Face Recognition in an Unconstrained Environment using ConvNet

Kunal Chaturvedi[1]
[1]Department of Applied Mathematics
Delhi Technological University
New Delhi, India
kunalchaturvedi98@gmail.com

Dinesh Kumar Vishwakarma[2]
[2]Department of Information Technology
Delhi Technological University
New Delhi, India
dinesh@dtu.ac.in

ABSTRACT
With the recent advancements in the discipline of Facial Recognition, it has made it easier to detect and identify multiple faces at a time in a situation where the subjects could have varying face posture, expressions, appearances in a dark or lighted background. In this paper, the detection of faces from the captured images having single or multiple faces in an unconstrained environment is done by Histogram of Oriented Gradients (HOG) feature descriptor and SVM classifier which is not only fast and accurate but also improves the accuracy of the proposed deep convolutional architecture which performs by learning feature representations to identify whether the subject is present in the image or not. The experiments are conducted on the DTU Biometric Research Lab, AT&T, and the Essex Face databases. In real-time, the proposed methodology performs excellent with high recognition accuracy.

CCS Concepts
• **Computing methodologies→Object recognition**

Keywords
Deep learning; face detection; face verification; HOG descriptors; biometrics; Convolution Neural Networks

1. INTRODUCTION
The need for authentication of an individual by employing unique physiological identifiers of the face has led to the subsequent developments in Face Recognition technology. The end-user requires minimal physical contact with this technology. It has proven to work more robust than its counterparts including DNA, fingerprints, voice, iris, hand geometry which are prone to flaws [1] and require high installation costs. Among all the aforementioned systems, face recognition provides a faster processing speed with higher efficiency. It was first implemented in the year of 1964 by Woodrow Bledsoe on RAND Tablet, a man-labeled machine that would extract the coordinates and calculate the width of mouth and eyes, distance from pupil to pupil and so on. Since then, the efficiency of face recognition technology has been significantly improved with the advancements in the algorithms. It is creating a positive influence on the daily lives of the current generation due to its user-friendly interface.

In the recent years, Face recognition has paved the way for real-world implementations by preventing fraud voters, quantifying human expressions, tracking attendance, finding solutions for security and the surveillance system that are unmatched with the human counterpart [2]. In most organizations such as offices and universities, attendance is taken manually which is a time-consuming process requiring heavy paperwork. The process of manual identification is being replaced with the face identification systems which perform the task in seconds. However, most of the biometric systems for face recognition works on one face at a time which when implemented on multiple subjects is exhausting and difficult to enact on hardware with limited configurations. Illumination conditions [3], occlusions, different expressions [4] pose a big challenge to the Face Recognition.

The paper proposes a fast and accurate approach that could detect and identify the faces of the individuals present in the frame using a Histogram of Oriented Gradients (HOG) and Convolutional Neural Networks. It has an advantage over other feature extractors as the HOG descriptor is invariant to the changes in geometric and photometric conditions. Histogram of Oriented Gradients (HOG) descriptors are collected from all blocks of a dense grid with fewer dimensions covering the Region of Interest (ROI) into a flattened combined feature vector from an image captured using a mobile device under uncontrolled environment. SVM classifier is then used to classify the faces, feature vectors of the extracted faces are then used to identify from the known faces present in the dataset using a combination of Deep Convolutional neural networks and SoftMax classifier. The identified subject is then marked present.

We evaluated the performance of the framework for the proposed challenging database named DTU Biometric Research Laboratory database. We also evaluated our model's performance on the traditional databases, At&t, Face94 & Face95. The paper has a major contribution in developing a fast and robust facial verification model that utilizes the hand-crafted features for detection and deep learned features for identification of faces.

The paper is organized as follows, Section 2 provides a review on the recent research conducted in the field of deep learning and face biometrics, a detailed description about the proposed efficient system is provided in section 3 and section 4 follows by results and discussion. At last, the conclusion is drawn in section 5.

2. RELATED WORK
The rapidly evolving field of Biometric systems has found its best use in multiple applications for the Government as well as private

BDET2020, January 3–5, 2020, Singapore.

