Convolutional Neural Network Based Multimodal Biometric Human Authentication using Face, Palm Veins and Fingerprint

Priti Shende, Yogesh H. Dandwate

Abstract: security access control systems and forensic applications. Performance of conventional unimodal biometric systems is generally suffered due to the noisy data, non universality and intolerable error rate. In propose system, multi layer Convolutional Neural Network (CNN) is applied to multimodal biometric human authentication using face, palm vein and fingerprints to increase the robustness of system. For the classification linear Support Vector Machine classifier is used. For the evaluation of system self developed face, palm vein and fingerprint database having 4,500 images are used. The performance of the system is evaluated on the basis of % recognition accuracy, and it shows significant improvement over the unimodal biometric system and existing multimodal systems.

Keywords: Convolutional Neural Network, Human authentication, Multimodal biometrics, Support Vector Machine

I. INTRODUCTION

Nowadays, biometric recognition is generally used in many modern human authentication systems such as criminal identification systems, forensic applications and secure access control systems etc. Biometric authentication is method of automatically recognizing the human being by using some computational algorithm with the aid of biometric features stored in database. The features are extracted for various internal or external biological parts of the human body such as face, fingerprint, eye retina, hand vein, finger vein, lips, palm prints, lips, iris, voice, dental radiographs and hand radiographs etc [1].

Biometric authentications are categorized in unimodal and multimodal biometric authentication systems. Unimodal biometric authentication system uses single biological entity which is less secure, less reliable and has inadequate usability [2]. Whereas, multimodal biometric authentication systems use multiple biological entities for the authentication which are more reliable, secure, accurate and robust. In multimodal biometric system the fusion of data can be done at sensor level, feature level, classifier level and rank level.

Priti Shende∗, Department of E&TC, Dr. D. Y. Patil Institute of Technology, Pimpri, Pune , India. Email:pritijawale30g@gmail.com
Yogesh H. Dandwate, Department of E&TC, Vishwakarma Institute of Technology, Pune, India.

Generally, biometrics systems are classified in static and dynamic biometric systems. Static features of the humans are rigid over the time such as fingerprint, iris, face etc.

While, dynamic features of the humans are changing over the time such as voice, electrocardiogram, keystroke and touch dynamics etc. Efficiency, accuracy, robustness, security and privacy are the major performance evaluation parameters for biometrics recognition systems [3]. The biometrics systems are often subjected to the impersonation and spoofing attacks which further degrades the performance of the recognition system. Therefore, there is need of secure, non-replicable, and non-spoofing systems to make person recognition systems robust and reliable [4][5].

In this paper, we proposed the convolutional neural network based multimodal biometric person authentication system which uses users face, palm vein and finger prints. In our work, three layers of convolutional neural network are used for hidden feature extraction for face, palm vein and finger print images. The fusion at feature level is done and for the classification multiple linear SVM is used.

The paper is organized as follow. Section II gives the information about previous work carried out on the multimodal biometric and convolutional neural network. Section III gives detailed description about proposed methodology. Further, section IV gives the experimental results and discussion. Last section V concludes the paper and explains the open background for the advancement in the proposed system

II. BACKGROUND

Most of the unimodal and multimodal biometric recognition generally uses the fingerprint, hand geometry, face, iris, signature, voice, gait, palm vein, ear and palm-print etc. Though the fingerprint based methods are popular, robustness of the system decreases due to age, dirty or scaly skin of finger [5]. Hand geometry based methods used edge, texture [6] or contour based methods [7][8]. Deployment of such a method is simple but it is quite expensive and cannot be applied to arthritic patient [9]. Various face recognition based systems have been implemented using 2-D or 3-D view of faces. Generally face recognition systems gives poor performance in occlusion [10][11]. Iris based recognition systems are highly expensive and are used in highly secured access control applications. Iris based systems are also used to detect the liveliness of the user [12]. Signature recognition can be online or offline [13] in nature. Signatures can be easily copied by the professionals.

Convolutional Neural Network Based Multimodal Biometric Human Authentication using Face, Palm Veins and Fingerprint

Voice recognition systems became more popular in recent years. Voice recognition systems are sensitive to illness and it can be easily spoofed using recorded voice [14]. Palm veins based recognition systems are based on the infrared images of palm vein and are also used for liveliness detection of user [15][16].

Thus, unimodal biometric human authentication systems are having challenges of noisy data, non universality, lack of individuality, vulnerability to circumvention and distinctiveness.

In past extensive work has been carried out on multimodal biometrics recognition systems. S. Singh et al [17] have presented the multimodal biometric recognition system using normalization of iris and face features in one domain. Ching H. C et al[18] have described the hidden markov model (HMM) system for fingerprint and online signature based person authentication. Finger print, hand geometry and voice based human authentication system have been proposed by Kowatko M. A et al which used sum rule based score level fusion of features [19]. B. E Manjunathswamy et al have implemented the multimodal biometric system using fingerprint and face which used z-score, tanh techniques for the normalization and max score method is used for the score level fusion [20]. Qing Zhang et al [21] have presented the multimodal biometric recognition system using semi-supervised techniques based on face and fingerprint. A finger multimodal biometric using three finger traits such as fingerprint, finger knuckle point and finger vein using graph fusion method [22].

