leads to overfitting that consequently gaining low accuracy. So the recent approach to address the issue regarding the lack of sufficient trained samples is transfer learning. (Weiss, khoshgoftar and Wang, 2016) In this approach instead of training the network with a small limited dataset and using the random weight, a pre-trained network learned by million samples is offered, then based on the final target of the network the last layers are designed. Based on the target of this research, just used a normal dataset with a limited number of images. This study uses AR dataset including occluded faces with 4000 images that are very small amount comparing to the datasets which are used in the Deep Neural Network. In this stage, ResNet is utilized as a framework. The reasons behind choosing ResNet as a frame network besides, high accuracy is Identity Skip Connection layers that prevent vanishing gradient in this network. In deep neural networks vanishing gradient happens when the number of layers is increased. Vanishing gradient causes the quality of learning to get reduced. However, ResNet can solve this issue by inserting Identity Skip Connection (He, Zhang, Ren and Sun , 2016).

In spite of a large number of layers in ResNet, it is light weighted and also generates discriminative descriptor. Figure 9-top indicates the failure in face recognition using image pixels' intensity. Figure 9-bottom denotes high accuracy of ResNet descriptor.

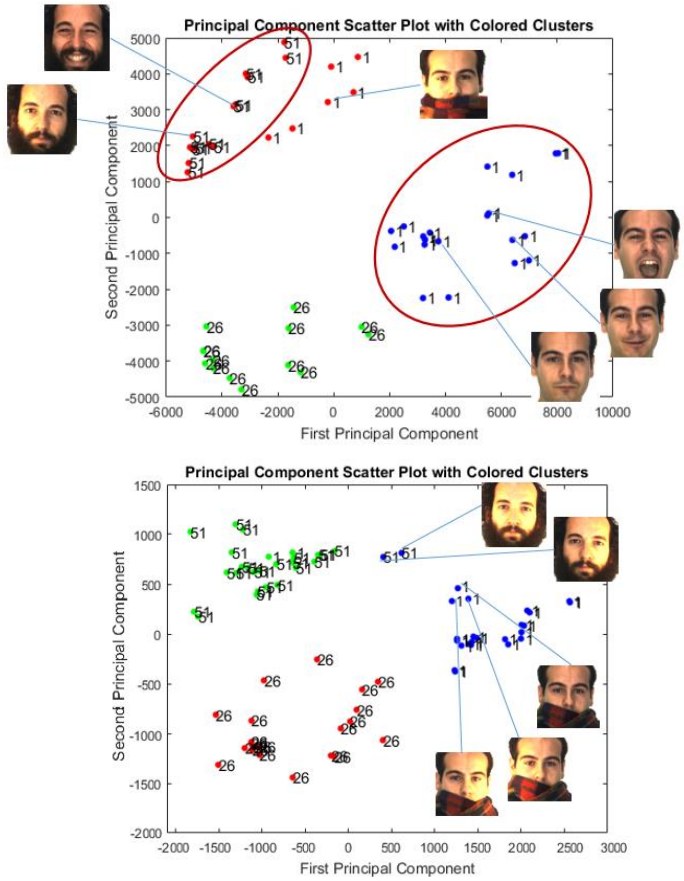


Figure 9. Top: face recognition based on intensity as a feature vector. Bottom: face recognition base on ResNet descriptor.

This research uses a pre-trained Residual Network(ResNet) (Gruber et al., 2017) and classifies 128 embed features descriptors to address the issue regarding face recognition under uncontrolled conditions. ResNet initially described by He et al in 2015 (Xie et al., 2016) to address image Recognition (Gruber et al., 2017). ResNet architecture offers extreme deep network trained by Standard Stochastic Gradient Descent (SGD). SGD is one of the most used optimization techniques in an iterative fashion. ResNet Convolution applies a set of convolutional filters. Equation 16 denotes how to calculate the output size of the convolution layer when

The first Convolution2D layer has 64 feature maps with the size of 7 7 as 64@7 7 and the size of the input image is 224 224 3 then output = 64@112 112 when padding=1 stride=2

Rectified linear unit (ReLU) allocates negative values to zero and maintains positive values. Batch Normalization accelerates training phase by Reducing Internal Covariate Shift (Ioffe and Szegedy, 2015) then MaxPooling2D pooling layer is configured with a pool size of 3 3 to take the maximum value. Pooling simplifies the output by performing nonlinear down sampling, reducing the number of parameters that the network needs to learn about. Pool 3 3 then out will be 64 112 112 These three operations are repeated over the layers, with each layer learning to detect different features. The first residual block starts by The convolution layer has 64 feature maps with the size of 1x1 (stride=2, padding=1) and a rectifier activation function output will be 64 56 56

