Mixture-Based Open World Face Recognition

Arthur Matta[1,2], João Ribeiro Pinto[1,2] [0000−0003−4956−5902], and Jaime S. Cardoso[1,2] [0000−0002−3760−2473]

[1] INESC TEC, Porto, Portugal
[2] Faculdade de Engenharia, Universidade do Porto, Porto, Portugal
up201609953@fe.up.pt, joao.t.pinto@inesctec.pt,
jaime.cardoso@inesctec.pt

Abstract. Face Recognition (FR) is a challenging task, especially when dealing with unknown identities. While Open-Set Face Recognition (OSFR) assigns a single class to all unfamiliar subjects, Open-World Face Recognition (OWFR) employs an incremental approach, creating a new class for each unknown individual. Current OWFR approaches still present limitations, mainly regarding the accuracy gap to standard closed-set approaches and execution time. This paper proposes a fast and simple mixture-based OWFR algorithm that tackles the execution time issue while avoiding accuracy decay. The proposed method uses data curve representations and Universal Background Models based on Gaussian Mixture Models. Experimental results show that the proposed approach achieves competitive performance, considering accuracy and execution time, in both closed-set and open-world scenarios.

Keywords: biometrics, face recognition, open set, open world

1 Introduction

A known limitation of conventional classification algorithms is using closed identity sets. Here, all identities seen during testing have been previously presented to the method during training or enrollment stages. The presence of unknown subjects during testing or deployment has a significant negative effect on recognition performance, since the algorithms are unable to correctly recognize their biometric data.

Open-Set Face Recognition (OSFR), introduced by Günther *et al.* [4], addresses this problem by thresholding confidence scores and assigning a single "unknown" label to all samples which do not meet the defined threshold. However, a limitation of OSFR is that it does not learn or otherwise take advantage of newly available data. On the other hand, Open-World Face Recognition (OWFR), introduced by Bendale and Boult [1], extends the concept of OSFR using Class Incremental Learning (CIL). Instead of assigning all unknown subjects to a single class, OWFR discriminates data from unknown identities and learns a new class for each unfamiliar subject.

Bendale and Boult [1] proposed the Nearest Non-Outlier (NNO) algorithm, an extension of the traditional Nearest Class Mean (NCM) approach that tackles

open space risk while balancing accuracy. However, Rosa *et al.* [7] argued that several metric learning algorithms, like NNO and NCM, estimate their parameters on an initial closed set and keep them unchanged as the problem evolves, contradicting the very own definition of OWFR. Hence, they extended three algorithms, the Nearest Class Mean, the Nearest Non-Outlier, and the Nearest Ball Classifier, to update their metric and novelty threshold online. Following this line of thought, Doan and Kalita [3] developed a similar approach, employing their solution for the incremental addition of new classes, but optimizing the nearest neighbor search to determine the closest local balls.

Lonij *et al.* [5] proposed a different approach, using knowledge graph embedding to add semantic meaning by employing smoothing constraints in the graph embedding loss function and an attention-based scheme to improve novel graph predictions. For the action recognition task, Shu *et al.* [9] proposed the Open Deep Network (ODN), which applies a multi-class triplet thresholding technique to detect new classes and then dynamically reconstruct the network's classification layers, continually appending predictors for new categories. Xu *et al.* [12] proposed a meta-learning algorithm that only requires a trained meta-classifier to continually include new classes when sufficient labeled data is available and detect/reject later unseen subjects.

OWFR is a relatively new concept and, therefore, the related literature is limited. However, the existing state-of-the-art approaches still present some issues, mainly regarding their accuracy compared to standard closed-set approaches. The algorithms proposed in [7], for example, achieved a maximum accuracy of approximately 50% on the ILSVRC'10 dataset. The need to retrain models also leads to long execution time, even with relatively small data sets.

Considering the current limitations in OWFR, this work proposes a fast and straightforward algorithm which models identity classes using mixture-based data representations, aiming for high recognition accuracy and time efficiency. The proposed algorithm is formulated with two variants regarding the data representation methodologies. The first variant represents biometric samples as curves and assigns identities based on feature-wise distances to each class' representation. The second variant employs Gaussian Mixture Models (GMM) to represent identities with a Universal Background Model (UBM) as a normalization factor. These approaches were directly compared to the state-of-the-art Online NNO algorithm [7].

Besides this introduction, four more sections compose the remainder of this paper. Section 2 details the mixture-based algorithm proposed in this paper, as well as its two variants. Section 3 describes the experimental settings and databases used for development and evaluation. Section 4 presents and discusses the results obtained from the conducted experiments. Section 5 gathers the conclusions drawn from this work and indicates some potential paths for future research.

2 Proposed Methodology

This section introduces the mixture-based OWFR methodology developed in this work. The first method variant focuses on representing N-dimensional data as visual curves in a bidimensional space. The second variant extends the previous approach by employing Gaussian Mixture Models and Universal Background Models for more efficient and complete representation of each identity class.

2.1 First variant: curve-based representations

The first variant of the proposed method employs a data simplification by representing any N-dimensional biometric sample as a curve on a bidimensional space. It then assigns an instance to an identity class by calculating a feature-by-feature distance between an instance's curve and a class' curve. Figure 1 illustrates the representative scheme of this approach.

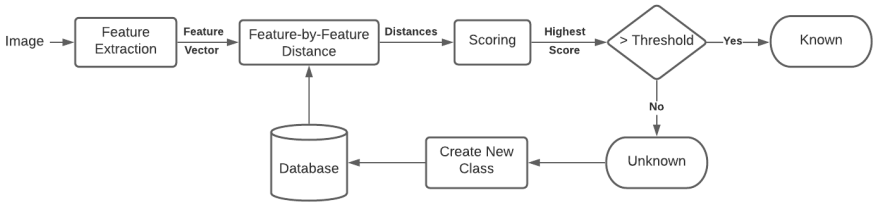


Fig. 1. Scheme of the first variant of the proposed method.