Convolutional neural network has been identified as important attribute learner and classifier in many object recognition, objet detection and classification methods. Xinyi Zhou et.al [23] used faster R-CNN for the object detection which has been used for ImageNet, PascalVoc, and COCO dataset. Faster R-CNN is faster than R-CNN as convolution is done only once rather than giving larger region with convolution per region. Tianming Zao et.al [24] used the capsule convolutional neural network for the iris recognition which is robust and efficient. Zhangyu Wang et.al [25] presented the rail surface area detection using convolutional neural network. They have used dilated cascade sampling and dilated convolution. Dilated convolution is used for multi scale feature learning of rail surface using CNN architecture. Cascade sampling is combination of average pooling, maximum pooling and convolution to down sample the image. The key challenges of the multimodal biometric systems are selection of feature extraction method, larger complexity of the feature extraction and classification algorithms.

III. PROPOSED METHODOLOGY

In proposed methodology, three layers of CNN are applied to the face, finger print and palm vein images. To minimize the computation time original images are converted to the gray scale image from RBG color space. Fully connected layer is applied after third CNN layer. The neurons obtained at the fully connected layers of CNN applied for the face, palm vein and finger prints are concatenated for feature level fusion before the classification. The features are given to the one verses all multiclass linear SVM classifier. The performance of the system is evaluated on the basis of the % recognition rate. The flow diagram of the proposed methodology is

shown in figure 1. The convolutional neural network consists of basic four layers such as convolution layer, rectified linear unit layer (ReLU), max pooling layer and fully connected layer for futuristic security applications. The architecture of the convolutional neural network is shown in figure 2.

A) Convolutional Layer

The first layer of CNN is convolution layer in which the original image (I_m) is convolved with the filter kernel (F_K) as given in (1). This layer is also called as hidden feature extractor which describes the internal connectivity of the image region [25]. The dimensions of the filter kernel used for convolution are 5×5×6. To maintain the size of convolutional map the original image is zero padded on the border. The filter weights are adjusted using mini-batch gradient descent learning method.

$$I_{CONV} = I_m \otimes F_{k(5 \times 5 \times 6)} \qquad (1)$$

B) Rectified Linear Unit (ReLU) Layer

ReLU is applied after convolutional layer which uses nonlinear activation function (2) to minimize the linearity introduced in the convolutional layer and. In this layer the all the neurons with negative weights are forced to zero.

$$I_{RELU}(x, y) = \max (I_{CONV}(x, y), 0) \qquad (2)$$

C) Max Pooling Layer

Max pooling layer acts as non-linear down sampling method to reduce the number of neurons in the ReLU layer output. It divides the image in to the non-overlapping region of N×N pixel and consider the maximum value of the local region [29]. This layer minimizes the computation time and also control the overfitting.

D) Fully Connected Layer

In fully connected layer, neurons from different layers are connected in one layer. Most of the classifiers needed the data in one dimensional therefore multidimensional feature map is converted in to 1-dimensional vector.

E) CNN Learning

For the learning of CNN mini batch gradient descent optimization algorithm is used. In general Gradient descent algorithm the optimization over error takes place on entire set of training data which is computationally very expensive. In mini-batch gradient descent algorithm n number of training dataset samples are divided into small batches b, then the model coefficients are updated using model error.[26].

In the mini batch sum or average of gradient is chosen. Average of gradient used to reduce the variation of the gradient. Robustness of stochastic gradient descent and the effectiveness of batch gradient descent are combined in mini batch gradient method. Because of cheaper computation it is frequently used in deep learning algorithms. Total T=n/b number of iteration per training epoch. The weights (w) of CNN are optimized using error function defined in equation (3).