Analyzing propagation formulation approves the forward and backward signals can be directly propagated from one block to any other block when using identity mappings as the skip connections and after-addition activation. Identity mapping makes training easier and improves generalization (He et al., 2016) After three layers identity mapping convolution layer 256 feature maps with kernelSize=1x1 stride=2, padding=1) and a rectifier activation function generates output 256 28 28 other residual blocks repeat as in the same way by features map (128,512),(256,1024) and (512,2048) and output size (28,14,7)

Next is a layer that converts the 2D matrix data to a vector called Flatten. It allows the output to be processed by standard fully connected layers. Next, a fully connected layer with 128 neurons and rectifier activation function are used. Figure 10 illustrates the original ResNet layers.

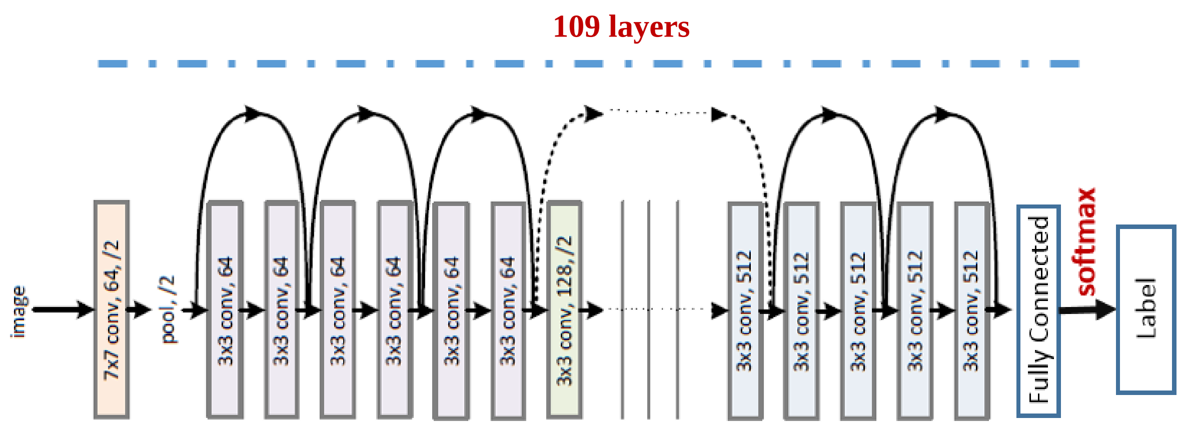


Figure 10: An overview of original ResNet

To achieve more accurate result this section relies on output descriptor of modified ResNet.

The designed network based on ResNet in this paper initially eliminate the last layer then insert a fully connected layer followed by sigmoid loss function. As figure 108 illustrates the first 108 layers have been frozen and only the two last layers have been updated. Figure 11 illustrates the modified ResNet layers.

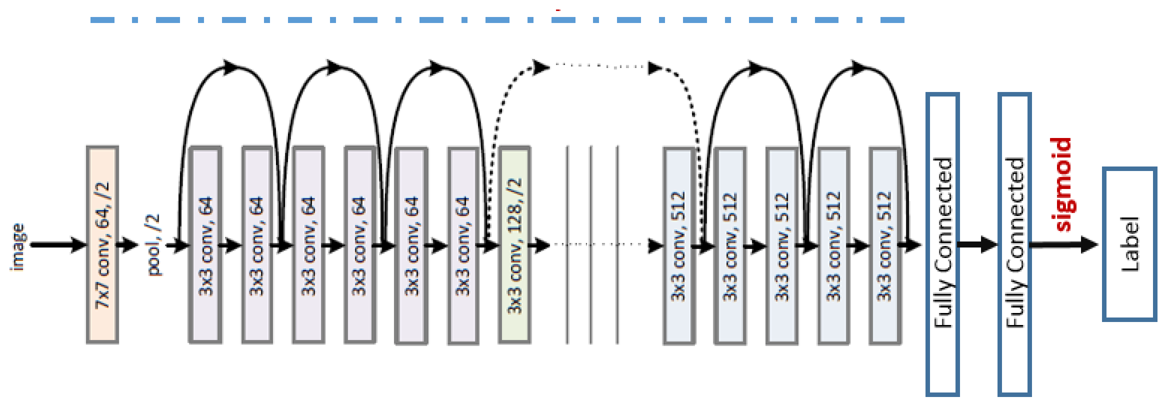


Figure 11: the modified ResNet layers by this research

Learning rate is modified from 0.001 to 0.0001. Setting the smaller learning rate assists network to learn with the smaller steps and prevent to lose global optima as it has been illustrated in Figure 12 shows the straight relation between the number of frozen layers and accuracy in this figure, the red color indicates the learning accuracy and the blue color denotes the network evaluation accuracy in different epochs with the frozen layers alternates from 30 to 90 with the step 30 respectively from left to right.