ACM ISBN 978-1-4503-7683-9/20/01…$15.00
DOI: https://doi.org/10.1145/3378904.3378905

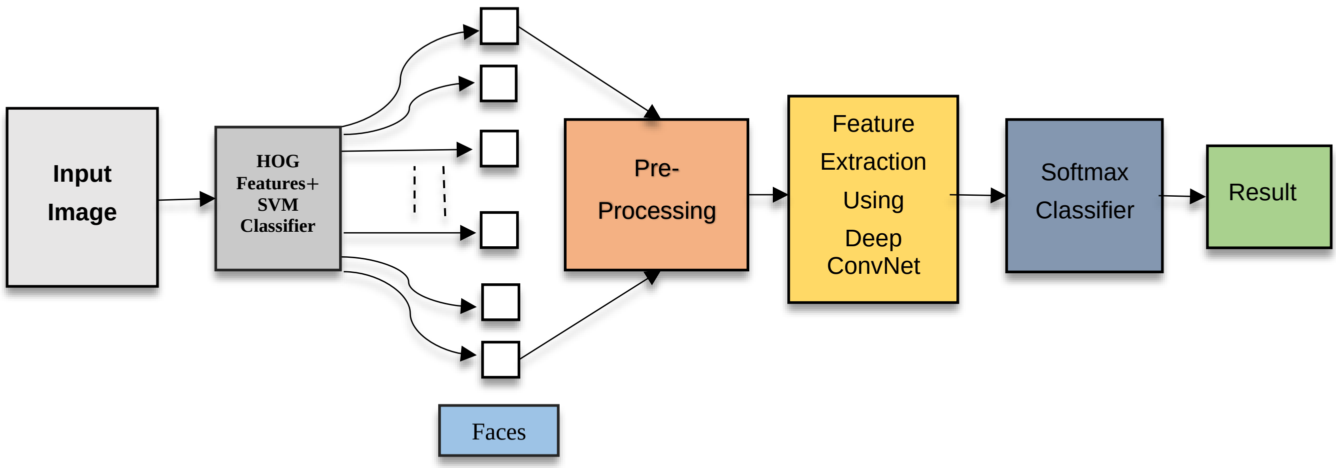

Figure 1. Block Diagram representation for the proposed Face Biometric system

organizations [5]. For this, Unique Behavioural and Physiological attributes such as hand geometry, fingerprint, iris, face, gait are responsible for the recognition of an individual. [6][7] Extensive research work has been conducted on Face Biometric systems. Face Recognition technology comes with an advantage over the other counterparts as facial features are found in all human beings and it makes no direct physical contact with the subject which is why it is used for surveillance [8][9].

Eigenfaces method for face identification given by M. Turk and A. Pentland requires uniform background and cannot deal with variations in pose and expression [10]. Paul Viola and Michael

Jones proposed Haar feature-based Cascade classifiers method that detects objects in an image [11]. However, they are rotation variant and are trained for a particular orientation, any change could lead to poor results [12]. CNN based face detection method outperforms all the state-of-the-art methods introduced by H. Li et al. [13]. Its robust method could detect faces in an unconstrained environment but due to the expensive computing cost of Neural Networks, it is not suitable to implement. Histogram of Oriented Gradients algorithm proposed by Dalal and Triggs [14] creates histograms of edge orientations from certain patches in images. The expression, color and other factors of the face may vary but the general edges around a face may remain relatively constant. This fast and robust technique revolutionized object detection. The handcrafted features for the purpose of face verification has been vastly explored such as Scale Invariant Feature Transformation (SIFT) which describe local features of the facial image, Local Binary Patterns (LBP) does the pixel labeling by considering threshold of its neighborhood pixel and the Support Vector Machines (SVMs) [15][16][17][18].

The idea of Deep learning was introduced earlier but it got its rebranding by G. Hinton et al. [19]. It could learn the categories through the hidden layer architecture incrementally. In 1980, Convolutional Neural Networks [20] was introduced, however, they were successfully implemented by LeCun for the task of handwritten digit recognition. It performs the important tasks of feature extraction followed by the classification. These deep learning models have been exploited for the purpose of object recognition tasks [21]. The use of Convolutional Neural Networks for face recognition has shown good results as it has higher accuracy over another classifier [25]. However, increasing the depth of layers in the architecture has resulted in improved identification rates but they are prone to the problem of overfitting and vanishing gradients. The introduction of the Dropout layer [22] and the Relu layer [23] solved this problem to some extent and hence improving the model's performance and speeding up the training process.

