

More Realistic and Efficient Face-Based Mobile Authentication using CNNs

Abhijit Das^{a,c,e}, Abira Sengupta^{b*}, Muhammad Saqib^{e*}, Umapada Pal^d, and Michael Blumenstein^e

^aInstitute for Integrated and Intelligent Systems, Griffith University, Queensland, Australia
abhijit.das@griffithuni.edu.au

^bDepartment of Computer Science, Kalyani Government Engineering College, Kalyani, India.
enggabira0609@gmail.com

^cInria Sophia Antipolis – Méditerranée, France

^dComputer Vision and Pattern Recognition Unit, Indian Statistical Institute, Kolkata, India, umapada@isical.ac.in

^eCenter for Artificial Intelligence, School of Software, University of Technology Sydney, Australia,
michael.blumenstein@uts.edu.au, Mohammad.Saqib@student.uts.edu.au

Abstract

In this work, we propose a more realistic and efficient face-based mobile authentication technique using CNNs. This paper discusses and explores an inevitable problem of using face images for mobile authentication, taken from varying distances with a front/selfie camera of the mobile phone. Incidentally, once an individual comes towards a certain distance from the camera, the face images get large and appear over-sized. Simultaneously sharp features of some portions of the face, such as forehead, cheek, and chin are changed completely. As a result, the face features change and the impact increases exponentially once the individual crosses a certain distance and gradually approaches towards the front camera.

This work proposes a solution (achieving better accuracy and facial features, whereby face images were cropped and aligned around its close bounding box) to mitigate the aforementioned identified gap. The work investigated different frontier face detection and recognition techniques to justify the proposed solution. Among all the employed methods evaluated, CNNs worked best. For a quantitative comparison of the proposed method, manually cropped face images/annotations of the face images along with their close boundary were prepared. In turn, we have developed a database considering the above-mentioned scenario for 40 individuals, which will be publicly available for academic research purposes. The experimental results achieved indicate a successful implementation of the proposed method and the performance of the proposed technique is also found to be superior in comparison to the existing state-of-the-art.

Keywords: Face biometrics, CNN, Mobile biometrics.

1. Introduction

Accessing information on mobile devices has become mainstream nowadays; besides the clear benefits that mobile devices provide as a means to improve efficiency, productivity, and user convenience, they, in turn, require proper methods for secure access control. For this reason, governments and private industry are turning to mobile biometrics to speed up processing of people and goods in financial services and banking, law enforcement, military, public transportation, border control, healthcare and commercial shipping. ¹Acuity Market Intelligence released a new research report which forecasts annual revenue from the mobile biometrics market will grow at nearly 41 percent CAGR from \$6.5 billion in 2016 to \$50.6 billion in 2022 [1]. In addition, there will be more than 1.37 trillion in payment and non-payment mobile device transactions that require some level of biometric authentication. To date, biometrics have been widely adopted as a convenient PIN alternative for unlocking smartphones. The major biometric traits used for mobile authentication are a face, iris, and fingerprint. Fingerprint requires an additional sensor, iris-based mobile authentication is intrusive. Whereas, the face trait does not possess these disadvantages.

For these reasons, the face trait is used for a wide variety of application scenarios, however there are some scenarios in real life applications where face recognition still poses certain limitations. One such phenomenon happens when an individual comes closer to the camera; the face image captured in such a scenario becomes large and over-sized, which appear as different types of facial appearance deformation. Face recognition is truly-based upon

¹ <http://www.biometricupdate.com/biometric-news/mobile-biometric-news>

* The second and the third author have an equal contribution towards the paper

anatomical/shape and texture-based facial components. Therefore, it can be easily assumed that such a scenario will highly impact the performance of the system. Such scenarios are also true for face image based mobile authentication. For face image-based mobile authentication, the front camera or the selfie camera of the phone is usually employed. Our observations suggest that the face images captured with this camera for mobile authentication gets similar over-sized and distorted effects once an individual comes too close to the camera. In real life, an individual can hold the camera at different distances while getting authenticated. Examples of face images in such a scenario are given below in Figure 1.

