ISSN 2278-3091
Volume 8, No.3, May - June 2019
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse06832019.pdf
https://doi.org/10.30534/ijatcse/2019/06832019

Development of a Face Recognition System Using Deep Convolutional Neural Network in a Multi-view Vision Environment

Jay Robert B. Del Rosario
Electronics and Communications Engineering Department, De La Salle University - Manila
2401 Taft Ave, Malate, Manila, 1004 Metro Manila, Philippines
jay.delrosario@dlsu.edu.ph

ABSTRACT

This paper presents the implementation of Face Recognition System for multi-view vision system consisting of three cameras. It captures input frames from RLC423 camera; the main structure composed of recognizing faces, embeddings computation and database vector comparison. OpenCV, Python dlib, face recognition or Openface and ResNet34 are utilized for this face recognition method. It illustrates the application of face recognition and compare it to different dataset with static and dynamic approaches such as MIT-CBCL face recognition dataset, Caltech faces and EPFL-RLC Multi-Camera Dataset which, obtained 1.5% training loss and 90.3% model accuracy.

Key words : face recognition; multi-view vision; neural network; quadrotor

1. INTRODUCTION

Face recognition is the way toward taking a face in a picture and really distinguishing whom the face appertains to. Face recognition is in this manner a type of individual recognizable proof [1]. Early face recognition frameworks were dependent on facial points of interest extricated from pictures, for example, the relative position and size of the eyes, nose, cheekbone, and jaw. In any case, these frameworks were regularly profoundly emotional and inclined to blunder since the PC researchers and managers running the face acknowledgment programming physically removed these measurements of the face.

A multi-vision environment uses more than one camera that acts as a sensor to capture the same scenes or a completely different view and the captured data are processed together. Through this, image mapping can be accomplished easier and faster. A system of multi-view can have two set-ups. First, cameras can be set right beside each other and second is to position the cameras with distance and with different angles.

Multi-camera calibration maps varying view in a single coordinate system. The topology of a camera network is important for transition time of objects. Another technology is object re-identification where two image regions from different views are compared and matched in order to identify if they belong to the same object [2][3]. Multi-camera tracking traces objects in the different camera views. Multi-camera activity analysis is the fusion of information from different cameras to recognize activities and detect abnormal activities.

This research will acquire data from multiple cameras thus performing a robust face recognition in a multi-view vision environment with adaptive control in multiple platform-based systems.

2. RELATED WORKS

There are various attempts and approaches dealing with face recognition since 1960's [4]. Most commonly, articles are based on searching the major features of the face to a training set.

Face recognition using Eigenfaces [5][6][7][8] recommended an appearance-based approach on their viola jones algorithm. Sirovich and Kirby [9], which produce a set of eigenvectors for face recognition by means of Principal Component Analysis, developed Eigen-faces algorithm.

The Eigenface technique aims to cut back the spatiality of the initial image area by exploitation Principal Component Analysis (PCA) to pick out a brand new set of unrelated variables. The aim is to decide on these new variables in such the way that retains the maximum amount of the variation as possible from the initial set of variables that outline the initial image area. This objective is then comparable to finding the principal parts of the image area. These principal parts, conjointly known as Eigenfaces, is thought of as a group of feature vectors that represent the characteristic options among the face set and along this set of Eigenfaces characterize the variation between face pictures [9]. However, this method is sensitive to lighting and scale; and requires a highly controlled environment which is difficult for real-time application.

Murtaza et. Al proposed a method to combine the Adaptive Margin Fisher's Criterion and Linear Discriminant Analysis (LDA) [10]. LDA is a methodology employed in pattern recognition and machine learning to search out a linear combination of options that characterizes or separates two or many categories of objects or events. The ensuing combination could also be used as a linear classifier or, a lot of ordinarily, for spatiality reduction before classification. A major challenge using LDA is that the high computational cost. Several researchers exploit each PCA and LDA to cut back their quality. They applied the two-way linear low dimensional feature classifiers for this purpose however still each technique has their own difficulty.