$$E_t[f(w)] = \frac{1}{b} \sum_{i=(t-1)b+1}^{tb} f(w, x_i) \qquad (3)$$

Where x_i is ith sample of training data. At each iteration the weights are updated by rule mini batch gradient descent update rule with learning rate μ given in equation (4).
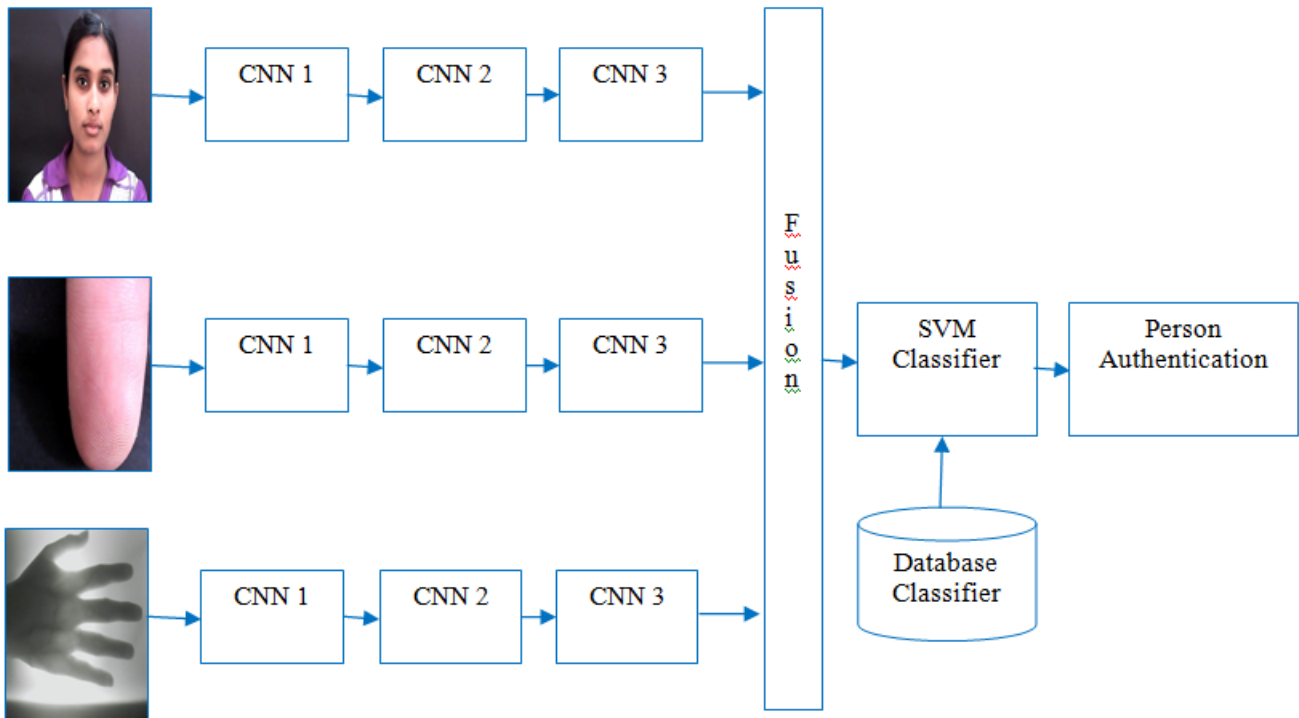
$$w^{t+1} = w^t - \mu \nabla_w E_t[f(w^t)] \quad (4)$$

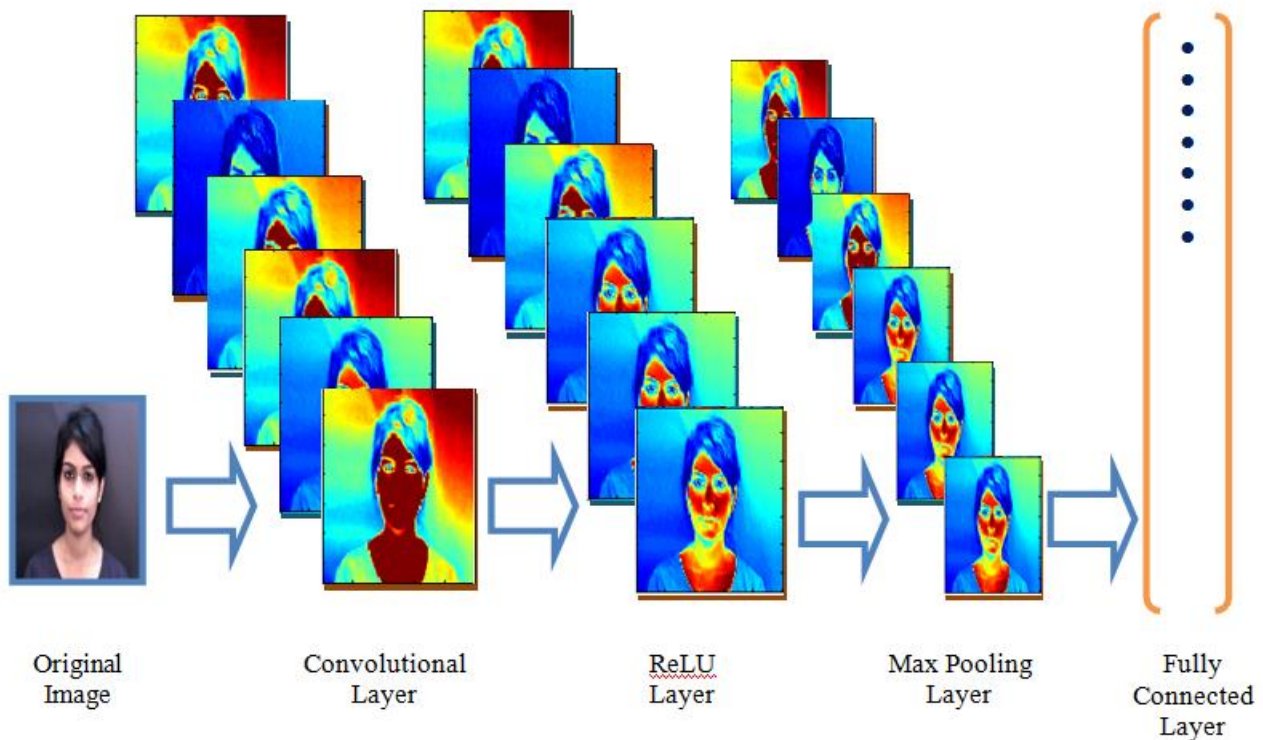**Figure 1 Flow diagram of the proposed system**

