Multiplet Selection: A Practical Study of Multi-Faces Recognition for Student Attendance System

ZHANG ZHIYAO
Department of Mechatronics & Systems Integration
School of Engineering, Nanyang Polytechnic
180 Ang Mo Kio Ave 8, Singapore 569830
zhangzy_bluebell@yahoo.com

Benjamin MA DI
Department of Mechatronics & Systems Integration
School of Engineering, Nanyang Polytechnic
180 Ang Mo Kio Ave 8, Singapore 569830
benjamin_ma@nyp.edu.sg

ABSTRACT
This paper presents a new approach, Multiplet Selection, for multi-faces recognition and its application in student attendance system. Instead of using a linear classifier such as SVM to classify face feature vectors, we adopt a "multiplet selection" approach such that Euclidean distances score between each identity's Anchor face [4] and a random input face are computed. Together with a pre-determined threshold parameter, this score is used for input face-identity pair association. We also develop a student attendance system based on the proposed multi-face recognition algorithm. And testing results video are available at the following URL: https://youtu.be/OZOgcw7B1YI.

CCS Concepts
Computing methodologies → Computer graphics → Image manipulation → Image processing • Image-based rendering.

Keywords
Multiplet selection; Face detection; Face verification; Face recognition; FaceNet; MTCNN (Multi-task cascaded convolution neural networks); Student attendance system..

1. INTRODUCTION
Face recognition is a popular biometric-based identity authentication technique. It has been applied in many areas. Such as CCTV surveillance, access control, 2FA (Two Factor Authentication), personal tracking, etc. It has been a long-standing research topic in the computer vision community. In the early 1990s, the author of [12] presented an Eigenface approach for face feature extraction and recognition. Eigenfaces is a set of eigenvectors derived from the covariance matrix of the probability distribution over the high-dimensional vector space of face images.

ICIGP 2019, February 23–25, 2019, Singapore, Singapore

ACM ISBN 978-1-4503-6092-0/19/02…$15.00

https://doi.org/10.1145/3313950.3313954

Many algorithms are developed based on this principle of low-dimensional representation of face images, such as [13,14,15]. However, the main problem of this approach is that it suffers from uncontrolled face images variation due to different viewing angle, illumination or shadowing. To address this challenge, researchers proposed many local-feature-filtering approaches to achieve robust face recognition [16, 17]. These approaches use some invariant properties for local filtering.

Recently, deep learning-based face recognition system performs outstanding performance compared to traditional methods. Deep learning-based methods leverages on deep convolutional neural network architect to extract face feature vectors (known as embeddings) from raw face images and classify these embeddings using linear classifier such as SVM [10].

With the help of the MTCNN [3], we achieve real-time face detection by using Multi-task cascaded convolutional neural networks, which has been provided a good foundation for recognizing multi-faces of different identities. Furthermore, Based on FaceNet [2], we can easily obtain and train the identities' face features (embeddings). However, when using transfer learning to convert in own dataset. It' s challenge to classify all identities' face feature vectors by using linear SVM classifier when the registered faces database is very huge.

Therefore, by leveraging the state-of-the-art face detection MTCNN [3] and face feature extraction model FaceNet [2], we propose a new approach is named by multiplet selection, it's highly accurate to classify different identities' face feature vectors. We test the proposed method in LFW [5] and VGGFace2 [6] two datasets. The rest of this paper are organized as follows. In section 2, the related works will be discussed. In section 3, we present the details of multiplet selection algorithm architecture. In section 4, we design the algorithm in real-time face recognition, we also implement the application of real-time face recognition for student attendance system. In section 5, experiment testing results are presented. Section 6 concludes the paper.

2. RELATED WORK
In this section, we will discuss whole procedure for face recognition implementation.

Perhaps the most famous early example of a face recognition system is due to Kohonen [1], who presents that a simple neural network could perform face recognition aligned and normalized face images. The network is used to compute a face description by approximating the eigenvectors of the face image's autocorrelation matrix, which is known as "eigenfaces", in literature, which is also known as "embeddings". However, Kohonen 's system was not a practical success, because of the need for precise alignment and normalization. In following years many researchers try to solve these difficulty in face recognition, therefore precise alignment and normalization are important research topics in face recognition. Fortunately, for example MTCNN [3]'s authors have been success to align and normalize face precisely, and based on the aligned and normalized face, the face features can be extracted easier.