Olszewska introduces some of the main challenges in automatic face recognition; narrating usual solutions based on machine learning methods. These challenges are pose variations, presence or absence of structuring elements or occlusions, facial expression changes, ageing of the face, varying illumination conditions, image resolution and modality, and availability and quality of face datasets [11]. The author also cited some methods to address these challenges: Eigenfaces, Active Appearance Models, Local Binary Patterns and SIFT.

Another algorithm in face recognition was proposed by Siegel-man and Vapnik [12][13]. SVM finds the hyperplane that separates the biggest viable fraction of factors of the same category on the same side, while maximizing the distance from either type to the hyperplane [13].

Another fine-tuning algorithm in supervised learning is Support Vector Machines (SVM). Face Recognition by Dong et. al [14] combined the use of SVM and KPCA to extract the face contours and curve details for successful recognition conducted in ORL and Yale. However, this method is accurate on small samples only in performing face recognition [14][15][16]. either SI (MKS) or CGS as primary units. (SI units are strongly encouraged.) English units may be used as secondary

3. METHODOLOGY

The project will follow the hierarchy as seen in Figure 1 for the face recognition system.

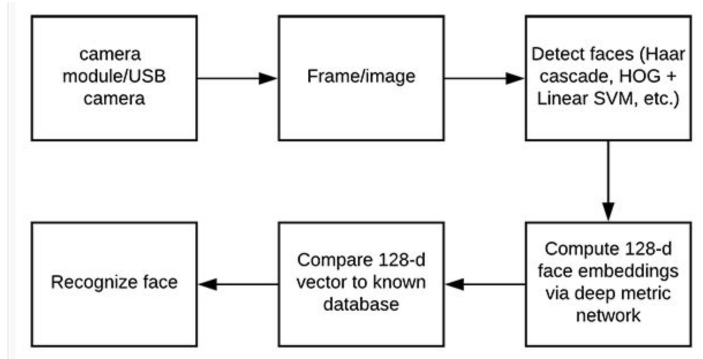

Figure 1: Face Recognition System

Beginning with capturing input frames from RLC423, the system comprised of face recognition through Haar cascades, HOG with Linear SVM, followed by computation of 12-d face matrix embedding then compare it to the existing face database. OpenCV, Python dlib, and face recognition or Openface are required for this face recognition method.

3.1 Data Events

There are 3 datasets included in the research: Caltech Faces [17], MIT-CBCL face recognition database [18], and EPFL-RLC Multi Camera Dataset [19] [20].

Table 1: Research Domain

Dataset	Samples Collected	Number of Person	Image Size
MIT-CBCL Face Recognition	200 images/subject	10 subjects	220x220 pixels
Caltech Faces	450 frontal face images	27 subjects	896x592 pixels
EPFL-RLC Multi-Camera Dataset	8000 from 3 camera views	>10 subjects	1920x1080 pixels

Figure 3 shows how the researchers adopted new dataset for approximately 150 samples with 6 persons for 250x250 pixels.

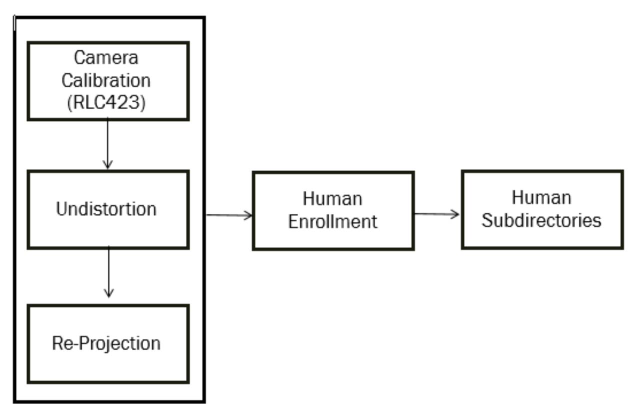

Figure 2: Human Enrollment Block Diagram

The first block of the diagram above entails the camera calibration method for image refinement using undistortion and re-projection [21]. Undistortion technique will return the image with little unwanted pixels while re-projection error gives the mean error and estimation on how accurate the parameters given the distortion and rotation matrices. Human enrollment is a manual gathering of human images or face images from different persons. Human subdirectories will separate the unique person to other dataset.