3. PROPOSED METHODOLOGY

The proposed Face Bio-metrics system framework is shown in the block diagram (Fig 1). Source image containing face in unconstrained environment is fed to the Bio-metric system. Dense feature extraction methodology of Histogram of Oriented Gradients is used for detecting face using an SVM classifier. This image is further divided into a number of blocks. A Block is composed of small regions known as cells. An individual cell may be contained in multiple blocks as the blocks overlap each other. Gradients represented by small arrows as shown in Figure 2. for each pixel.

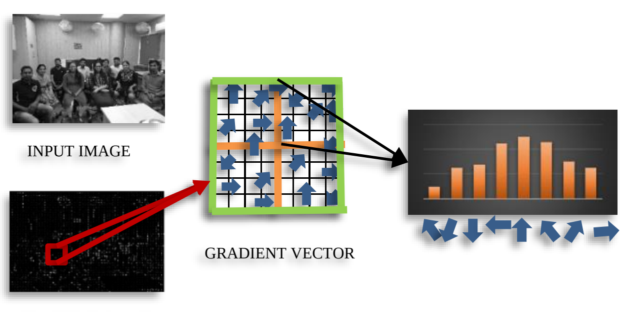

Figure 2. HOG feature extraction for face detection

Horizontal and vertical components of the gradients are calculated for the pixels, present inside the cell are calculated by using Eq. (1) and (2).

$$G_i(i,j) = I(i+1,j) - I(i-1,j) \quad (1)$$

$$G_j(i,j) = I(i,j+1) - I(i,j-1) \quad (2)$$

The magnitude and the direction of the gradient at coordinate (i,j) is given by Eq. (3) and Eq. (4).

$$G(i,j) = \sqrt{G_i(i,j)^2 + G_j(i,j)^2} \quad (3)$$

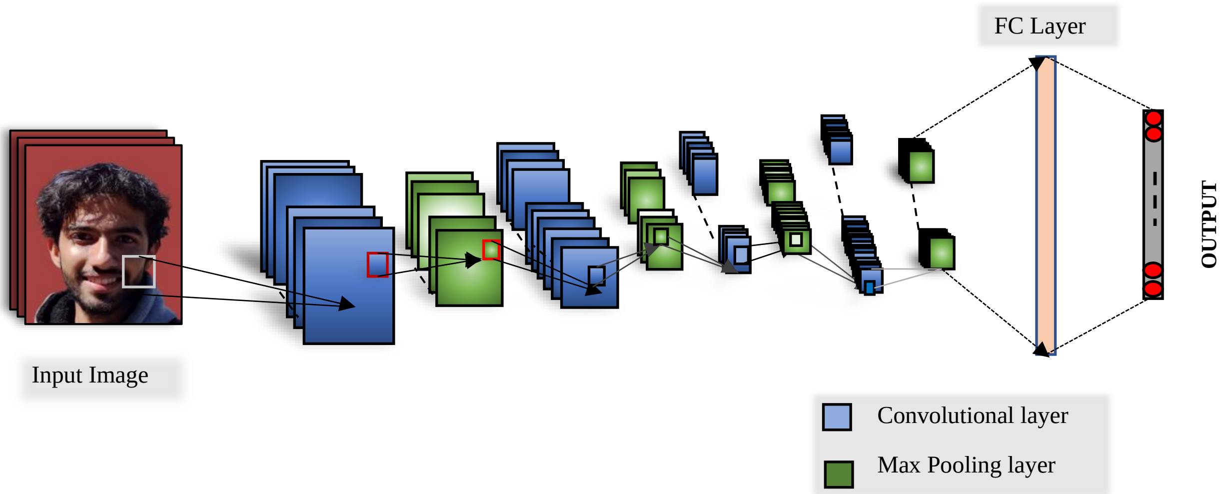

Figure 3. The proposed CNN architecture for the purpose of feature extraction

$$\theta = \tan^{-1}\frac{G_j(i,j)}{G_i(i,j)} \tag{4}$$

For each cell, a Histogram of oriented gradients is created as shown in the block diagram where the orientation is divided into Z bins. Further contrast normalization is performed on the histogram vector for each block using the L2 norm using Eq. (5).