With the help of [3]'s structure, through P-Net (Proposal Network), R-Net (Refinement Network), O-Net (Output network), three kinds of networks are illustrated in Fig.1, we success achieve face detection with superior accuracy over face alignment.

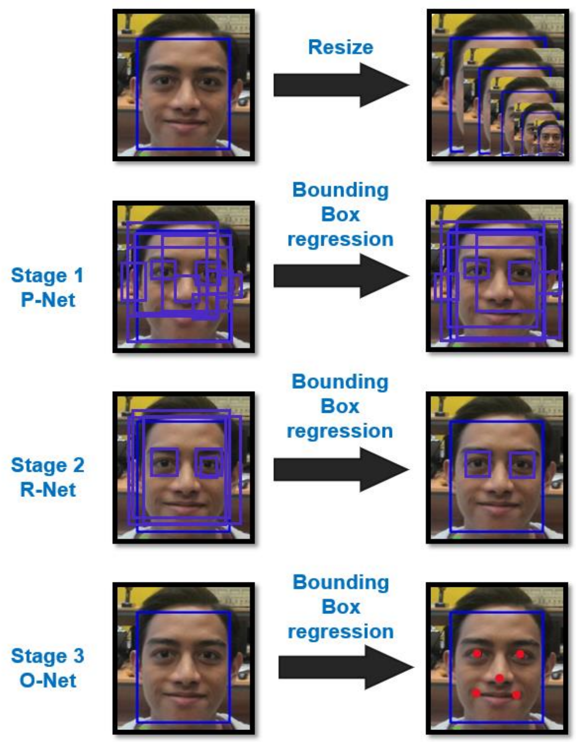

Figure 1. Cascaded framework includes three-stages multi-task deep convolutional networks to output final bounding box and five-facial

Through face detection, we would like to achieve face recognition which is illustrated in Fig 3, one to K (many identities) situation [8] rather than face verification which is illustrated in Fig 2, one to one situation [8] in order to achieve multi-faces recognition simultaneously in one frame. Before this paper, during the survey, we found there are also some algorithms and applications aren't able to recognize multiple faces. For example, Apple Inc iPhone X FaceID [7] in IOS 11 platform in 2017. It is the state-of-art face verification algorithm and application with high payable level accuracy. However, it can't handle second identity to load the face information in the system, which only meets the one to one face verification situation and brings inconvenience to the public's life. Thus, face recognition is a better choice to implement multi-faces recognition. Since, the FaceNet has been proposed by [2]'s authors, which provides a good direction to extract face features and analyze face embeddings.

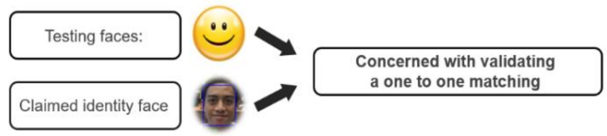

Figure 2. Face verification structure

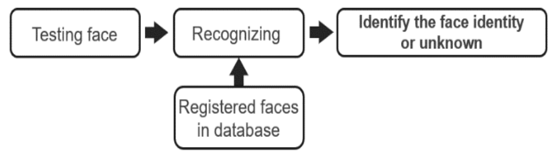

Figure 3. Face recognition structure

3. ALGORITHM ARCHITECTURE

3.1 Triplet Loss & Model Structure [2]

In this section, we adopt the state-of-the-art network FaceNet [2] to end-to-end train model weights, by leveraging of FaceNet model structure as shown Fig 4, we train a model with the testing accuracy 98.69% in LFW testing dataset and VGGFace2 training dataset. This model structure consists of input batch layer and the Inception ResNet-V1, which results in a 128-demisional embedding array, and followed by triplet loss which is illustrated in Fig 5.

