2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN) | 978-1-7281-8337-4/20/$31.00 ©2020 IEEE | DOI: 10.1109/ICACCCN51052.2020.9362764

Person identification in low resolution CCTV footage using deep learning

Sumantu Powale
Dept. of Electronics and Telecomm.
Don Bosco Institute of Technology
Kurla (W), Mumbai, India
ssp19298@gmail.com

Abhijeet Dhanawade
Dept. of Electronics and Telecomm.
Don Bosco Institute of Technology
Kurla (W), Mumbai, India
abhijeet02dhanawade@gmail.com

Siddhesh Bagwe
Dept. of Electronics and Telecomm.
Don Bosco Institute of Technology
Kurla (W), Mumbai, India
bagwesiddhesh02@gmail.com

Shreyas Kawale
Dept. of Electronics and Telecomm.
Don Bosco Institute of Technology
Kurla (W), Mumbai, India
shreyaspkawale@gmail.com

Nitin L. Chutke
Regional Forensic Science Laboratory
Solapur, Maharashtra, India
nitichutke@gmail.com

Satishkumar Chavan
Dept. of Electronics and Telecomm.
Don Bosco Institute of Technology
Kurla (W), Mumbai, India
satyachavan@yahoo.co.in

***Abstract*—The advancement in camera technology, high speed connectivity, and availability of huge information over media resulted into developing better algorithms for person identification. However, the recognition of a person in low resolution cameras like CCTV is still a challenging problem. This paper aims at identifying a person in low resolution image captured by webcam or in various frames of CCTV footage by using deep learning convolutional neural network (CNN). The proposed system uses facial image of a person for this task. The designed CNN pipeline has six convolution layers, one flattened layer and two fully connected layers. Total 6667 images of 62 subjects are used for training and validation of CNN framework with 500 epochs. The designed CNN attained 99.99% and 98.45% of training and validation accuracy, respectively. The network was tested on 1599 test images and was able to achieve a testing accuracy of 96.03%. The proposed CNN framework is also tested on low resolution face recognition benchmark dataset (TinyFace) for which it achieved 94.55% testing accuracy.**

Keywords— Face recognition; person identification; face detection; convolutional neural network; classification; low resolution CCTV footage; Haar cascade classifier.

I. INTRODUCTION

Biometrics has been much popular research area for reliable authentication of a person in computerized access control. Face recognition is one of the biometrics which is preferred tool for video surveillance and analysis with the goal of person identification. Face recognition is passive and non-invasive system used to recognize a person. It is treated as a difficult problem to be solved and high accuracy results are still awaited. This problem is modularized as face detection followed by face recognition. Conventional approaches in face recognition have used frontal faces (constraint), extracted features using various methods, and correlated these features or trained the neural network on these features to identify the best match of face in a database to authentic person. The development of robust automated face recognition systems has been challenging even though large number of algorithms have been proposed with high performance metrics.

The major hurdles in development of robust face detection and recognition system are situations where subjects are non-cooperative and the method of acquisition of face. Other factors such as posture of person, illumination of face, and image quality due to camera (resolution, optical zoom, noise, background, occlusion, etc.) also play significant role in deteriorating the performance of the face recognition system.

Due to the need of public safety, large number of surveillance cameras like close circuit television (CCTV) are installed in public area like supermarkets, banks, streets, stores, housing societies, etc. Obviously, authorities need automatic and robust face recognition system for law enforcement applications. The mounted CCTV camera is at longer distance from subjects which results in capture of smaller i.e. low resolution facial images. Even though, the better algorithms for person identification have been evolved, face recognition from such poor quality and smaller low resolution facial images is still a technological challenge. Person identification in such scenarios has numerous applications like surveillance in crowded area, identifying and locating criminals, authentication of evidences in CCTV footages, time attendance in premises, etc.

The work presented in this paper is primarily interested in identifying a person in image captured by webcam (low resolution image) or in the frames of CCTV footage using deep learning convolutional neural network (CNN) approach. Deep learning approaches have better capabilities to elevate the performance of face recognition work to a whole new level. It is a system that can detect and recognize a person in the CCTV footage with better accuracy.