3.2 Image Pre-Processing

The noise reduction method would primarily deal with Gaussian noise, and salt and pepper noise due to these being the most common noises. As such the main methods used would be Gaussian blur for Gaussian noise, where the frame is convolved with a Gaussian matrix [22]. The median filter block will also be used to eliminate the black and white speckles in the image, following proposed median filter algorithm:

1. Determine the size of the image
2. Sort the pixels into ascending order by odd-neighborhood and/or even- neighborhood pixel
3. Select the value of the middle pixel as the new value for pixel [i, j].

The project utilized the thresholding technique to identify the sub images that represent the object for detection and surveillance. The partitioning of an image is also known as segmentation.

After determining the sub-images or objects during the initial phase, geometric properties will focus on the following: size, position and orientation. To further identify the image, the research investigated different morphological technique to properly structure the input image such as dilation, erosion, opening and closing.

3.3 Multi View Vision System (MVVS)

A multi-vision environment of the project incorporates three cameras as seen in Figure 3 that acts as instrument to capture the similar scenes or completely varying views then they are administered together [23].

MVVS uses multiple cameras that are directed in the identical field or in distinctive location, where the views can be processed together, photo mapping becomes a lot less difficult and faster. The cameras no longer always want to be subsequent to each other, multi-view purposes can be developed with the cameras far from each other. Changes in viewpoints of varying vehicles and their particular occlusions are two challenges in multi-view vision system [24].

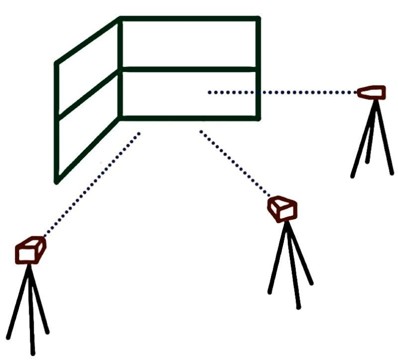

Figure 3: MVVS Set-up

3.4 Face Recognition

The project includes the pre-trained ResNEt34 model for dlib facial recognition network depicted in Figures 4 and 5, which outputs 128 real valued numbers. In Figure 6, ResNet34 architecture [25] introduced a concept called “skip connections”. Typically, the input matrix calculates in two linear transformations with ReLU activation function.

Dlib is a machine learning toolkit that contains Basic Linear Algebra Subprograms (BLAS Support). It includes anomaly detection, Bayesian network, clustering and feature ranking [26]

Facial recognition using ResNet34 encompasses a “triplet” procedure as seen in Figure 4 and 5. The triplet contains three distinct face images — two of the three are the same person. The neural network outputs a 128-d vector for each image.

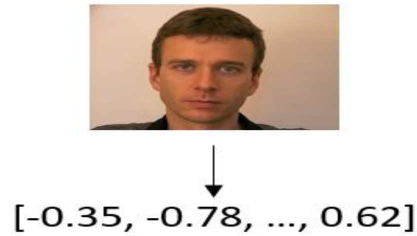

Figure 4: 128d real-valued number generated for each image

In order to represent faces numerically, we quantify all faces in the dataset with a 128-d feature vector generated by a neural network.