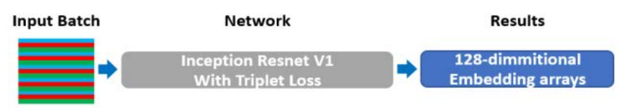

Figure 4. FaceNet Model structure

Figure 5. Triplet loss structure

And the triplet loss function is:

$$\mathcal{L} = \sum_i^n \left[\left\| f(x_i^a) - f(x_i^p) \right\|_2^2 - \left\| f(x_i^a) - f(x_i^n) \right\|_2^2 + \alpha \right]_+ \quad (1)$$

$f(x_i^a)$: the score of anchor face image

$f(x_i^p)$: the score of positive face image

$f(x_i^n)$: the score of negative face image

$[\,]_+$:Max <= 0 $\quad \|\ \|_2^2$: Euclidian distance (L2)

In [2], Triplet loss is recognized as the most suitable for face verification. So, we try to use triplet loss function to achieve real-time face recognition to differentiate multiple identities in one frame. During training, the model structure would minimize the distance between an anchor face image (Sample) with positive face image (Positive Sample) and maximize the distance between anchor face image between negative image (Negative Sample), where α is a margin that try to enhance the distance between positive pairs and negative pairs.

After training settle down, model will output 128-dimmesional embeddings array which represents the face common features. In [2], the result has shown the best feature representation is in 128-dimmtional array. However, after this, even if train a classifier for own dataset based on registered identities' face (Anchor face) to classify each face feature vectors. we adopt traditional deep learning method linear SVM classifier [10] to classify fifty different identities, unfortunately, the real-time testing rate performs very low:

- When the registered identities in database are huge (more than fifty people)
- When multiple identities appear in one frame simultaneously (more than one person)

which result will be released in section 5 experiment and evaluation. Therefore, in order to solve the above existed problems, we propose a new approach named by "Multiplet Selection" to classify multiple identities face feature vectors and improve the accuracy further.

3.2 Multiplet Selection

As we have described in section-3.1, the linear SVM classifier is not a good choice to converge all 128-dimetional array into a quantifiable value to represent all identities to recognize multiple faces when in a huge database. so that, with the inspired by the [2]'s triplet loss and L2 normalization [11], we propose a new approach "Multiplet Selection" to achieve recognizing multiple faces in huge database and clear with accurate differentiate multiple faces in one same frame simultaneously. The method of features visualization for multiplet selection is illustrated in Fig 7.

As shown as Fig 6, this is the simple graph to describe how multiplet selection convergent all identities' 128-dimmtional array into quantifiable value. Originally, every identities' face comes in FaceNet and output from structure's last layer the 128-dimmtional embedding array are normalized array, we use Euclidean distance calculate them into one single value in order to compare every

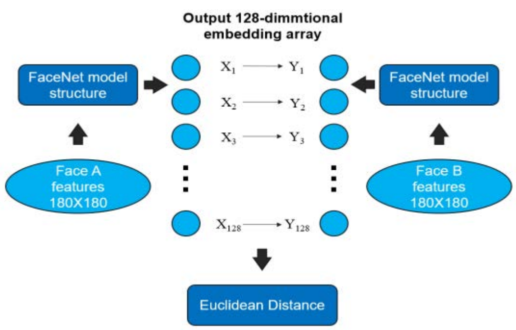

Figure 6. Multiplet selection features

identities' face quantifiable value and pull apart the distance of each identity. When comparing the testing identity's face with registered identity in database, we propose this way in multiplet selection step one is to calculate two identities distance as shown in function (2) and (3):

$$\text{Distance} = \sqrt{\sum_{i=1}^{n=128}(x_i - y_i)^2} \quad \textbf{(2)}$$

$$= \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \ \ldots + (x_{128} - y_{128})^2} \quad \textbf{(3)}$$

Hence, our proposed solution in multiplet selection step two as illustrated in Fig 7 is to compare testing identities' face with registered faces in database with the advanced constant value α. Which α is a value to pull apart the distance between same identity with different identity with preventing the mis-recognition between two different identities. Therefore, by using multiplet selection to be the last layer to convergent all identities so that we can differentiate different identities with clear and high accuracy.

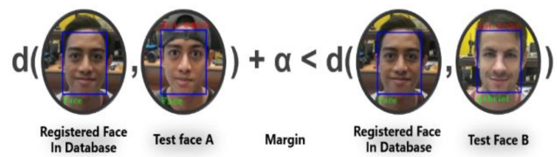

Figure 7. Multiplet selection distance

4. APPLICATION ARCHITECTURE

Maintaining the attendance is vital important in all institutions for checking the performance of presenters (students, staffs). Traditionally, presenters' attendance is taken manually by using attendance sheet given by the faculty members in class or meeting, which is time consuming event. it is very difficult to verify one by one presenters in a large environment with distributed branches whether the authenticated presenters are responding or not. So, face recognition system is gradually evolving to a universal biometric solution. Hence, we develop a student attendance system based on the proposed multi-faces recognition algorithm multiplet selection. The flowchart of student attendance flowchart is illustrated in Fig 8.