The paper is organized as Section II gives background of the work in the area of face detection and recognition. Section III details the methodology adopted for the presented work.

ISBN: 978-1-7281-8337-4/20/$31.00 ©2020 IEEE

Section IV provides the experimental results and its analysis. Section V concludes the work presented in this paper.

II. Literature Survey

Widely used applications of face recognition are biometric authentication, automatic tagging friends in pictures uploaded on social media, surveillance, etc. [1]. Many approaches for image preprocessing, feature extraction and matching were proposed for face recognition systems. There were various attempts to improve face recognition accuracy in early 2000s using local feature based approaches. Local descriptor feature based systems were designed to achieve robust performance which was better than handcrafted feature learning. However, face recognition remained a challenge for complex data of faces.

Most popularly used feature extraction techniques were eigenfaces [2], active shape models [3], etc. Zou and Yuen presented super resolution method followed by Eigen face, kernel PCA and SVM for face recognition in very low resolution images [4]. Wang and Deng have given very comprehensive survey on current trends in face recognition with methodologies, face databases used, protocols followed, and various practical uses [5].

Face detection and recognition were treated as unstable performance systems with large number of false positives in real time scenario. However, this has changed since 2012 with first deep learning approach as DeepFace with AlexNet architecture [6]. In many computer vision applications, convolutional neural network (CNN) provided much better results. Therefore, researchers started developing approaches to solve this problem with the help of CNN models. Rasti et al. have improved facial image using super resolution method and applied CNN to achieve face recognition for surveillance videos [7].

Further, facial image recognition has achieved very high performance with DeepID networks like DeepID2 [8] and DeepID3 [9]. Large face database was used by Parkhi et al. for face recognition using VGG16 model [10] with adverse condition faces. Grm et al. [11] presented comparative performance evaluation of four CNN models for face verification namely AlexNet, VGG-Face, GoogLeNet, and SqueezeNet with face degradations. Lu et al. [12] developed Deep Coupled ResNet along with coupled mapping loss function for recognition of faces in varying resolution images. Other networks like Facenet [13], Openface [14], and GoogleNet [15] provided much promising recognition accuracies for face recognition.

However, all these networks were trained and tested for good quality facial images and achieved very excellent results. Better performance systems for face recognition with low resolution images are still awaited. This paper is an attempt to develop robust system with good performance for person identification based on faces captured using webcam and CCTV cameras.

III. Methodology

The proposed person identification system using face recognition consists of three stages namely database creation, design and training of CNN model, and classification and recognition of a person using trained model as given in Figure 1. The database is created by capturing facial images of subjects using low resolution webcam. The face is detected using Haar cascade classifier algorithm. The detected faces i.e. cropped images of faces are stored in the database. The CNN model is designed and trained on these images from database. The trained model is tested for its performance using separate test images from database as well as test images extracted from low resolution CCTV camera. The result image will have bounding box around face with the person's name (attribute) that is used during training.

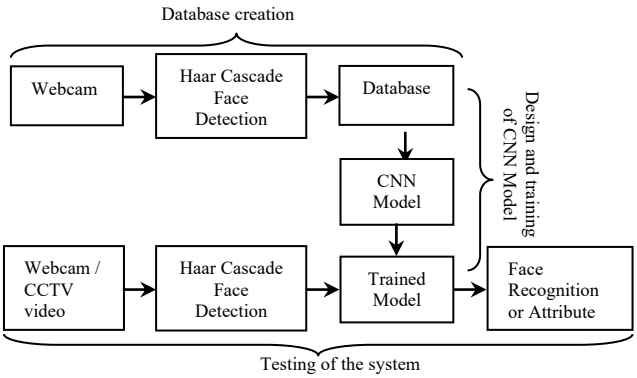

Fig. 1. Block schematic of proposed face recognition system

A. Dataset Creation

Database creation is an important task in the face detection and recognition system. In this work, Haar cascade classifier [16] is used for face detection in CCTV footage or live webcam video capturing process. Haar features like line and edge were extracted using various masks as shown in Figure 2 for faithful face detection [17]. Viola and Jones used the concept of cascade of classifiers with Adaboost optimizer for reducing the required number of features to be extracted for high performance face detection [16].