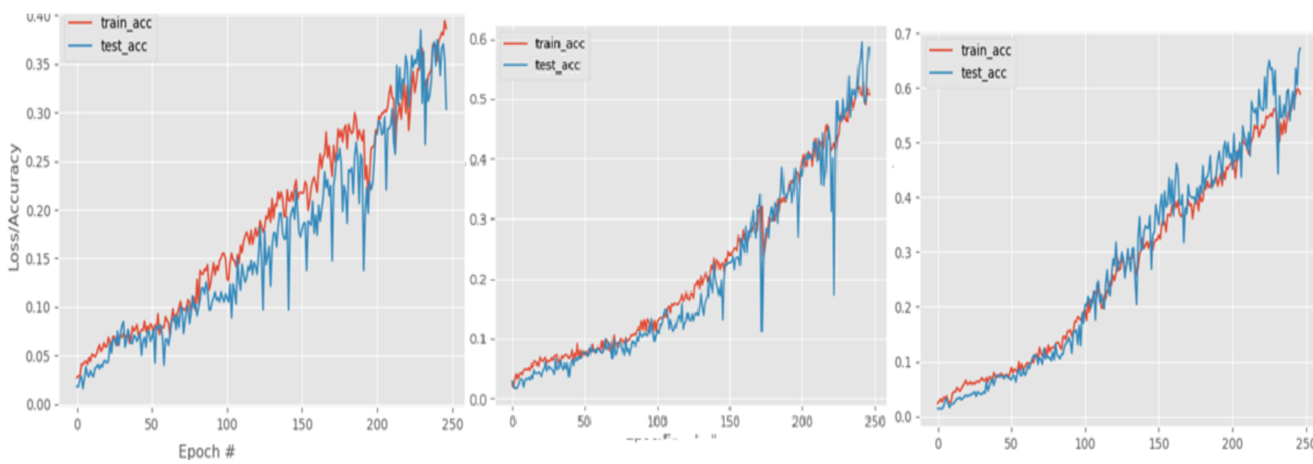


Figure 12. Illustration of the transfer learning main concept regarding freezing the hidden layers. From left to right respectively shows 30, 60 and 90 frozen layers in modified ResNet method proposed by this research which trained on Yale-B dataset. The comparison of three figures with different frozen layers proved that the network accuracy has straight proportional relation with the number of frozen layers. The red color is used as an indicator of training accuracy nevertheless blue color is representer of evaluation cross-validation. the horizontal axis is responsible to present the number of epochs and the vertical axis(y) is the illustrator of the network accuracy in evaluation or training.

In this deep approach utilizing sigmoid function class recognition and occlusion, classification can be performed at the same time. It means during training network each sample has two labels. Since occlusion probability and the class of person are two independent parameters then can be trained synchronizing. Using softmax as a loss function the network only be able to predict class label. In this stage, similar to SRC family method the prediction accuracy based on the percentage of occlusion has been calculated. In order to evaluate and compare the result of this part with SRC family the random block of monkey faces has been added to the training dataset in the different location.

Deep face identification approach has been tested on four different datasets including AR, UMB, Yale-B and ShirazU. As stated before, UMB dataset contains random and non-consistent occlusion made by things such as a hat, scarf, glasses or some unpredictable occlusion made by a hand or Cup (Cusano and Schttin , 2012).
Additionally, Yale-B dataset contains the variety in illumination with the different angle of light (Yale , 2001)
The next section contains the experimental results and evaluation of this deep face network method.

3.4. Evaluation and Experimental Result of deep face method

Table 1 indicates the implemented results of deep face method on AR dataset for sunglass and Scarf occlusion. these results have been compared to SRC family methods.

Table 1: Prediction result of ResNet deep method on AR and comparison to SRC family

Accuracy occlusion	SRC	ESRC	SSRC	GSRC	Our method
10*10 random block	100	100	100	100	100
20*20 random block	92.6±0.12	96.5±0.52	98.8±0.13	97.2±0.61	98.01±0. 15
30*30 random block	86.3±0.26	87.5±0.04	94.6±0.61	86.5±0.06	97.2±0.15
40*40 random block	72.1±0.06	78.9±0.21	81.4±0.18	75.9±0.12	84.4±0.31
50*50 random block	65.3±0.41	68.2±0.11	74.8±0.32	74.4±0.31	79.6±0.29

Table 2: The results of the recognition of occlusion type on AR dataset.