Figure 2 Architecture of convolutional neural network

F) Feature Level Fusion

Fusion of neurons obtained from the third layer of CNN applied to face, palm vein and fingerprints is performed using concatenation of all three types of features to increase the discriminant nature of the feature set as shown in equation (5).

$$Feat_{Final} = \lfloor Feat_{Face} \quad Feat_{Fingerprint} \quad Feat_{palmvein} \rfloor \quad (5)$$

G) Support Vector Machine (SVM) Classifier

Output of fully connected layer of CNN is given to support vector machine classifier for prediction of output class. SVM is supervised learning discriminative classifier [27][28]. While training multiclass SVM, one vs. all architecture is used in which the data is divided in one to other classes. The selected class is given as label +1 and for group of other classes label is assigned as -1 as given in (6,7). (x,y) is set of training data, x is the CNN feature set and y is class label. w is normal vector and b_0 is bias value.

$$\langle w.x \rangle + b_0 \geq 1 \ , \ \forall y = 1 \qquad (6)$$

$$\langle w.x \rangle + b_0 \geq -1, \ \forall y = -1 \qquad (7)$$

Linear kernel is used for the generation of separating hyper-planes as shown in equation (8).

IV. EXPERIMENTAL RESULTS & DISCUSSION

The proposed system is implemented using MATLAB software on the intel core i3-3110M processor with 2.64 GHz speed, 4 GB RAM, windows environment.

For the experimentation self developed database is used which consists of 1500 images of face, palm vein and fingerprint each. 70% of total image database is used for the training and 30 % database is used for the testing. The database is collected for the 100 users (50 male and 50 female) which belongs to different age group and profession.

The facial images are low resolution images captured for various face poses, alignments, expressions and at different intensity variations. Likewise, fingerprints and palm veins images are captures for different alignments. To maintain the uniformality in the database, all the images are resized to the 64×64 dimension. This system requires 25.3458 Hrs for the training of total 1050 face, palm vein and fingerprint training images. The performance of the system is evaluated on the basis of % recognition accuracy as given in equation (9).

$$\% \, Accuracy$$
$$= \frac{Number \ of \ Correctly \ \mathrm{Re}\, cognized \ Samples}{Total \ Number \ of \ Samples} X \, 100$$
$$(9)$$

The initial parameters considered for the implementation and training of convolutional neural network with mini batch stochastic gradient descent method are given in Table I. The feature maps for the different layers of CNN are given in Table II.

Table I. CNN initial parameters

Parameter	Specification
Filter Size	5×5
Number of Filter	6
Stride	[2 2]
Padding	[1 1]
Initial Bias	1
Weight Initialize Function	Glorat Initializer
Initial Learning Rate	0.01
GradientThreshold Method	L2 Norm
L2 Regulization	10^4
Gradient Threshold	Inf
Maximum Epoch	20
Mini Batch Size	64
Weight Learning Rate	0.02
Bias Learning Rate	0.02

The visualization of the all three layers feature map is shown in the figure 3. Increasing the number of CNN layers increase the number of neurons and connectivity maps. Thus, by increasing the number of layers multiple representations with connectivity information between local image regions is obtained. Fusion of CNN feature vector is formed by concatenation of face, palm vein and fingerprint to improve the discriminative nature of the feature vector as shown in figure 4.

Experimental results show that smaller filter kernel size for convolution layer gives poor performance because of inability to describe the large edges of the image region whereas larger window sized filters neglects the occlusion of the image region as shown in figure 5 (a). Therefore, 3x3 window is selected for the implementation which gives the proper correlation between the local image region. Increasing the pooling size decrease the image size by the pooling size factor. For 2x2 pooling the image decreases to the half of its original dimensions. Thus, increasing it beyond 3x3 decreases the local connectivity of the local image region which results in degradation of performance as shown in figure 5 (b). Number of filter kernel increases the number of neurons in each layer which increases the connectivity map or discriminant nature of the feature map as shown in figure 5 (c). For larger number of layer learning time if more and there is chance of under-fitting of the network. Thus for the implementation six number of filter kernels are selected. Striding the filter kernel by one pixel over the original image takes more time and fails to give connectivity between smaller edges whereas larger pixel striding neglects the fine changes in the local region as shown in figure 5 (d). Therefore, two pixel stride is considered for the convolution. For every parameter variations, three layered CNN gives better output than the single and dual layer CNN because increasing the number of CNN layers increases the feature variability. CNN has gradient vanishing problem if number of CNN layers are increased.