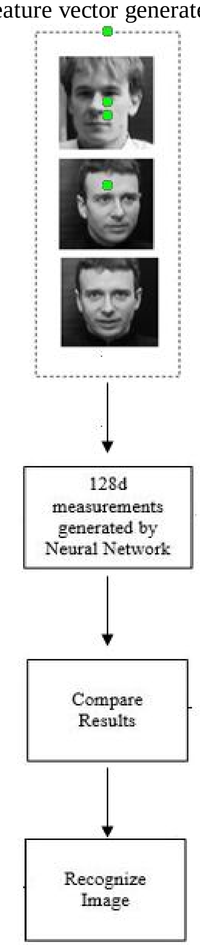

Figure 5: 128D Representation Algorithm

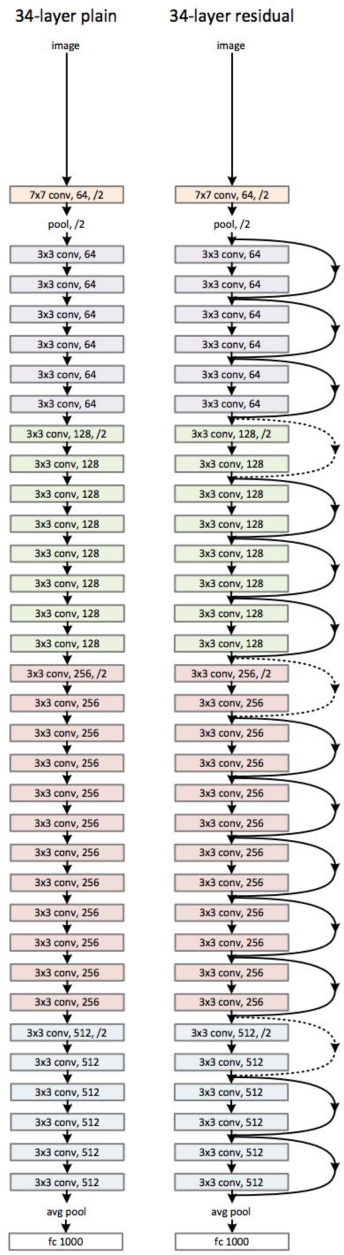

Figure 6: ResNet34 Architecture [10]

ResNet34 or Residual Network is a deep convolutional neural network for image classification that aims to resolve dificulties in recognition accuracy. It won 1st place in ILSVRC 2015 with 3.57% error rate; replacing VGG-16 layers in Faster R-CNN [25].

4. RESULTS AND DISCUSSION

Figures underneath demonstrate the rate error in face acknowledgment with 90 percent or more certainty level. As observed, there are still a few mistakes in the model since there are occasions that influence the outcome, for example, shadows from the dividers and people that obstruct the view from the camera and other dim regions making it harder for the neural system to prepare the face precisely.

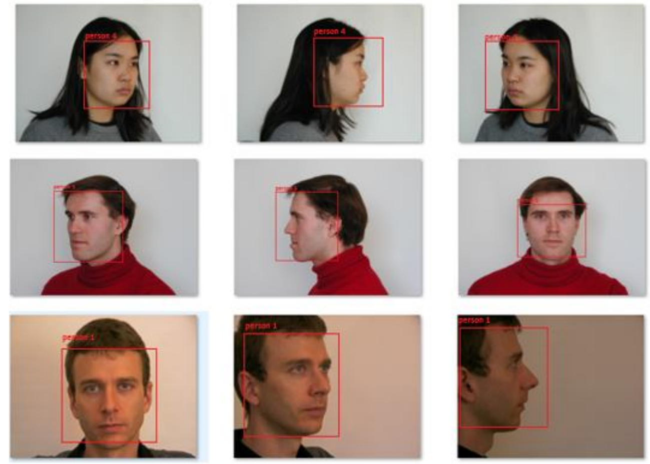

Figure 7: Sample Frames

After 100 epochs, it can be seen in Figures 8 and 9 that the model has obtained a high accuracy on recognition tasks. The proposed algorithm has obtained 1.5% training loss and 90.3% model accuracy.