Accuracy Method	Illumination	Sunglasses	Scarf	Non-Occlusion
Our method	92.14±0.11	99.51±0.47	79.27±0.02	82.28±0.43

Table 3 presents the accuracy of recognition on Yale-B dataset occluded by monkey face random

Accuracy Method	Illumination	Sunglasses	Scarf	Normal + Expression
SRC		81.23±0.40	67.11±0.21	
GSRC		79.83±0.13	85.45±0.32	
ESRC		84.18±0.01	87.17±0.11	
SSRC		85.72±0.11	87.74±0.13	
ResNet based method	97.21±0.18	87.51±0.43	91.13±0.09	98.16±0.28

block.

Table 3: the recognition accuracy of Yale-B occluded by monkey face.

Table 4 includes the accuracy recognition of ResNet based method on UMB dataset

Table 4: ResNet based recognition accuracy on UMB dataset

Accuracy / Method	Hand	Glass	Scarf	Hat	Non-Occlusion
Our method	79.49 ± 0.11	85.13 ± 0.06	76.08 ± 0.22	81.15 ± 0.17	87.32 ± 0.12

Table 5 denotes the accuracy of the type of occlusion recognition on UMB Dataset

Table 5: Occlusion type recognition accuracy on UMB dataset.

Accuracy / Method	Hand	Glass	Scarf	Hat	Non-Occlusion
Our method	73.56 ± 0.25	34.66 ± 0.12	76.33 ± 0.06	83.92 ± 0.11	89.25 ± 0.46

The experimental results of the occlusion type recognition prove that in the most cases the occlusion relevant to glasses are not recognized and the sample with such occlusion has been recognized as a normal sample without any occlusion.

Conclusion

In order to investigate the issues regarding occlusion in face identification and address single sample trained dataset, this research first offered an approach base on feature extraction, mask projection and kNN. Then it recommended a method based on transfer learning to generate discriminative descriptors. To generate the random occluded trained dataset this study augmented occluded faces with random blocks. Face recognition and occlusion classification can be performed at the same time using the ResNet network as a frame and sigmoid as a loss function. The research conducted as part of an ongoing project to prevent cheating/fraud on remote online assessment or increase the level of security on online banking. The proposed method by this research to address this issue is MultiFactor Authentication by face as a biometric factor. Since face recognition is usually affected by uncontrolled conditions, thus this research proposed two methods as a solution with reasonable accuracy and speed.

References

Abhishree, T., J. Latha, K. Manikantan, and S. Ramachandran

2015. Face recognition using gabor filter based feature extraction with anisotropic diffusion as a pre-processing technique. Procedia Computer Science, 45(Supplement C):312 – 321. International Conference on Advanced Computing Technologies and Applications (ICACTA).

Alyuz, N., B. Gokberk, and L. Akarun
2008. A 3d face recognition system for expression and occlusion invariance. In 2008 IEEE Second International Conference on Biometrics: Theory, Applications, and Systems, Pp. 1–7.

Amos, B., B. Ludwiczuk, and M. Satyanarayanan
2016. Openface: A general-purpose face recognition library with mobile applications. CMU School of Computer Science.

Azam, M., M. A. Anjum, and M. Y. Javed
2010. Discrete cosine transform (dct) based face recognition in hexagonal images. In 2010 The 2nd International Conference on Computer and Automation Engineering (ICCAE), volume 2, Pp. 474–479.

Chen, C., A. Dantcheva, and A. Ross
2013. Automatic facial makeup detection with application in face recognition. In 2013 International Conference on Biometrics (ICB), Pp. 1–8.

Choi, D. H., I. H. Jang, M. H. Kim, and N. C. Kim
2007. Color image enhancement based on single-scale retinex with a jnd-based nonlinear filter. In 2007 IEEE International Symposium on Circuits and Systems, Pp. 3948–3951.

Colombo, A., C. Cusano, and R. Schettini
2011. UMB-DB: A database of partially occluded 3D faces. In 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), Pp. 2113–2119.

Deng, W., J. Hu, and J. Guo
2012. Extended src: Undersampled face recognition via intraclass variant dictionary. IEEE Transactions on Pattern Analysis and Machine Intelligence, 34(9):1864–1870.

Drira, H., B. B. Amor, A. Srivastava, M. Daoudi, and R. Slama
2013. 3D Face Recognition under Expressions, Occlusions, and Pose Variations. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(9):2270–2283.