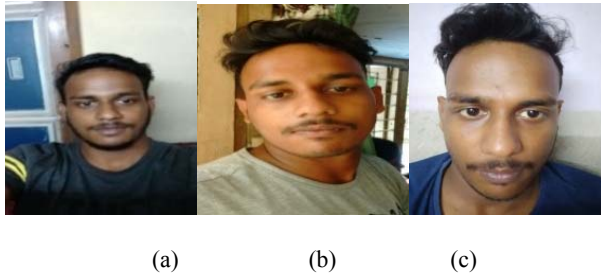


Figure 1: Face images of an individual collected by himself at a different distance while authenticating himself on his mobile.

To the best of our knowledge, this problem of face-based mobile authentication has not been investigated in the literature in terms of facial key point detection, landmark detection and by 3D face image reconstruction. The main reason for this distortion was assumed to be the optical phenomenon of refraction of the object through the camera front lens. If the object crosses the focal length of the front lens, the distortion is found. Furthermore, when the images are captured in the wild, it can contain external specular reflection, and other external objects (for example in Figure 1(a & b)) that is assumed to affect the working principle of key point detection, landmark detection and 3D face image reconstruction. To fulfil the aforementioned gap, the present work is proposed.

Over the last few decades in the machine learning literature, Deep Convolutional Neural Networks have been thoroughly investigated. We assume that Deep Convolutional Neural Network-based alignment can be a significant advancement. The specific contributions of this work are:

1. A novel technique to handle the impact/deformation of face images captured from the front camera of a mobile phone at different distances is proposed.
2. A real-life dataset of face images of individuals captured from the front camera of a mobile phone is proposed.

The remainder of the paper is organized as follows:

Section 2 summarizes the literature, Section 3 highlights the proposed methodology, followed by Section 4 which highlights the experimental setup and results, and Section 5 draws an overall conclusion and future scope.

2. Literature Review

For face-based person identification, a significant amount of work has been undertaken in the literature. Many different methods for face identity establishment have been proposed during the past few decades [2]. For feature extraction techniques, seminal works were proposed for holistic matching methods such as Principal Component Analysis (PCA) [3], Eigenfaces, Probabilistic Eigenfaces [4, 5], Fisher faces/subspace LDA [6, 7, 8], SVM [9], Evolution pursuit [10], Feature lines, ICA [11], etc. The next category of feature extraction techniques used were Feature-based methods such as pure geometry-based methods [12], Dynamic link architecture [13], Hidden Markov model [14], and Convolutional Neural Networks [15]. Other categories of feature extraction techniques used include Modular Hybrid LFA [16], Shape-normalized [17], component-based [18], etc. Some other categories of feature extraction techniques used were LDA/FLD [19, 20], PDBNN etc. Reduced Eigenface is used in [21], and DLBP is used in [22, 23].

In the area of facial deformation research topics such as facial key point detection, landmark detection and 3D face image reconstruction have been investigated. Without considering some of the major challenges, a face database is proposed in [25], and a distance estimation techniques has been proposed in [24]. In the context of key-point and landmark detection from around main facial components such eyes, nose, chin, and mouth. The facial feature points are rich, discriminative and cover the most important parts of the face. The manually labelled feature points represent a shape, and these points along with their corresponding images are given to the training system to learn the model between appearance variation and shape variation. The trained model is used to localize the shape of the facial feature points according to facial appearance. However, detection of such shapes is a challenging task due to the deformation of the face. There are some methods based on carefully designed features such as SIFT [30], pixel difference features [28], local binary features [31]. However, all these methods cannot handle the variations in shape and appearance. A comprehensive survey on facial feature point detection can be found in [32]. The survey categorized the existing methods into the following four groups: constrained local mode, active appearance model, regression-based, and other methods.

Though significant progress has been made, facial feature point detection is limited in its success by in-the-wild and real-world conditions e.g. variations across poses.

expressions, illuminations, and occlusions. Although some state-of-the-art methods are ostensibly comparable to humans on some databases, there remain challenges in detecting occluded faces or those with large shape variations. Furthermore, most existing real-world databases are composed of frontal or near frontal images.