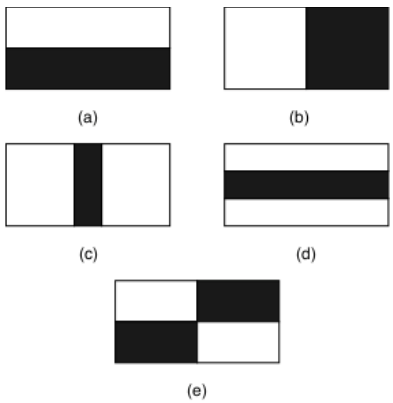

Fig. 2. Haar feature masks: (a) Bihorizontal, (b) Bivertical, (c) Trivertical, (d) Trihorizontal, (e) Diagonal

The dataset created is consistent with the cascade of classifiers method and consists of the photographs of the face of the subject itself. The subject was made to sit at an approximate distance of one meter away from the laptop. Total of 125 images of each subject are captured using

webcam, out of which 100 are used for training & validation and 25 images used for testing. Also 467 sample images of same 62 subjects are also taken from CCTV frames for training. For each subject, the time that the machine took to extract the facial portion of the images was almost 25-30 seconds. Every subject's image is labeled in the format as 'name.number' and each subject's dataset folder was labeled with its name. Figure 3 shows detected faces of one subject during the dataset creation using Haar cascade classifier.

Fig. 3. Sample database images of one subject using Haar cascade classifier

B. CNN6 Model

The data, which is given to machine learning methods to learn, is features of the objects to be classified. These features are extracted automatically or using hand-crafted techniques. It is proven that automatic extraction of features achieved excellent results compared to hand-crafted techniques in recognition or classification problem. This automatic generation of feature vectors through successive convolution layers is key to the success of deep learning algorithms in computer vision applications. This complex process makes network to learn at higher level of features for classification like human does [18]. Deep neural network is developed through large number of ways. Most widely preferred deep learning approach is a convolutional neural network which is based on framework of multilayer perceptron (MLP) [19].

CNN has neurons for processing of small patch of input image i.e. small field of view. It learns the pattern through these patches. These neurons are cascaded to make the structure more complex and effective for learning patterns. It has very large number of weights to learn. Most importantly, CNN learns explicitly with local pixel level correlation without using complete image.

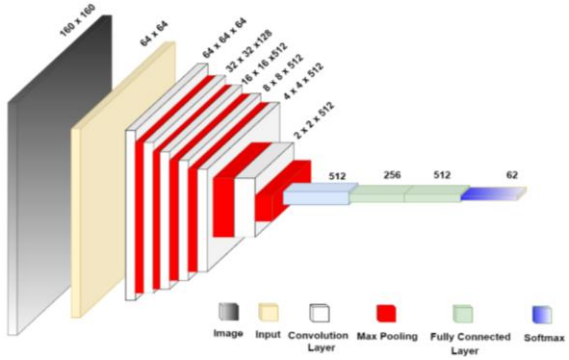

Fig. 4. Proposed face recognition CNN6 architecture

In presented deep learning CNN pipeline, the number of class labels was defined and the model chosen was sequential. The images used for training were of the size 160 × 160. The patches of size 64 × 64 were generated from the input image for the training of network. The proposed CNN consists of 6 convolutional layers and hence the network is named as CNN6. ReLu is the activation function used in this process [20]. Significant and relevant features were extracted using the max-pooling layer with a pooling size of 2 × 2. All of these form a single convolution layer. The convolution layers were followed by a single flattened layer and two fully connected layers. The architectural details of the CNN6 network are presented in the Figure 4 and 5.

The time taken by the machine to train the network was nearly 4-5 days for 500 epochs. The training accuracy of presented network after training was 99.99% with the parameters as shown in Table I.