$$L2\ norm,\ v_b' = \frac{v_b}{\sqrt{|v_b|_2^2+\epsilon^2}} \tag{5}$$

Here, v_b is the eigenvector within each block and ϵ is a small positive value.

Each detector window has a descriptor that consists of all the cell histograms contained in blocks. The descriptors are used by SVM (Support Vector Machines) classifier to classify faces in the image. After using pre-processing techniques such as normalization and mirror symmetry, the extracted faces from the images are further trained on a Convolutional Neural Networks (CNN) with a batch size of 32. The proposed Deep ConvNet architecture given in Figure 3. is composed of multiple layers in which the input layer takes an image of fixed dimensions and the output layer gives the prediction of classes. For Feature extraction, the CNN architecture introduced is composed of 8 layers containing Convolutional as well as Max pooling layers. The optimization of the objective function is done using Stochastic Gradient Descent due to its higher rate of convergence. It is a further modification of the traditional batch gradient descent algorithm. 0.01 is the default learning rate. The parameters are updated at each iteration of the training using the SGD optimizer as shown in Eq. (6) where x^irepresents training example and y^i , a label.

$$\emptyset := \emptyset - \eta\, \nabla_{\emptyset} J(\emptyset; x^i; y^i) \tag{6}$$

In the testing phase, 20% of the dataset is used. The faces from the captured images are detected and extracted using the HOG feature extractor and SVM classifier. The Deep ConvNet architecture then performs identification of the individual from the known faces in the Database. As the number of classes is more than 2, we use a multinomial logistic regression or Softmax classifier just before the output layer.

4. RESULTS AND DISCUSSION

In this section, the results are evaluated on four datasets namely the DTU Biometric Research Lab database, AT&T Database, Face94, and Face95 Database. After passing the images in the databases through HOG feature extractor, the input face image is resized to 10×10, 25×25, 50×50, 100×100 and 200×200 dimensions for the best results. The datasets are split into the ratio of 70:10:20 for training, validation and testing sets. The following datasets used for the experimental evaluation are described below.

4.1 At&t Database

Figure 4. AT&T Database

Earlier known by The ORL Database of faces, it was collected in the *AT&T* Laboratories at Cambridge. It is one of the most popular standard datasets. The faces of 40 individuals were captured between 1992 and 1994. 10 images of each subject having a uniform background with variable postures and uniformly dark background as shown below in Figure 4.

4.2 DTU Biometric Research Lab Database

For the purpose of challenging real-world datasets, the 10 images with variable expressions of 20 individuals were collected from Biometric Research Lab, DTU in an unconstrained environment in two different sessions. Each image has a resolution of 180×200 pixels. A few extracted Faces using HOG and SVM classifier collected from the dataset are shown in Figure. 5.

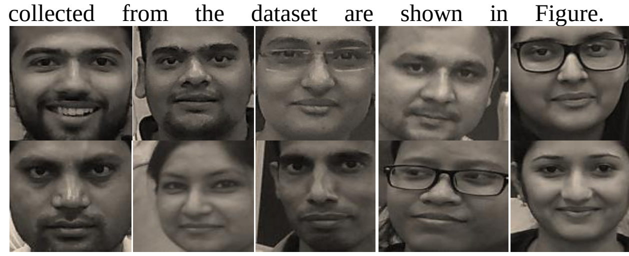

Figure 5. Extracted faces from the database using HOG feature descriptor and SVM classifier.

4.3 Face94 and Face95 (Essex) Database

Face94 [26] & Face95 [27] Databases contains images of 153 and 72 individuals respectively. Both the datasets are collected in one session with a difference of 0.5s between successive frames in a sequence. However, Face95 is more challenging as compared to the Face94 Database. Face95 shown in Figure 6(b) has a complex background with some translation of the face in the image while Face94 as shown in Figure 6(a) has a uniform background with no translation in face.