To address this scenario, a cascaded elastically progressive model for accurate face alignment is proposed in [23] in recent work. All of these scenarios is found to be true in our experimental scenario.

Currently, deep learning provides state-of-the-art results on face detection and recognition due to its ability to learn the relevant features [31]. The learned features perform much better than hand-crafted features due to the rich, expressive, and discriminative nature of the learned features. Thus, we hypothesize to propose a technique based on Deep-CNNs to solve the above-highlighted challenge.

3. Proposed Methodology

Apart from the general face recognition challenges, additional challenges can be faced in the scenario of face capture at a varying distance by mobile smartphone front camera. The main challenges that can be faced in the close-up distance are intra-class variance due to optical phenomena, i.e. refraction. Conventionally, human face image appearance has potentially very large intra-subject variations due to 3D pose, illumination (including indoor/outdoor), facial expression, occlusion due to other objects or accessories (e.g., sunglasses, scarf, etc.), facial hair and aging. These types of inter-subject variances are dealt with in the literature with robust feature and classification techniques. Excluding these challenges, in the scenario where face images are acquired close up, the intra-subject variance in face images captured from close-up include distortions of the face, and they appear over-sized. By observing the swelling formation or abnormal enlargement of a face part, it appears to have an abnormal impact on the personal identification scenario. The main reason for this distortion is assumed to be the optical phenomenon of refraction of the object through the camera lens. If the object crosses the focal length of the lens, then a distortion is found.

On observation, we can find that in this scenario of facial features getting distorted, these mostly appear around the perimeter of the face, and hence the face images were pre-processed by cropping and aligning them around the face boundary.

The proposed face pre-processing as well as the recognition system employed, are discussed in the next subsections.

3.1. Face Pre-processing

We have employed a couple of advanced face detection techniques. They are as follows:

Viola-Jones face detection: The first detection technique used is the MATLAB built-in class and function. Based on Viola-Jones face detection algorithm [27], the computer vision system toolbox contains the vision.CascadeObjectDetector System object which detects objects based on the above-mentioned algorithm. For fast and accurate detection and for efficient computational resource allocation, we used three components namely: a) The integral image for feature computation, b) Adaboost for feature selection and c) An attentional cascade. For extracting features, this algorithm uses Haar features, which is scalar product between the image and some Haar templates. For face detection, Adaboost assumes the detection, where 1 means that there is a face and -1 signifies the opposite, as well as the number of Haar-like features extracted from an image. Adaboost aims to iteratively drive down an upper bound of the empirical loss. For more efficiency, the simpler classifiers are used to reject the majority of sub-windows, then more complex classifiers by cascade are called upon to achieve low false positive rates. It can detect the face bounding box from which the face images are cropped by hard coding.

CNNs: For face detection and alignment, we have employed CNNs. The performance of the alignment process is degraded by variation in poses, lighting conditions and occlusion. In recent years, Convolutional Neural Networks (CNNs) have shown significant improvements over other methods based on traditional machine learning. Traditional machine learning techniques use hand-crafted features which have a low representation capability for objects. Whereas CNNs learn the generic features which encode the domain knowledge in a much better way.

Therefore, we have adopted a cascaded multi-task CNN for the face detection and alignment problem [27], and finally to crop the image around the face boundary. The adapted CNN works in three stages. In the first stage, a shallow network is used to generate potential proposals for faces. A Non-Maxima Suppression (NMS) is applied to remove redundant bounding boxes. The second stage acts as a refinement for further removal of false proposals. Finally, the third stage extracts the detailed information and outputs five landmark positions.

3.2. Face Recognition

The recognition techniques used are summarised below.

CNN: We have employed the FaceNet model proposed in [26], that directly learns a mapping from face images to a compact Euclidean space where distances directly correspond to a measure of face similarity.

Once this space has been produced, tasks such as face recognition, verification and clustering can be easily implemented using standard techniques with FaceNet embedding as feature vectors.