TABLE I. PARAMETERS OF PROPOSED CNN6 MODEL

Parameters	Values
Input Size	64 × 64
Trainable Parameters	8, 177, 062
Loss Function	Sparse categorical cross entropy
Key feature	Low resolution dataset
Number of images per class	125
Optimizer	Adam Optimizer
Number of epochs	500

IV. RESULTS AND DISCUSSIONS

The facial images of 62 subjects were captured using low resolution webcam. A total of 125 images were taken of each subject which were split into 100 images for training & validation and 25 images for testing. The proposed convolutional neural network (CNN6) is trained as well as tested on these database images. We also tested the trained CNN6 model for images of low resolution CCTV video footages. The frames from the video were extracted and each frame was given as an input to the trained model to predict the detected face.

The presented system used facial image of a person for identification task. During testing phase 1, the person's image was captured using webcam. The face of a person is detected using Haar cascade classifier and extracted. The facial image is given to trained model for identification. The result image had detected face with bounding box around face with name of the subject.

In testing phase 2, CCTV video footage was given to the system; the frames from CCTV footage were extracted and applied as an input. Haar cascaded classifier detected face and this facial image was tested using trained CNN6 model. The recognized faces had bounded boxes around the faces with their names displayed in each frame. The system performed excellent in recognizing the person in low resolution CCTV footage. Sample recognized faces of 4 subjects with their names are shown in Figure 6.

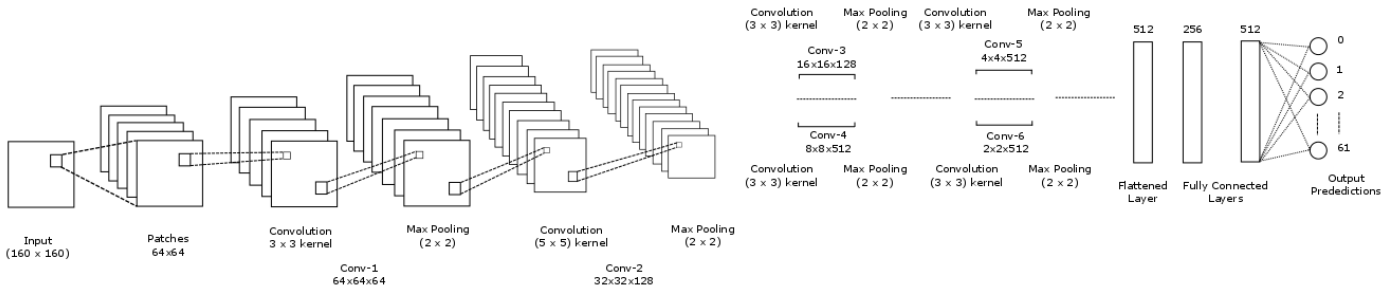

Fig. 5. Detailed layer-wise CNN6 pipeline

Fig. 6. Face recognition results

Total 6667 images of 62 subjects (6200 images from low resolution webcam and 467 images from CCTV footage) were used for training and validation of CNN6 framework with 500 epochs. The designed CNN attained 99.99% and 98.45% of training and validation accuracy, respectively. The network was tested on 1599 test images (webcam and CCTV footage) and was able to achieve a testing accuracy of 96.03%.

The proposed CNN6 pipeline is compared with VGG16 [10] and Local Haar Binary Pattern (LHBP) [21] as face recognition systems. The recognition accuracy achieved with CNN6 is higher than that of VGG16 and LHBP. The performance analysis of these networks is presented in the Table II.

TABLE II. PERFORMANCE ANALYSIS OF VARIOUS MODELS

Model	Training (%)	Validation (%)	Testing (%)
LHBP [21]	98.13	74.07	70.70
VGG16 [10]	97.21	96.47	90.89
Proposed CNN6	99.99	98.45	96.03

The proposed CNN6 model is also tested on TinyFace database [22] to evaluate its robustness for face recognition. We selected 20,556 images from 5139 classes randomly for training of the proposed model. The classification accuracy of the training is 96.85%. The classification accuracy achieved by CNN6 model for 5139 test images is 94.55% which is substantially satisfactory.