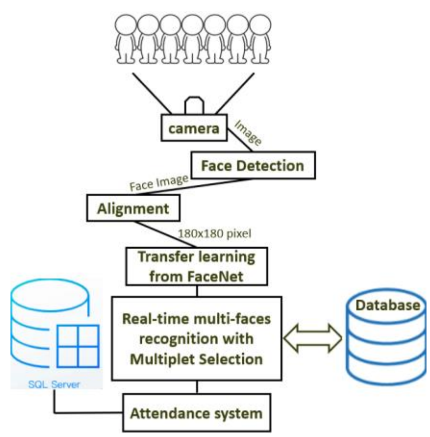

Figure 8. Student attendance system flowchart

Multiplet selection is a proposed method to improve last layer's convergence of embedding array and help to implement the application to handle multi-faces recognition practically. So, we create a new attendance system to record students' attendance by face recognition.

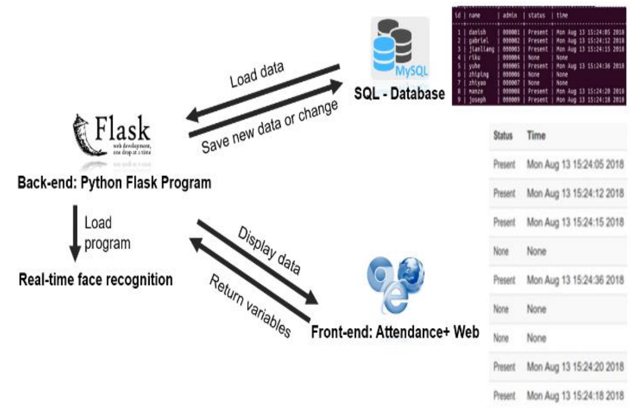

Figure 9. Student attendance system web-based structure

As shown in Fig 9, we adopt the MySQL server to be our database to store every students' information (Name, Identity number, Present time, Status). we also adopt Flask library to develop back-end of system and named with "Attendance+" in front-end of system (Web-based).

In the back-end Flask program, we load the program real-time face recognition with multiplet selection function to recognize students' face when run the student attendance system "Attendance+" in website.

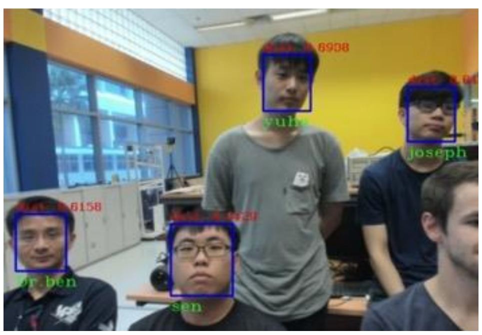

Figure 10. Testing for real-time multi-face recognition

5. EXPERIMENT AND EVALUATION

5.1 Algorithm Performance on Datasets

With the help of FaceNet, we select VGGFace2 becomes training set and LFW becomes validation set. And, we end to end train the model with triplet loss on Inception ResNet-V1 on GeForce GTX 1070 for 107 hours, the accuracy is 98.69% as well as the validation rate is 92.57%.

We evaluate our real-time multi-faces recognition with multiplet selection approach by using standard protocol for unrestricted on two different datasets (LFW and VGGface2 with input size 180x180 pixel) to record error rate and real-time testing rate. In order to display the effect of algorithm more intuitively, we define error rate (False positive rate) and the real time testing rate (Ture positive rate) are:

$$FPR = \frac{FP}{FP+TN} \ ;TPR = \frac{TP}{FP+FN}$$

Both this two-validation rate is focused on every frame with comparing registered identity's face with testing identity's face to obtain whether the algorithm's judgement is correct or wrong. we randomly choose fifty identities with per identity 300 images from VGGface2, so we made the fifty-identities becomes our registered face in the database, the LFW dataset becomes testing faces. And we set the threshold margin in our multiplet selection is 0.28, we execute the experiment for twenty times testing in LFW dataset. Fortunately, the result for error rate is 0.0.