Table II. Feature maps for various CNN layers

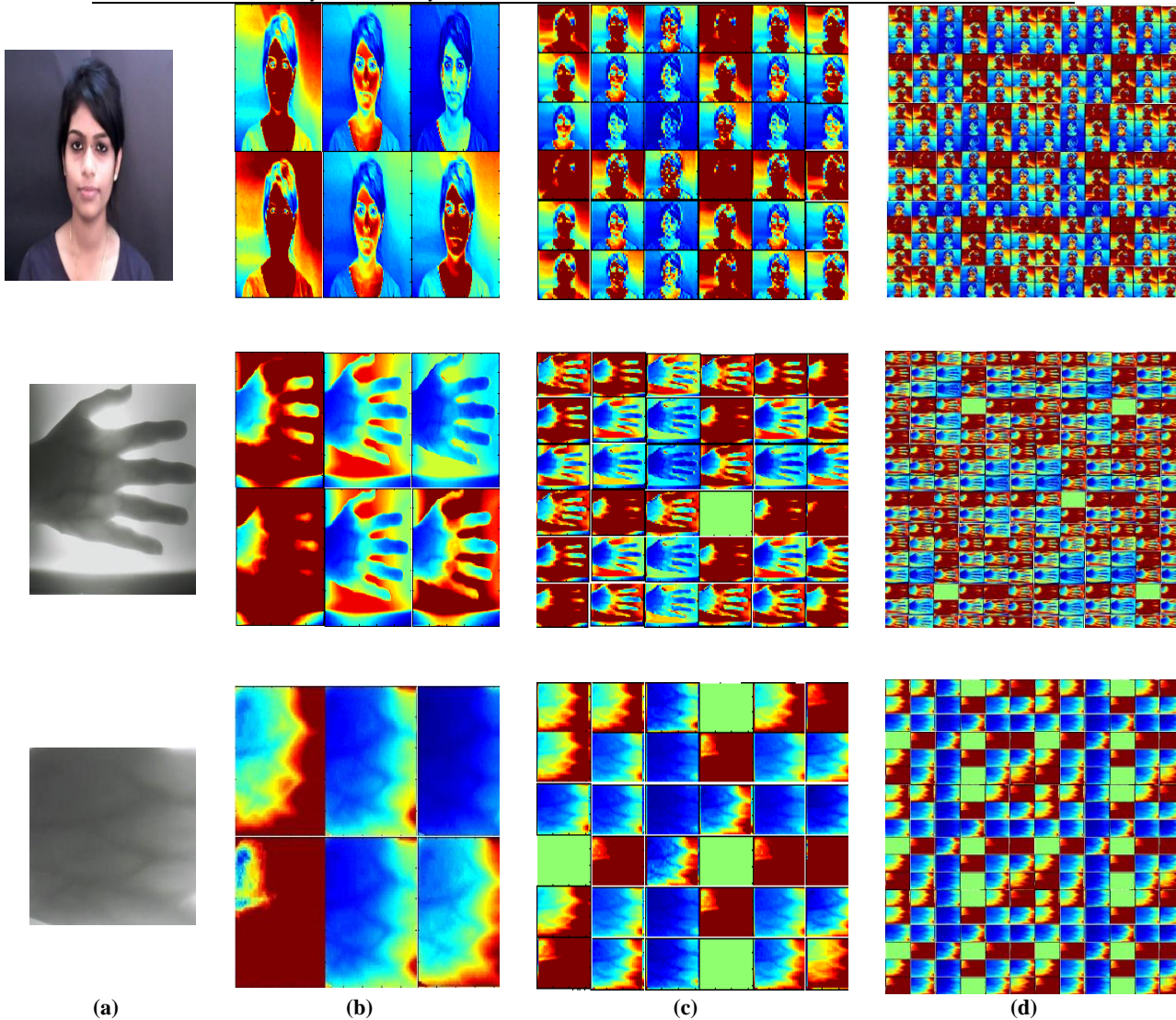| CNN layer	Sub Layer	Feature map size (Face)	Feature map size (Palm Vein)	Feature map size (Fingerprints)
CNN-1	Convolution Layer	$64 \times 64 \times 6$	$64 \times 64 \times 6$	$64 \times 64 \times 6$
	ReLU Layer	$64 \times 64 \times 6$	$64 \times 64 \times 6$	$64 \times 64 \times 6$
	Max Pooling Layer	$32 \times 32 \times 6$	$32 \times 32 \times 6$	$32 \times 32 \times 6$
CNN -2	Convolution Layer	$32 \times 32 \times 36$	$32 \times 32 \times 36$	$32 \times 32 \times 36$
	ReLU Layer	$32 \times 32 \times 36$	$32 \times 32 \times 36$	$32 \times 32 \times 36$
	Max Pooling Layer	$16 \times 16 \times 36$	$16 \times 16 \times 36$	$16 \times 16 \times 36$
CNN-3	Convolution Layer	$16 \times 16 \times 216$	$16 \times 16 \times 216$	$16 \times 16 \times 216$
	ReLU Layer	$16 \times 16 \times 216$	$16 \times 16 \times 216$	$16 \times 16 \times 216$
	Max Pooling Layer	$8 \times 8 \times 216$	$8 \times 8 \times 216$	$8 \times 8 \times 216$
	Fully Connected layer	1×13824	1×13824	1×13824