V. CONCLUSION

Face recognition has come a long way in the last twenty years with the increasing computational power and deep research in new technologies. In this paper, a person identification system for face recognition in low resolution frames of CCTV video footage is presented. We created our own dataset of 62 subjects in which the faces of a person are extracted using Haar cascade classifier. The designed CNN6 model is trained for 500 epochs and achieved classification accuracy of 96.03%. This face recognition system performed better than that of VGG16 and LHBP face recognition algorithms with higher accuracy. The proposed model has achieved 94.55% classification accuracy when tested on TinyFace benchmark database. The presented work can be used for surveillance and criminal identification in CCTV footages.

REFERENCES

[1] T. Sattler, B. Leibe, and L. Kobbelt, "Efficient & effective prioritized matching for large-scale image-based localization," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 9, pp. 1744–1756, 2016.

[2] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[3] L. Du, P. Jia, Z. Zhou, and D. Hu, "Human face shape classification method based on active shape model," *Journal of Computer Applications*, vol. 29, no. 10, 2009.

[4] W. W. Zou and P. C. Yuen, "Very low resolution face recognition problem," *IEEE Transactions on image processing*, vol. 21, no. 1, pp. 327–340, 2011.

[5] I. Masi, Y. Wu, T. Hassner, and P. Natarajan, "Deep face recognition: A survey," in *2018 31st SIBGRAPI conference on graphics, patterns and images (SIBGRAPI)*. IEEE, 2018, pp. 471–478.

[6] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 1701–1708.

[7] P. Rasti, T. Uiboupin, S. Escalera, and G. Anbarjafari, "Convolutional neural network super resolution for face recognition in surveillance monitoring," in *International conference on articulated motion and deformable objects*. Springer, 2016, pp. 175–184.

[8] Y. Sun, X. Wang, and X. Tang, "Deep learning face representation by joint identification and verification," in *Technical report*, 2014. [Online]. Available: arXiv:1406.4773

[9] Y. Sun, D. Liang, X. Wang, and X. Tang, "Deepid3: Face recognition with very deep neural networks," 2015. [Online]. Available:arXiv:1502.00873

[10] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *British Machine Vision Conference. British Machine Vision Association*, 2015.

[11] K. Grm, V. ˇ Struc, A. Artiges, M. Caron, and H. K. Ekenel, "Strengths and weaknesses of deep learning models for face recognition against image degradations," *IET Biometrics*, vol. 7, no. 1, pp. 81–89, 2017.

[12] Z. Lu, X. Jiang, and A. Kot, "Deep coupled resnet for low-resolution face recognition," *IEEE Signal Processing Letters*, vol. 25, no. 4, pp. 526–530, 2018.

[13] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 815–823.

[14] B. Amos, B. Ludwiczuk, M. Satyanarayanan et al., "Openface: A general-purpose face recognition library with mobile applications," *CMU School of Computer Science*, vol. 6, no. 2, 2016.

[15] R. Anand, T. Shanthi, M. Nithish, and S. Lakshman, "Face recognition and classification using googlenet architecture," in *Soft Computing for Problem Solving*. Springer, 2020, pp. 261–269.

[16] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001*, vol. 1. IEEE, 2001, pp. I–I.

[17] A. Roy and S. Marcel, "Haar local binary pattern feature for fast illumination invariant face detection," in *British Machine Vision Conference*, 2009.

[18] M.-I. Georgescu, R. T. Ionescu, and M. Popescu, "Local learning with deep and handcrafted features for facial expression recognition," *IEEE Access*, vol. 7, pp. 64 827–64 836, 2019.

[19] O. Rudenko, O. Bezsonov, and O. Romanyk, "Neural network time series prediction based on multilayer perceptron," *Development Management*, vol. 17, no. 1, pp. 23–34, 2019.

[20] X. Hu, P. Niu, J. Wang, and X. Zhang, "A dynamic rectified linear activation units," *IEEE Access*, vol. 7, pp. 180 409–180 416, 2019.

[21] R. Ji, P. Xu, H. Yao, Z. Zhang, X. Sun, and T. Liu, "Directional correlation analysis of local haar binary pattern for text detection," in *2008 IEEE International Conference on Multimedia and Expo*. IEEE, 2008, pp. 885–888.

[22] Z. Cheng, X. Zhu, and S. Gong, "Low-resolution face recognition," in Asian Conference on Computer Vision. Springer, 2018, pp. 605–621.