Du, S. and R. Ward
2006. Face recognition under pose variations. Journal of the Franklin Institute, 343(6):596 – 613. Winners of the student paper competition at the 2005 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP) held in Philadelphia PA, provide the state-of-art in their fields of research.

Fierrez, J., J. Ortega-Garcia, A. Esposito, A. Drygajlo, and M. Faundez-Zanuy
2009. Biometric ID Management and Multimodal Communication: Joint COST 2101 and 2102 International Conference, BioID_MultiComm 2009, Madrid, Spain, September ... Vision, Pattern Recognition, and Graphics), 1st edition. Springer Publishing Company, Incorporated.

Geng, C. and X. Jiang
2009. Sift features for face recognition. In 2009 2nd IEEE International Conference on Computer Science and Information Technology, Pp. 598–602.

Grant, M. and M. Stewart
2012. Asystems model for probabilistic risk assessment of improvised explosive device attack. Int. J. of Intelligent Defence Support Systems, 5: rt Systems.

Gruber, I., M. Hlaváˇc, M. Železný, and A. Karpov
2017. Facing Face Recognition with ResNet: Round One, Pp. 67–74. Cham: Springer International Publishing.

Guo, T., X. H. Tan, and C. X. Chao
2016a. Robust Face Recognition via Occlusion Detection and Masking. In MATEC Web of Conferences, volume 44. EDP Sciences.

Guo, T., X. H. Tan, and C. X. Chao
2016b. Robust Face Recognition via Occlusion Detection and Masking. In MATEC Web of Conferences, volume 44.

He, K., X. Zhang, S. Ren, and J. Sun
2016. Identity Mappings in Deep Residual Networks. CoRR, abs/1603.05027.

Huang, G. B., M. Ramesh, T. Berg, and E. Learned-Miller
2007. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical report, Technical Report 07-49, University of Massachusetts, Amherst.

Ioffe, S. and C. Szegedy
2015. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. CoRR, abs/1502.03167.

Jain, A. K., A. Ross, and S. Prabhakar
2004. An introduction to biometric recognition. IEEE Transactions on circuits and systems for video technology, 14(1):4–20.

Jia, H. and A. M. Martinez
2008a. Face recognition with occlusions in the training and testing sets. In Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on, Pp. 1–6.

Jia, H. and A. M. Martinez
2008b. Face recognition with occlusions in the training and testing sets. In Automatic Face
& Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on, Pp. 1–6. IEEE.

Martinez, A. M.
1998. The AR face database. CVC technical report.

Milborrow, S., J. Morkel, and F. Nicolls
2010. The MUCT Landmarked Face Database. Pattern Recognition Association of South Africa.

Ou, W., X. You, D. Tao, P. Zhang, Y. Tang, and Z. Zhu
2014. Robust face recognition via occlusion dictionary learning. Pattern Recognition, 47(4):1559–1572.

Parkhi, O. M., A. Vedaldi, and A. Zisserman
2015. Deep Face Recognition. In British Machine Vision Conference.

Sahoolizadeh, H. and Y. A. Ghassabeh
2008. Face recognition using eigen-faces, fisher-faces, and neural networks. In 2008 7th IEEE International Conference on Cybernetic Intelligent Systems, Pp. 1–6.

Turk, M. and A. Pentland
1991. Eigenfaces for recognition. J. Cognitive Neuroscience, 3(1):71–86.

Wattenberg, M., F. ViÃ©gas, and I. Johnson
2016. How to use t-sne effectively. Distill.

Weiss, K., Khoshgoftaar, Taghi M.,Wang,D.2016.A survey of transfer learning,Journal of Big Data

Wen, Y., W. Liu, M. Yang, Y. Fu, Y. Xiang, and R. Hu
2016. Structured occlusion coding for robust face recognition. Neurocomputing, 178:11–
24.

Wright, J., A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma
2009. Robust face recognition via sparse representation. IEEE transactions on pattern analysis and machine intelligence, 31(2):210–227.

Xie, S., R. B. Girshick, P. Dollár, Z. Tu, and K. He
2016. Aggregated residual transformations for deep neural networks. CoRR, abs/1611.05431.
Yale. (2001). The Extended Yale Face Database B. Retrieved from

http://vision.ucsd.edu/~iskwak/ExtYaleDatabase/ExtYaleB.html

Yang, M. and L. Zhang
2010. Gabor feature based sparse representation for face recognition with gabor occlusion dictionary. Computer VisionâŁ“ ECCV 2010, Pp. 448–461.