(a)   **(b)**   **(c)**   **(d)**

Figure 3 Visualization of different CNN feature maps a) Original Images b) CNN 1st layer outputs c) CNN 2nd layer outputs d) CNN 3rd layer outputs
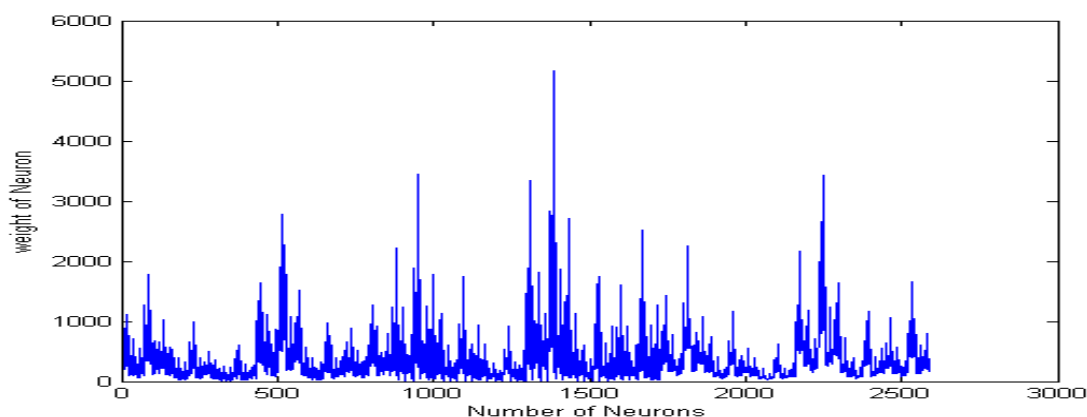
Figure 4 Representation of fusion of face, palm vein and fingerprint CNN feature vector
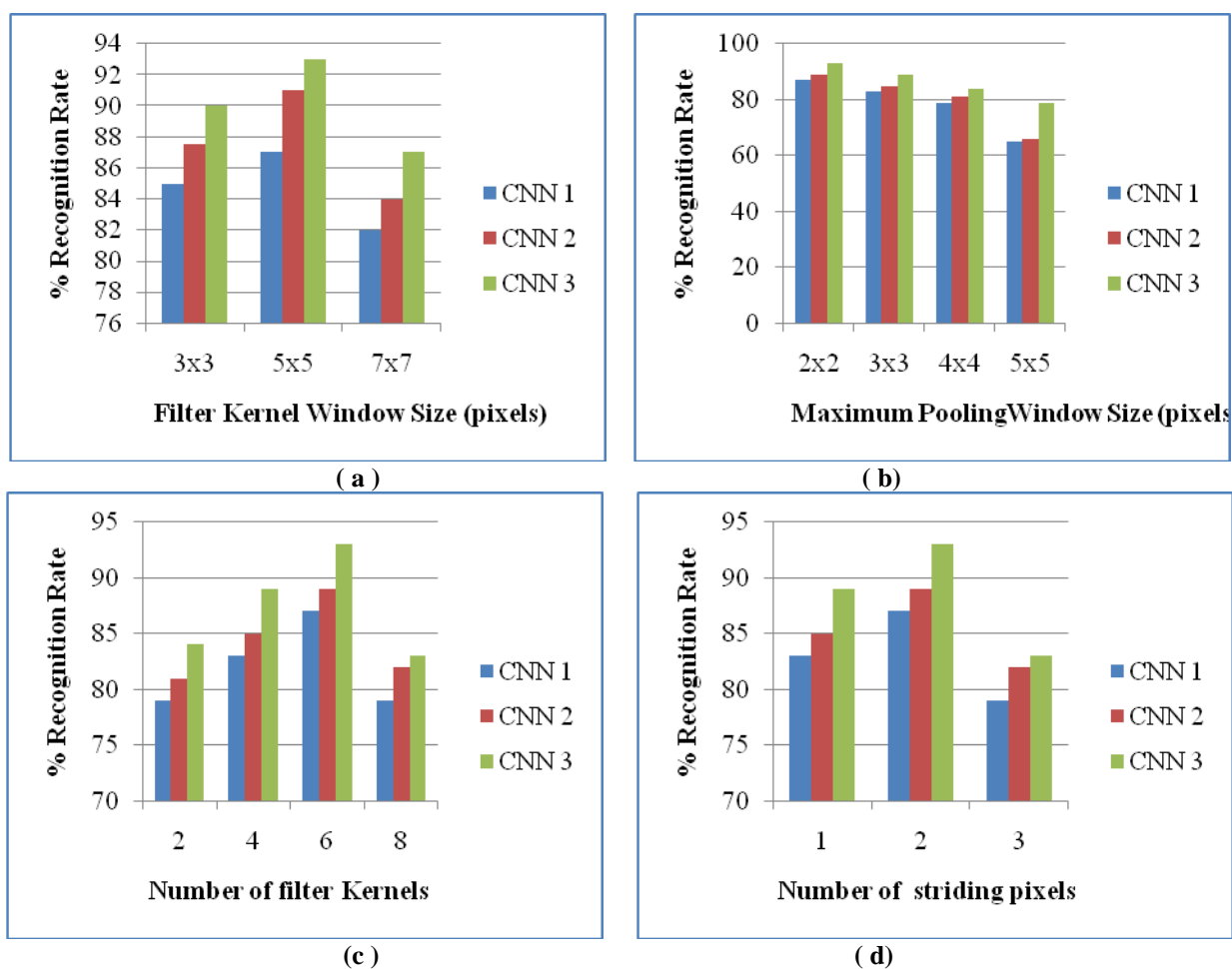