(a)

(b)

Figure. 6(a) Face94 Database (b) Face95 Database

The best performance was observed at 50×50 input size as shown in Table 1. for DTU Database while it was 25×25 for Face94 and Face95 Datasets and 200×200 for *AT&T* Database. The model gives a satisfying recognition rate. The curve for the evaluated verification accuracy with input image size of (128,128) is shown in Figure 7.

Table 1. Mean verification accuracy for the different datasets

Input Image Size	*DTU*	*AT&T*	*Face94*	*Face95*
(10,10)	72.83	82.3	97.84	98.06
(25,25)	92.39	93	98.37	98.93
(50,50)	94.57	93.5	98.24	96.53
(100,100)	92.39	95	97.91	96.67
(200,200)	90.22	95.5	97.71	95.14

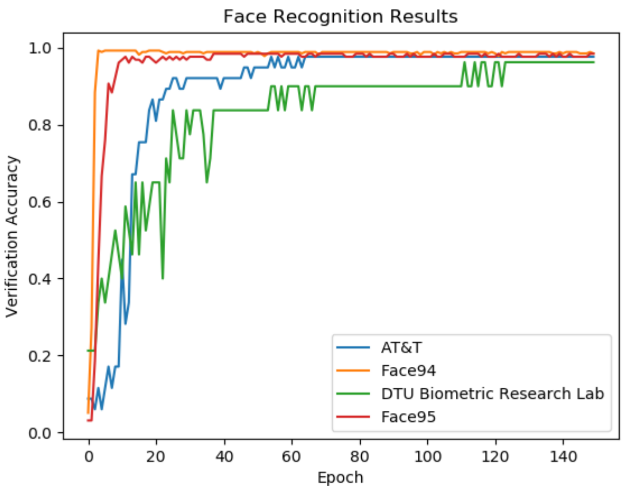

Figure 7. Curve on Verification accuracy over number of epochs.

Table 2. The proposed framework compared with the state-of-the-art methods for the following datasets

Methods	Features	*AT&T*	*Face94*	*Face95*
S´anchez D. et al. [28]	Modular Neural Network (MNN)	93.4	97.49	-
A. Mashhoori et al. [29]	((2D)2 PCA)	97.52	-	85.01
E. Hidayat et al. [30]	PCA	91.30	99.99	87.00
	LDA	94.40	99.99	90.80
Proposed Method	HOG+Deep ConvNet	95.5	98.37	98.93

A comparison of the propsed methodolgy with the previously given approaches is shown in Table 2. Our model performs better than other developed algorithms for the purpose of detection and identification of the known facial images.

5. CONCLUSIONS

The present paper proposes an efficient face biometric system based on Histogram of oriented Gradients and Convolutional Neural Networks that can work on real-time data. The Deep learning model attempts to learn and can interpret new incoming faces to the system that can handle illumination changes, various expressions, and poses.

In the near future, our aim would be to implement better architectures on a larger and complex dataset.

6. REFERENCES

[1] J.A.Unar, Woo C.S., Almas A., “A Review Of Biometric Technology Along With Trends And Prospects”, Pattern Recognition, vol.47, Issue 8, pp. 2673–2688, 2014.

[2] Zhao W, Chellappa R, Rosenfeld A, et al. Face recognition: A literature survey. ACM Comput. Surv., 2003, 35: 399-458.

[3] R Alferez and Y F Wang, "Geometric and Illumination Invariants for Object Recognition", IEEE Trans. on PAMI, Vol. 21, pp. 52-58, 1999.

[4] M. Pantic and I. Patras, "Dynamics of facial expression: Recognition of facial actions and their temporal segments from face profile," IEEE Transactions on Systems, Man, and Cybernetics,, vol. 36, no. 2, pp. 433-449, 2006.

[5] A K. Jain, AA Ross, K. Nandakumar, Introduction to Biometrics, Springer.

[6] A K. Jain, P. Flynn, A A Ross, Handbook of Biometrics, New York: Springer, 2010.