The method uses a deep convolutional network trained to directly optimize the embedding itself, rather than an intermediate bottleneck layer as in previous deep learning approaches. To train the Inception-Resnet-v1 model, it was used as a classifier, i.e., not using Triplet Loss as described in the FaceNet paper [26]. The initiation trained/weight was performed on the CASIA/Facescrub datasets, looking at the accuracy and the validation rate ($VAL@FAR=10^{-3}$). A batch size of 1000, iterations 300, and epoch 1 was used.

4. Experimental Results

The experimental result conducted are detailed in this section along with the proposed dataset.

4.1. Proposed Dataset

A dataset was created to conduct the experiments of the proposed methodology. To the best of our knowledge, no dataset exist in the literature which covers the proposed subject of research.

The proposed dataset was collected from 40 individuals, including 28 males, and 12 females all varying in age between 19-65. Different types of mobile cameras available in the market (Motorola, iPhone, Samsung, etc.) were used for data collection to test the robustness of the technique.

Each individual captured all their images by one type of sensor, using the front camera. The images were captured at different times of the day and in the time period of 30 days.

Due to this long phase of the collection, a huge intraclass variation can be noted in the dataset (an individual is wearing different, dress, makeup, etc.). Additionally, the individuals captured their images holding the phone as if they are attempting to get authenticated to their mobile via a face image.

The images in the dataset were collected at three different distances: close to camera ($< 1\text{ft}$ and $> 0.7\text{ft}$) distance 3, moderately far ($< 1\text{ft}$ and $> 1.3\text{ft}$) distance 2 and far ($< 1.3\text{ft}$ and $> 1.6\text{ft}$) distance 1. All this makes the dataset more realistic and close to a real life scenario.

In addition to the major challenges highlighted, the other basic challenges such as, blurring (when a person takes a picture using the front camera lens, due to hand instability, the picture may get blurred).

Furthermore, different intensities of light from the background (which may degrade the quality of the face image as the intensity of light plays a vital role in capturing a quality image). Persons with different types of makeovers are also included in the database. Some examples of such images are included in Figure 3.

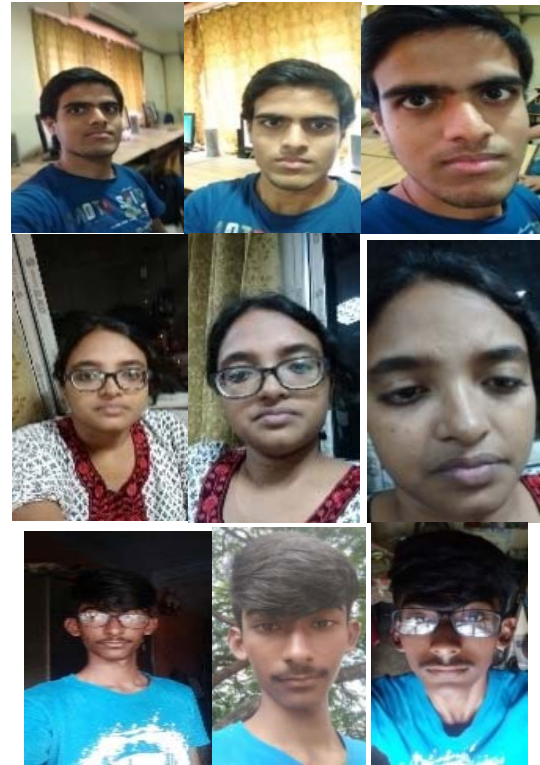


Figure 2: Examples of face images – some individuals from the dataset at different distances from the camera

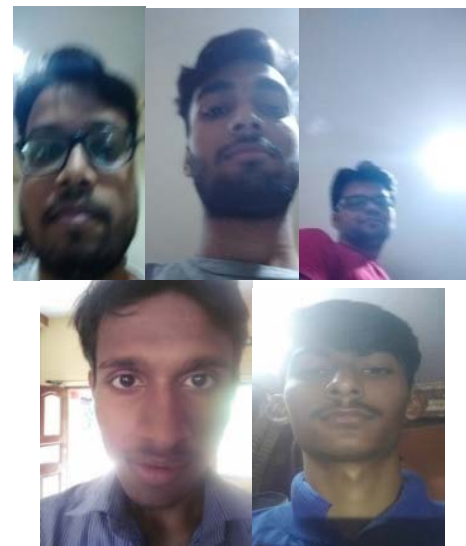


Figure 3: Images with noisy backgrounds, uneven illuminations etc.