(a)

(b)

(c)

(d)

Figure 5 a) Graph of filter kernel window size vs recognition rate b) Graph of maximum pooling window size vs recognition rate c) Graph of number of filters vs recognition rate d) Graph of number of striding pixels vs recognition rate.

The performance of proposed algorithm is compared with the previous methods for the multimodal biometric human authentication using face, palm vein on the basis of % recognition rate and fingerprints and it outperforms the previous methods as shown in Table III.

Table III. Performance comparison of proposed system with existing methods

Methods	Recognition Rate (%)
Eigen Method (Face) [29]	89.33
HOG (Face) [30]	95.00
Gabor Transform (Fingerprint) [31]	85.00
Curvlet Transform (Palm Vein) [32]	90.00
CNN (Face Recognition) [33]	98.40
Proposed multi Layer CNN (Face)	98.00
Proposed multi Layer CNN (Palm Vein)	95.50
Proposed multi Layer CNN (Fingerprint)	97.00
Proposed multimodal multi Layer CNN (Face + Palm Vein + fingerprint)	**99.10**

V. CONCLUSION

Thus, three layer convolutional neural network based multimodal biometric human authentication system using face, palm vein and fingerprints shows the feature variability, robustness to the noise and universality. For the implementation of the system three layers of CNN using 3x3 filter kernel, 1pixel striding, 2x2 max pooling window size, and six filter kernels are used. The performance of the system is evaluated on self developed face, palm vein and fingerprint database on the basis of % recognition rate. Multimodal biometric approach shows the improvement in the recognition rate over the unimodal systems. Future scope of this work consists of minimization of the feature length using rule based fusion rule to minimize the complexity of the system.