[7] Jain, A.K., Li, S.Z.: Handbook of face recognition. Volume 1. Springer (2005)

[8] A. Hadid, M. Pietikainen, "From still image to video-based face recognition: an experimental analysis", Automatic Face and Gesture Recognition 2004. Proceedings. Sixth IEEE International Conference on, pp. 813-818, 2004.

[9] L. An, M. Kafai, B. Bhanu, "Dynamic bayesian network for unconstrained face recognition in surveillance camera networks", IEEE Journal on Emerging and Selected Topics in Circuits and Systems, vol. 3, pp. 155-164, 2013.

[10] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on, volume 1, pages I– 511. IEEE, 2001.

[11] Mahdi Rezaei, Reinhard Klette, “Novel Adaptive Eye Detection and Tracking for Challenging Lighting Conditions”, Asian Conference on Computer Vision Workshops (ACCV), pp. 427-440, 2013.

[12] M. Turk, and A. Pentland, "Face Recognition Using Eigenfaces," Proc. IEEE Conf. on Computer Vision and Pattern Recognition, 1991, pp. 586-591.

[13] H. Li, Z. Lin, X. Shen, J. Brandt, G. Hua, "A convolutional neural network cascade for face detection", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5325-5334, 2015.

[14] Navneet Dalal and Bill Triggs, ?Histograms of Oriented Gradients for Human Detection, ? Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005.

[15] He K, Zhang X, Ren S, Sun J (2015) Delving deep into rectifiers: surpassing human-level performance on imageNet classification. In: Proceedings of the IEEE international conference on computer vision, pp 1026–1034

[16] . Sun Y, Chen Y,Wang X, Tang X (2014) Deep learning face representation by joint identificationverification. In: Advances in neural information processing systems, pp 1988–1996

[17] Kavukcuoglu K, Sermanet P, Boureau YL, Gregor K, Mathieu M, Cun YL (2010) Learning convolutional feature hierarchies for visual recognition. In: Advances in neural information processing systems, pp 1090–1098

[18] Phillips PJ, Flynn PJ, Scruggs T, Bowyer KW, Chang J, Hoffman K, Marques J, Min J, Worek W (2005) Overview of the face recognition grand challenge. In: 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR’05), vol 1, pp 947–954

[19] LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. Nature 521, 436–444 (2015)

[20] K. Fukushima, “Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position,” Biological cybernetics 36(4), pp. 193–202, 1980.

[21] C. Szegedy, A. Toshev, D. Erhan, "Deep neural networks for object detection", Proc. Adv. Neural Inf. Process. Syst., 2013.

[22] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," Journal of Machine Learning Research,, vol. 15, pp. 1920-1958, 2014.

[23] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzman machines," in International Conference on Machine Learning (ICML), Haifa, Israel, 2010.

[24] M.A. Hearst, "Support Vector Machines", IEEE Intelligent Systems, pp. 18-28, July/Aug. 1998.

[25] M. Coskun, A. Ucar, O. Yildirim, Y. Demir, "Face recognition based on convolutional neural network", International Conference on Modern Electrical and Energy Systems, pp. 376-379, 2017.

[26] Face Recognition Data, University of Essex, UK, Face 94, http://cswww.essex.ac.uk/mv/all faces/faces94.html

[27] Face Recognition Data, University of Essex, UK, Face 95, http://cswww.essex.ac.uk/mv/all%20faces/faces95.html

[28] S´anchez D, Melin P, Castillo O (2015) Optimization of modular granular neural networks using a hierarchical genetic algorithm based on the database complexity applied to human recognition. Information Sciences 309:73–101

[29] A. Mashhoori, M. Z. Jahromi, "Block-wise two-directional 2DPCA with ensemble learning for face recognition", *Neurocomputing*, vol. 108, no. 2, pp. 111-117, May 2013.

[30] E. Hidayat, F. Nur, A. Muda, C. Huoy, and S. Ahmad, “A comparative study of feature extraction using PCA and LDA for face recognition,” in Proc. 7th International Conf. on Information Assurance and Security, December 2011, pp. 354 - 359.