For each individual, 40 images were collected, 5 were used for training and 35 for testing (this was done to evaluate the real-life scenario where the individual will enroll once and will validate for an infinite number of

occasions). Examples of some individuals from the dataset are in Figure 2.

4.2. Results of Preprocessing Experiments

The major challenges that can exist in individual face recognition for the above-mentioned scenarios seem to be automatic segmentation, cropping, and alignment of the face images.

In this work segmentation s performed in manual mode as well as automatic mode by computer vision techniques to crop the face images along the closure boundary box of the face. Manual segmentation or ground truth is undertaken for qualitative and quantitative analysis with respect to the automatic one. Examples of these images are shown in Figure 4.

The images in the first column are the original image, the 2nd column shows the manual crop image. The 3rd column shows automatic crop image by Viola-Jones face detection and 4th column contains automatic crop images by CNN-based alignment.

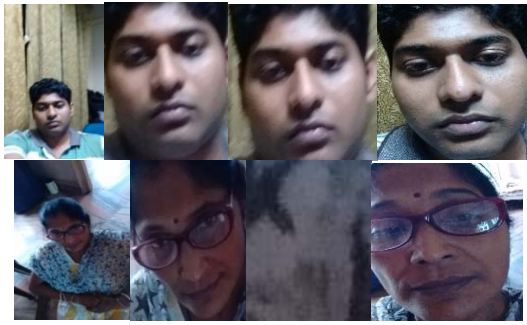


Figure 4: Examples of manually and automatically segmented images

The CNN method works well for face detection; a comparison result of both methods are summarized in Table 1. We can observe from the table that a large number of face images were not detected properly; the miss rate was high for the Viola-Jones-based face detection method.

Table 1: Results of the detection techniques where a number of images were not correctly detected

Techniques	Distance 1	Distance 2	Distance 3
CNN	6	101	783
Viola-Jones	1393	1455	1730

It can also be observed that the miss rate increased substantially with distance measures 1 to 2 and 3 ~ the 50% miss rate again reduced to 35% at distance 3. The reason might be the deformation in the face shape for distance 2 and 3 as well as illumination. Major attention for future research should be directed towards scenario 2 and 3.

Examples of some images that are missed by both the methods are shown in Figure 5.



Figure 5: Examples of some images that are missed by both face detection methods.

It is quite clear from the above discussion that there is significant room for research to advance the state-of-the-art for this subject of research.

4.3. Recognition and Verification Experimental Setup

The recognition experiment was performed with 3 sets of features. For the original and manual set, 5 images were used for training and the rest for testing. Among the previously-mentioned break down, only the aligned face images were used. The three sets of experiments were performed for each type of image in the dataset (original, automatic crop image and manual crop image).

Among the automatically cropped images, only the CNN outputs were used as it performed better. The performance of each category and the cross-distance combination mentioned, were tested to portray the realistic performance for the proposed face-based mobile authentication (an individual can hold the mobile camera from a different distance in real life). The legend used for the experiments is:

- 1st original- the original image at the far distance,
- 2nd original- the original image at the moderately far distance,
- 3rd original- the original image at the near distance,
- 1st Auto- the automatically cropped image by CNN at the far distance,
- 2nd Auto- the automatic cropped image by CNN at the moderately far distance,
- 3rd Auto- the automatically cropped image by CNN at the near distance,
- 1st Manual- manually cropped at the far distance,
- 2nd Manual - manually cropped at the moderately far distance,
- 3rd Manual - manually cropped at the near distance.

The detailed results of the recognition and verification stages are summarized in Tables 2 and 3, respectively.