REFERENCES

1. A. K. Jain, P. Flynn, A. Ross, "Handbook of Biometrics," Springer Publishing Company,2010.
2. Oloyede M. O and Hancke G. P, "Unimodal and Multimodal Biometric Sensing Systems: A Review," IEEE Access, vol. 4, pp. 7532-7555, 2006.
3. K. Tripathi, ``A comparative study of biometric technologies with reference to human interface," Int. Journal Comp. Appl., vol. 14, no. 5,2011, , pp. 10-15.
4. Zureik E and Hindle K, ``Governance, security and technology: The case of biometrics," Stud. Political Econ., vol. 73, no. 1, pp. 113-137, 2004.
5. Lati S and Solayappan N, ``A survey of unimodal biometric methods," Proc. Int. Conf. Secur. Manage, pp. 57-63, 2006.
6. Chien Le and Raj Jain, ``A survey of biometrics security systems," EEUU. Washington Univ. St. Louis, St. Louis, Mo, USA, Technical Report, pp. 1-10, 2009.
7. Kaur G, Yadav A. K, and Chaudhary S., ``An improved approach to multi-biometrics security," Int. J. Comput. Sci. Commun., vol. 5, no.1, pp. 181-187, 2014.
8. Yang F and Ma B., ``Notice of retraction two models multimodal biometric fusion based on fingerprint, palm-print and hand-geometry," in Proc. 1st Int. Conf. Bioinformat. Biomed. Eng. (ICBBE), 2007, pp. 498-501.
9. G. Kaur, A. K. Yadav, and S. Chaudhary, ``An improved approach to multibiometrics security," Int. J. Comput. Sci. Commun., vol. 5, no.1, 2014, pp. 181-187.
10. S. Liu, D. Yi, B. Li, and S. Z. Li, ``Face alignment under partial occlusion in near infrared images," in Proc. Chin. Conf. Pattern Recognit. (CCPR), 2010, pp. 1-5.
11. K. B. Bhangale, K. Jadhav and Y. Shirke, "Robust pose invariant face recognition using DCP and LBP', Int. J. Management, Tech. and Engg (IJMTES), Vol. 8, Issue 11, pp. 1026-1034, 2019.
12. Dong W, Sun Z, and Tan T, ``Iris matching based on personalized weight map," IEEE Trans. Pattern Anal. Mach. Intell., vol. 33, no. 9, pp. 1744-1757, 2011.
13. P. Porwik and T. Para, ``Some handwritten signature parameters in biometric recognition process," in Proc. 29th Int. Conf. Inf. Technol. Interfaces, (ITI), pp. 185-190, 2007.
14. K. B. Bhangale, P. Titre, R. Pawar and S. Bhavsar, " Synthetic speech spoofing detection using MFCC and radial basis function SVM," IOSR J. of Engg. (IOSRJEN), Vol. 8, Issue 6, pp. 55-62, 2018.
15. Athale S. S, Patil D, Deshpande P, Yogesh H. Dandawate, "Hardware Implementation of Palm Vein Biometric Modality for Access Control in Multi-layered Security System" Science direct, journal ,procedia ,Vol 58, 2015, Pages 492-498, 2015.
16. Shah G, Shirke S, Sawant S and Y.H. Dandawate " Palm Vein pattern based Biometric recognition system" Int. J. of Comp. Appl. in Tech., Inderscience , Vol.51, No 2 pp 105-111, 2015.
17. S. Singh, A. Gyaourova, G. Bebis, and I. Pavlidis, ``Infrared and visible image fusion for face recognition," in Proc. Defense Secur., vol. 5404, pp. 585-596, 2004.
18. H. C. Ching and C. Eswaran, ``A nature based fusion scheme for multimodal biometric person identity veri_cation at a distance," in Proc. ICSAP, pp. 94-97, 2009.
19. M. A. Kowtko, ``Biometric authentication for older adults," in Proc. IEEE Long Island Syst., Appl. Technol. Conf. (LISAT), pp. 1-6, 2014.
20. B. E. Manjunathswamy, J. Thriveni, and K. R. Venugopal, ``Bimodal biometric verification mechanism using fingerprint and face images(BBVMFF)," in Proc. IEEE 10th Int. Conf. Ind. Inf. Syst. (ICIIS), pp. 372-377, 2015.
21. Q. Zhang, Y. Yin, D. Zhan and J. Peng, "A Novel Serial Multimodal Biometrics Framework Based on Semisupervised Learning Techniques," in IEEE Trans. on Info. Forensic and Sec., vol. 9, no. 10, pp. 1681-1694, Oct. 2014.
22. H. Zhang, S. Li, Y. Shi and J. Yang, "Graph Fusion for Finger Multimodal Biometrics," in IEEE Access, vol. 7, pp. 28607-28615, 2019.
23. X. Zhou, W. Gong, W. Fu and F. Du, Application of deep learning in object detection, IEEE/ACIS 16th Int. Con. on Comp.and Info. Sci. (ICIS), Wuhan 2017;631-4.
24. T. Zhao, Y. Liu, G. Huo and X. Zhu, "A Deep Learning Iris Recognition Method Based on Capsule Network Architecture," IEEE Access, 7:49, pp. 691-11, 2019.
25. Z. Wang, X. Wu, G. Yu and M. Li, "Efficient Rail Area Detection Using Convolutional Neural Network," IEEE Access, 6:77, pp. 656-8, 2018.
26. D. Ito, T. Okamoto and S. Koakutsu, "A learning algorithm with a gradient normalization and a learning rate adaptation for the mini-batch type learning," 2017 56th Ann. Conf. of the Soc.of Instr. and Control Eng. of Japan (SICE), Kanazawa, 2017, pp. 811-816.
27. Bhangale K., and Shekokar R. "Human body detection in static images using hog & piecewise linear svm." Int. J. Inno. Research & Dev., Vol. 3. Issue. 6, 2014.
28. V. N. Vapnik, "An overview of statistical learning theory," IEEE Trans. on Neural Networks, vol. 10, no. 5, pp. 988-999, Sep 1999.
29. M. Turk and A. Pentland, "Eigenfaces for Recognition," Journal of Cognitive Neuroscience, vol. 3, no. 1, pp. 71–86, 1991.
30. Calvillo A.D., Vazquez R.A., Ambrosio J., Waltier A, "Face Recognition Using Histogram Oriented Gradients," Intell. Comp. Sys. ISICS 2016, Comm. in Comp. and Info. Sci., vol 597, 2016.
31. Munir, Muhammad Umer, and Muhammad Younas Javed. "Fingerprint matching using gabor filters." Nat. Conf. on Emerging Tech., Vol. 3. 2004.
32. Sun, Junwen & Abdulla, Waleed, "Palm vein recognition using curvelet transform," ACM Int. Conf. Proc. Series. 435-439, 2012.
33. Coskun, M., Ucar, A., Yildirim, O., & Demir, Y, "Face recognition based on convolutional neural network," Int. Conf. on Modern Elect.l and Energy Sys. (MEES), pp. 376-379, 2017.

Retrieval Number: C8467019320/2020©BEIESP
DOI: 10.35940/ijitee.C8467.019320
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

777