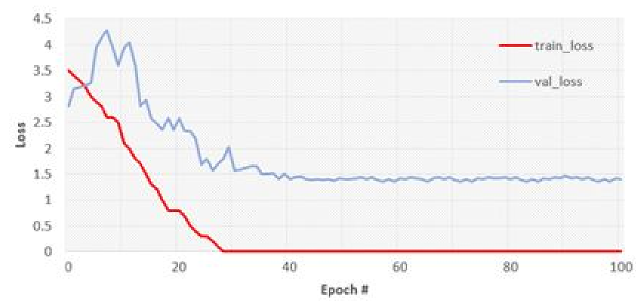

Figure 8: Training Loss of the System

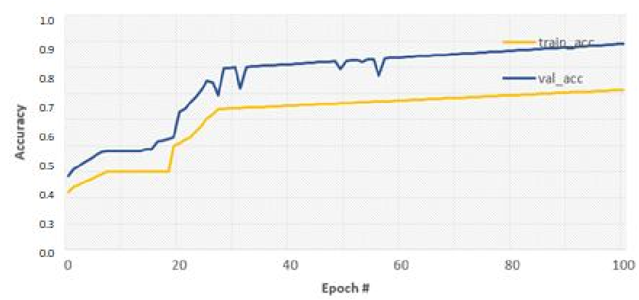

Figure 9: Model Accuracy

5. CONCLUSION

The installation of surveillance cameras is said to be important in the aspects of guaranteeing safety and safeguarding belongings; potential criminal activities could be avoided when individuals know they are being watched. Furthermore, tracking a specific individual's whereabouts within a vicinity by sifting through numerous CCTV footages is a tedious job and would take a while before the rightful action towards it is taken.

This research aims to implement a system of three security cameras connected as a network. These cameras are to be considered smart as they could communicate and transmit data such as images of a specific individual from one camera to another. It could also operate using Raspberry Pi and allows a real-time response for the user. The software, OpenCV, is utilized as well to perform human detection and recognition. This research hopes to address the problem of tracking specific individuals in an area using security cameras.

ACKNOWLEDGMENT

The authors would like to thank the University Research Coordination Office (URCO) of De La Salle University-Taft, Manila and Department of Science and Technology-Engineering Research and Development for Technology (DOST- ERDT).

REFERENCES

[1] Ahonen, T., Hadid, A., & Pietikainen, M. (2006). Face description with local binary patterns: Application to face recognition. IEEE Transactions on Pattern Analysis & Machine Intelligence, (12), 2037-2041. https://doi.org/10.1109/TPAMI.2006.244

[2] Krumm, J., Harris, S., Meyers, B., Brumitt, B., Hale, M., & Shafer, S. (2000). Multi-camera multi-person tracking for easy living. In Visual Surveillance, 2000. Proceedings. Third IEEE International Workshop on (pp. 3-10). IEEE. https://doi.org/10.1109/VS.2000.856852

[3] Knudsen, T. (2017). U.S. Patent No. 9,688,200. Washington, DC: U.S. Patent and Trademark Office.

[4] Fagan, J. F. (2017). The origins of facial pattern recognition. In *Psychological development from infancy* (pp. 83-113). Routledge. https://doi.org/10.4324/9781315163130-4

[5] Yin, D. B. M., Kamal, M. I., Azmanuddin, N. S., Ali, S. H. S., Othman, A. T., & Wan-Chik, R. Z. (2016, January). Electronic Door Access Control using MyAccess Two-Factor Authentication Scheme featuring Near-Field Communication and Eigenface-based Face Recognition using Principal Component Analysis. In *IMCOM* (pp. 1-1).

[6] Imran, M. A., Miah, M. S. U., Rahman, H., Bhowmik, A., & Karmaker, D. (2015). Face recognition using eigenfaces. *International Journal of Computer Applications*, *118*(5). https://doi.org/10.5120/20740-3119

[7] Kshirsagar, V. P., Baviskar, M. R., & Gaikwad, M. E. (2011, March). Face recognition using Eigenfaces. In *Computer Research and Development (ICCRD), 2011 3rd International Conference on* (Vol. 2, pp. 302-306). IEEE.https://doi.org/10.1109/ICCRD.2011.5764137

[8] Saha, R., & Bhattacharjee, D. (2013). Face recognition using eigenfaces. *International Journal of Emerging Technology and Advanced Engineering*, *3*(5), 90-93.