Table 2: Results of the recognition accuracy (in scale of 1)

Training	Testing	CNN
1 st original	1 st _original	0.084
	2 nd _original	0.069
	3 rd _original	0.046
1 st _Auto	1 st _Auto	0.981
	2 nd _Auto	0.973
	3 rd _Auto	0.958
1 st _Manual	1 st _Manual	0.028
	2 nd _Manual	0.029
	3 rd _Manual	0.023
2 nd _original	1 st _original	0.078
	2 nd _original	0.076
	3 rd _original	0.077
2 nd _Auto	1 st _Auto	0.988
	2 nd _Auto	0.978
	3 rd _Auto	0.990
2 nd _Manual	1 st _Manual	0.094
	2 nd _Manual	0.055
	3 rd _Manual	0.057
3 rd _original	1 st _original	0.046
	2 nd _original	0.076
	3 rd _original	0.089
3 rd _Manual	1 st _Manual	0.138
	2 nd _Manual	0.083
	3 rd _Manual	0.154
3 rd _Auto	1 st _Auto	0.961
	2 nd _Auto	0.952
	3 rd _Auto	0.955

4.4. Result Discussion of Recognition Experiments

It can be concluded from the above Table 1 that a very substantial recognition accuracy is obtained. When the alignment and the manual cropping of the face images are done, the cross distance accuracy attended in those scenarios was very impressive. CNNs-based features were found to be the best of all the features used. For the auto scenario, the recognition accuracy is ~ 0.97 and for verification ~ 0.99 .

The pattern for the results of the cross-distance scenario is also very promising; for auto only, a difference of ~ 0.01 can be found in the recognition scenario and ~ 0.005 for verification. The pattern for the result of the cross-distance scenario is also very promising for auto only, a difference of ~ 0.01 can be found in the recognition scenario, and ~ 0.003 for the verification scenario.

Whereas, the performance difference between auto and original is quite high for the recognition scenario and verification. Moreover, the performance of the auto in the cross-distance is also very high for recognition and the verification scenario. The performance of the cross-distance of the original and manual was around ~ 0.35 and ~ 0.4 for recognition. The performance for verification cross-distance differs by ~ 0.04 and ~ 0.005 for original and manual, respectively.

Table 3: Results of the verification accuracy (in scale of 1)

Training	Testing	CNN
1 st original	1 st _original	0.940
	2 nd _original	0.909
	3 rd _original	0.901
1 st _Auto	1 st _Auto	0.991
	2 nd _Auto	0.991
	3 rd _Auto	0.998
1 st _Manual	1 st _Manual	0.921
	2 nd _Manual	0.930
	3 rd _Manual	0.930
2 nd _original	1 st _original	0.911
	2 nd _original	0.916
	3 rd _original	0.907
2 nd _Auto	1 st _Auto	0.988
	2 nd _Auto	0.998
	3 rd _Auto	0.995
2 nd _Manual	1 st _Manual	0.941
	2 nd _Manual	0.950
	3 rd _Manual	0.917
3 rd _original	1 st _original	0.941
	2 nd _original	0.966
	3 rd _original	0.989
3 rd _Manual	1 st _Manual	0.981
	2 nd _Manual	0.988
	3 rd _Manual	0.983
3 rd _Auto	1 st _Auto	0.988
	2 nd _Auto	0.981
	3 rd _Auto	0.995

4.5. Result Discussion of Verification Experiments

The verification recognition performance of all the non-cross-distance scenario cases is very promising. As an overall conclusion, it can be stated that the recognition performance was poorer than the verification scenario. We can also state that the cropping and alignment of the face images have highly boosted the performance of both the

verification and the recognition scenario. In this context, we must also mention that the performance of the auto experimental scenario is slightly boosted because of the smaller number of test samples, as almost 50% of the images could not be aligned. Examples of false accepted and false rejected images from the auto experiments for each distance are given in Figures 6 and 7, respectively.

It can be assumed from the above, given face images after alignment and cropping, that the misclassification of the face images is mostly due to the illumination and other noise, not due to the face shape or feature deformation. Therefore, it can be considered that the proposed solution is successfully proven by the experiments.