As for the real-time testing rate, based on randomly selected fifty-identities (ID), we select one of image per identity becomes registered face in database, and remaining part becomes the corresponding testing faces. Hence, the real time testing rate (compared with trained linear SVM classifier) is:

Table 1. Real-time testing accuracy

Identity	Multiplet selection	Linear SVM classifier
ID1	**74.48%**	**20.69%**
ID2	**67.34%**	**30.62%**
ID3	**63.79%**	**17.96%**
…	**Average 66.72%**	**Average 30.04%**
ID50	**71.26%**	**38.62%**

5.2 Real-Time Face Recognition Testing Speed

Hence, we analyse the error rate and testing rate of multiplet selection based on FaceNet, as for our approach real-time face recognition program's running speed is illustrated as follow, we set the testing identities' face two meters away from camera and the resolution of running frame is 640x480 pixel. Furthermore, the screenshot of testing for real time multi-face recognition is shown in Fig 10.

Table 2. Program running speed

Identity per frame	Testing speed Highest	Testing speed lowest
1	**19.617 FPS**	**9.710 FPS**
2	**15.519 FPS**	**8.892 FPS**
3	**13.737 FPS**	**9.726 FPS**
4	**5.467 FPS**	**3.052 FPS**
5	**4.323 FPS**	**3.565 FPS**

6. CONCLUSION

In this paper, to achieve face feature vectors classification instead of using linear SVM classifier and real-time multi-faces recognition. We propose, implement and test the new approach "multiplet selection". And we also develop an application "Attendance+" student attendance system based on proposed new method and reliable testing results. Moreover, we promote it in one of electrical laboratory of Nanyang Polytechnic. In the future, we plan to implement a dynamic variable of margin α in the multiplet selection distance function to achieve 100 % real-time testing rate. The testing videos of application are available at URL: https://youtu.be/OZOgcw7B1YI.

7. REFERENCE

[1] T. Kohonen, *Self-organization and Associative Memory*, Springer-Verlag, Berlin, 1989.

[2] Schroff, F., Kalenichenko, D. and Philbin, J. 2015. FaceNet: A unified embedding for face recognition and clustering. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2015).

[3] Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, Yu Qiao. 2016.

Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Networks. 2016. IEEE Signal Processing Letters.

[4] Yandong Wen, Kaipeng Zhang, Zhifeng L. 2016. A Discriminative Feature Learning Approach for Deep Face Recognition.

[5] Gary B. Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Miller. 2007. Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments.

[6] Qiong Cao, Li Shen, Weidi Xie, Omkar M. Parkhi and Andrew Zisserman. VGGFace2: A dataset for recognizing faces across pose and age.

[7] Face ID Security. NOV 2017. Apple Inc.

[8] Core Calistra. May 2018. Face Identification, Verification, and Authentication Explained. KAIROS

[9] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. CoRR, abs/1409.4842

[10] Dmitriy Fradkin, Ilya Muchnik. Support Vector Machines for Classification

[11] Charles H Martin, PhD. June 2016. NORMALIZATION IN DEEP LEARNING

[12] M. Turk and A. Pentland. Eigenfaces for recognition. Journal of cognitive neuroscience, 3(1):71–86, 1991.

[13] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. IEEE Trans. Pattern Anal. Mach. Intell., 19(7):711–720, 1997.

[14] X. He, S. Yan, Y. Hu, P. Niyogi, and H.-J. Zhang. Face recognition using laplacianfaces. IEEE Trans. Pattern Anal. Mach. Intell. 27(3):328–340, 2005

[15] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma. Robust Face Recognition via Sparse Representation. IEEE Trans. Pattern Anal. Machine Intell., 31(2):210–227, 2009.

[16] T. Ahonen, A. Hadid, and M. Pietikainen. Face description with local. binary patterns: Application to face recognition. IEEE Trans. Pattern. Anal. Machine Intell., 28(12):2037–2041, 2006.

[17] C. Liu and H. Wechsler. Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. Image processing, IEEE Transactions on, 11(4):467–476, 2002