[9] Lee-Morrison, L. (2015). Eigenface: The image and the machine. In *The Berkeley Conference on Precarious Aesthetics*.

[10] Murtaza, M., Sharif, M., Raza, M., & Shah, J. (2014). Face recognition using adaptive margin fisher's criterion and linear discriminant analysis. *International Arab Journal of Information Technology*, *11*(2), 1-11.

[11] Olszewska, J. I. (2016). Automated face recognition: challenges and solutions. In *Pattern Recognition-Analysis and Applications*. InTech. https://doi.org/10.5772/66013

[12] Guo, G., Li, S. Z., & Chan, K. (2000). Face recognition by support vector machines. In *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on* (pp. 196-201). IEEE.

[13] V. N. Vapnik. Statistical learning theory. John Wiley & Sons, New York, 1998.

[14] Jianhua Dong, Jason Gu, Xin Ma and Yibin Li, "Face recognition based on KPCA and SVM," *Proceeding of the 11th World Congress on Intelligent Control and Automation*, Shenyang, 2014, pp. 1439-1444. https://doi.org/10.1109/WCICA.2014.7052930

[15] Zhao Lihong, Song Ying, Zhu Yushi, Zhang Cheng and Zheng Yi, "Face recognition based on multi-class SVM," *2009 Chinese Control and Decision Conference*, Guilin, 2009, pp. 5871-5873. https://doi.org/10.1109/CCDC.2009.5195250

[16] T. Wang, "A Novel Face Recognition Method Based on ICA and Binary Tree SVM," *2017 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC)*, Guangzhou, 2017, pp. 251-254. https://doi.org/10.1109/CSE-EUC.2017.52

[17] Griffin, G., Holub, A., & Perona, P. (2007). Caltech-256 object category dataset.

[18] B. Weyrauch, J. Huang, B. Heisele, and V. Blanz. Component-based Face Recognition with 3D Morphable Models, First IEEE Workshop on Face Processing in Video, Washington, D.C., 2004.

[19] EPFL-RLC Multi-Camera Dataset | CVLAB. [online] Cvlab.epfl.ch. Available at: https://cvlab.epfl.ch/data/rlc [Accessed 17 Apr. 2018].

[20] R.Y. Tsai, An Efficient and Accurate Camera Calibration Technique for 3D Machine Vision, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Miami Beach, FL, pp. 364-374, 1986.

[21] Datta, A., Kim, J. S., & Kanade, T. (2009, October). Accurate camera calibration using iterative refinement of control points. In *Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on* (Vol. 27, pp. 1201-1208). IEEE. https://doi.org/10.1109/ICCVW.2009.5457474

[22] Lee, J. S. (1983). Digital image smoothing and the sigma filter. Computer vision, graphics, and image processing, 24(2), 255-269. https://doi.org/10.1016/0734-189X(83)90047-6

[23] G. Kertesz and Z. Vamossy, "Current challenges in multi-view computer vision", 2015 IEEE 10th Jubilee

International Symposium on Applied Computational Intelligence and Informatics, 2015. https://doi.org/10.1109/SACI.2015.7208206

[24] D. Chen and C. Fookes, "Labelled silhouettes for human pose estimation," 10th International Conference on Information Science, Signal Processing and their Applications (ISSPA 2010), Kuala Lumpur, 2010, pp. 574-577. https://doi.org/10.1109/ISSPA.2010.5605435

[25] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778). https://doi.org/10.1109/CVPR.2016.90

[26] King, D. E. (2009). Dlib-ml: A machine learning toolkit. Journal of Machine Learning Research, 10(Jul), 1755-1758.