4.6. State-of-the-Art Comparison

For a *state-of-the-art* comparison, we performed all the experiments with PCA, LDA, Fisher face, Eigen's face, and they were found to obtain ~13% and ~7% less performance than CNNs for recognition and verification, respectively.

We also compared the proposed work with the most recent and related work that can be found in the literature related to facial alignment [30] and facial key-point detection [23]. The proposed work was found to outperform in most of the scenario, and especially in the cross-distance scenario by ~3.5% and ~2% for recognition and verification, respectively. The main failure of these algorithms can be found in the scenario of challenging illumination and pose change.

5. Conclusions and Future Scope

In this work, we propose a robust mobile authentication system using face images. The face images in this investigation can get affected by optical phenomena due to the distance of the individual from the camera/sensor. We identified the challenges and the reason for face image deformation in this scenario. Further, we proposed the pre-processing scenario to resolve the problem by cropping and aligning the face around its close boundary. This helps to achieve higher accuracy by reducing the impact of deformation on the face, and by reducing background and redundant features. The face is detected then cropped manually and aligned along the face boundary for the image captured using CNNs.

The various challenges of this scenario are identified, and a dataset for this investigation was developed. Different detection and recognition techniques of face recognition were used, and CNNs were found to be effective. The developed database will be publicly available for academic research. The initial investigation substantiates the possibility of using cross-distance face images after certain pre-processing techniques.

Future work will involve expanding the database and employing a more robust method for face detection to resolve the identified gaps and open research areas identified.

Acknowledgement

The authors sincerely appreciate the volunteers who submitted their samples for dataset development, and those who helped in the collection process. Upon request, the database is available for download to the research community.



Figure 6: Examples from false accepted auto images at different distances



Figure 7: Examples from falsely rejected auto images at different distances

References

- [1] The U.S. mobile app report | Tech. Rep., comScore (August 2014).
- [2] I. Craw, P. Cameron, Face Recognition By Computer. In Proceedings, British Machine Vi-Sion Conference. 489–507. 1996.

- [3] M. Kirby, L. Sirovich, Application Of The Karhunen-Loeve Procedure For The Characterization Of Human Faces. *IEEE Trans. Patt. Anal. Mach. Intell.* 1990.
- [4] M. Turk, A. Pentland. Eigenfaces For Recognition. *Cogn. Neurosci.* 3, 72–86, 1991.
- [5] B. Moghaddam, A. Pentland. Probabilistic Visual Learning For Object Representation. *IEEE Trans. Patt. Anal. Mach. Intell.* 19, 696–710, 1997.
- [6] P. N. Belhumeur, J. P. Hespanha, D. J. Kriegman. Eigenfaces Vs. Fisherfaces: Recognition Using Class Specific Linear Projection. *IEEE Trans. Patt. Anal. Mach. Intell.* 19, 711–720, 1997.
- [7] D. L. Swets, J. Weng, Using Discriminant Eigenfeatures For Image Retrieval. *IEEE Trans. Patt. Anal. Mach. Intell.* 18, 831–836, 1996.
- [8] W. Zhao, R. Chellappa, A. Krishnaswamy, Discriminant Analysis Of Principal Components For Face Recognition. In *Proceedings, International Conference On Automatic Face And Gesture Recognition.* 336–341, 1998.
- [9] P. J. Phillips, Support Vector Machines Applied To Face recognition. *Adv. Neural Inform. Process. Syst.* 11, 803–809, 1998.
- [10] C. Liu, H. Wechsler, Evolutionary Pursuit And Its Application To Face Recognition. *IEEE Trans. Patt. Anal. Mach. Intell.* 22, 570–582, 2000.
- [11] Z. Li, J. Lu, Face Recognition Using The Nearest Feature Line Method. *IEEE Trans. Neural Netw.* 10, 439–443, 1999.
- [12] M. S. Bartlett, H. M. Lades, T. Sejnowski, Independent Component Representation For Face Recognition. In *Proceedings, SPIE Symposium On Electronic Imaging: Science And Technology.* 528–539, 1998.
- [13] K. Etemad, R. Chellappa. Discriminant Analysis For Recognition Of Human Face Images. *J. Opt. Soc. Am. A* 14, 1724–1733, 1997.
- [14] S. H. Lin, S. Y. Kung, L. J. Lin, Face Recognition/Detection By Probabilistic Decision-Based Neural Network. *IEEE Trans. Neural Netw.* 8, 114–132, 1997.
- [15] T. Kanade, Computer Recognition Of Human Faces. Birkhauser, Basel, Switzerland, And Stuttgart, Germany, 1973.
- [16] M. D. Kelly, 1970. Visual Identification Of People By Computer. Tech. Rep. AI-130, Stanford AI Project, Stanford.
- [17] I. J. Cox, J. Ghosn, P. N. Yianilos, Feature-Based Face Recognition Using Mixture-Distance. In *Proceedings, IEEE Conference On Computer Vision And Pattern Recognition.* 209–216, 1996.
- [18] B. S. Manjunath, R. Chellappa, C. V. D. Malsburg. A Feature Based Approach To Face Recognition. In *Proceedings, IEEE Conference On Computer Vision And Pattern Recognition.* 373–378, 1992.
- [19] K. Okada, J. Steffans, T. Maurer, H. E. Hong, E. Lagin, H. Neven, C. V. D. Malsburg, The Bochum/USC Face Recognition System And How It Fared In The Feret Phase III Test. In *Face Recognition: From Theory To Applications*, 186–205, 1998.
- [20] N. Brümmer, J. du Preez, Application-independent evaluation of speaker detection, *Computer Speech & Language*, Volume 20, Issues 2–3, April–July 2006, Pages 230–275.
- [21] A. Das, “Face Recognition in reduced Eigen plane”, *International Conference on Communications, Devices and Intelligent Systems*, pp-620–623, 2012.
- [22] A. Das, A. Sengupta, U. Pal, M. F. Ballaster and M. Blumenstein, Linking Face Images Captured from the Optical Phenomenon in the Wild for Forensic Science, at *International Joint Conference on Biometrics*, 2017.
- [23] W. Yang, X. Sun, Q. Liao, Cascaded Elastically Progressive Model for Accurate Face Alignment, *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, Volume: 47, Issue: 9, Sept. 2017.
- [24] A. Flores, E. Christiansen, D. Kriegman, and S. Belongie, Camera Distance from Face Images, *ISVC 2013, Part II, LNCS* 8034, pp. 513–522, 2013.
- [25] C. McCool, S. Marcel, A. Hadid, M. Pietikäinen, P. Matějka, J. Černocký, N. Poh, J. Kittler, A. Larcher, C. Lévy, D. Matrouf, J. Bonastre, Phil Tresadern, and Timothy Cootes, “Bi-Modal Person Recognition on a Mobile Phone: using mobile phone data”, in *IEEE ICME Workshop on Hot Topics in Mobile Multimedia*, 2012.
- [26] F. Schroff, D. Kalenichenko, J. Philbin, FaceNet: A Unified Embedding for Face Recognition and Clustering (Submitted on 12 Mar 2015 (v1), last revised 17 Jun 2015 (this version, v3))
- [27] K. Zhang, Z. Zhang, Z. Li, Y. Qiao, Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Networks, Submitted to *IEEE Signal Processing Letters*.
- [28] Ren S, Cao X, Wei Y, Sun J (in press, 2014) Face alignment at 3000 fps via regressing local binary features. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*
- [29] Cao X, Wei Y, Wen F, Sun J (2012) Face alignment by explicit shape regression. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp 2887–2894.
- [30] Xiong X, De la Torre F (2013) Supervised descent method and its application to face alignment. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp 532–539
- [31] Bengio Y, Courville A, Vincent P (2013) Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35(8):1798–1828
- [32] N. Wang · X. Gao · D. Tao · X. Li, Facial Feature Point Detection: A Comprehensive Survey, *arXiv:1410.1037v1 [cs.CV]* 4 Oct 2014.