This approach allows the same class to have multiple clusters, each with its centroid, using a label encoder that maps each group to an index. Therefore, given $Y = \{y_1, ..., y_j\}$ the set of all classes, the index representation is given by $I = \{i_1^1, i_1^2, i_2^1, ..., i_k^g\}$, where i_k^g is the index pointing to cluster g of class k, with u_k^g representing the centroid of cluster i_k^g. For simplification, consider $I = \{i_1, i_2, ..., i_c\}$ the set of all indexes with μ_c representing the center corresponding to i_c. The algorithm first step calculates a threshold for each feature indicating the interval to which an element can be considered related to that cluster:

$$T^n = F \times \sigma^n, \tag{1}$$

where F is a constant scale factor used to adjust the interval's width, and σ^n is the standard deviation of the nth-feature considering all μ_c. The algorithm's second step is calculating a feature-by-feature distance between the instance $x \in X = \{x_1, x_2, ..., x_i\}$, with $X \in \mathbf{R}^N$, and each cluster' center μ_c, as follows:

$$D_{xc}^n = |\mu_c^n - x^n|, \tag{2}$$

where μ_c^n and x^n are the nth-feature of μ_c and x, respectively. This distance is employed to calculate a score value between [0, N] to an instance x for cluster c using the formulation:

$$S_{xc} = \sum_{n=1}^{N} \left[D_{xc}^n \leq T^n \right], \tag{3}$$

where [ ] symbolize the Iverson Brackets which return 1 if the condition inside is true and 0 otherwise. The algorithm accepts the instance x as belonging to cluster c with highest score $S_{xc_{max}}$ if $S_{xc_{max}} \geq H$, where H is a threshold in $[1, N]$, updating the respective centroid using the equation:

$$\mu_{c_t}^{t+1} = \left(1 - \frac{1}{n(c_t)}\right) \mu_{c_t}^t + \frac{1}{n(c_t)} x_t, \tag{4}$$

where $n(c_t)$ is the number of instances of cluster c_t at time t (including the current sample). Note that the initial value for μ^1 is equal to the first sample x_1. However, if $S_{xc_{max}} < H$, it assigns x to one of the unknown classes $U = \{u_1, u_2, ..., u_i\}$ by repeating the same procedure described above but replacing the set I for U. If $S_{xu_{max}} < H$, then a new unknown class u_{i+1} is created and x assigned to it.

The final step is, after some period, converting u_i into a known class when it reaches a minimum number of samples or discarding it otherwise. When a new cluster is created, an index pointing to it is also generated.

2.2 Second variant: GMM-UBM representations

The aforedescribed first variant employs a distance-from-cluster-center approach and therefore is susceptible to the curse of dimensionality. To avoid this, the second variant replaces the curve representation using a Gaussian Mixture Model (GMM) to represent each class and using the likelihood ratio as biometric comparison score. Figure 2 illustrates the representative scheme of this approach.

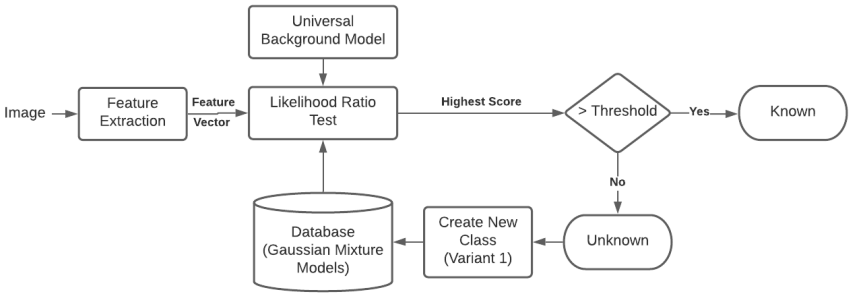


Fig. 2. Scheme of the second variant of the proposed method.

The GMMs apply a finite number of Gaussian distributions to model any arbitrarily-shaped cluster more accurately, increasing the method's robustness. They employ an expectation-maximization (EM) algorithm to calculate a weight encoding the probability of membership to each component for each point and then use it to update the corresponding component' parameters, ensuring the convergence to a local optimum. The general notation of a GMM is $\lambda = \{\omega_i, \mu_i, \Sigma_i\}$, with $i = 1, ..., M$, where M is the number of Gaussian components used, and ω_i, μ_i, and Σ_i are the component's weight, mean vector, and covariance matrix, respectively. The mean vector defines the Gaussian distribution's location in space, while the covariance matrix determines its density contours' direction and length. This approach employs diagonal covariance matrices.

The likelihood ratio (LR) test assesses the fitness between two models by employing a hypothesis test: given an observation, O, and a person, P, define the hypothesis $H_0 = O$ *is from* P, and $H_1 = O$ *is not from* P. Then, the ratio between the probability density function (or likelihood) for both hypotheses can be computed through:

$$LR(O) = \frac{p(O|H_0)}{p(O|H_1)} \begin{cases} \geq \theta \text{ accept } H_0 \\ < \theta \text{ reject } H_0 \end{cases} \tag{5}$$

In this approach, the GMM describes a feature's distribution derived from the corresponding person, hence characterizing a hypothesis, and each feature vector represents an observation. Therefore, the LR test becomes:

$$LR(x, P) = \frac{p(x|\lambda_P)}{p(x|\lambda_{\overline{P}})}, \tag{6}$$

where λ_P and $\lambda_{\overline{P}}$ are parameters denoting the weights, means, and covariance matrices of the corresponding GMM. The $p(x|\lambda_P)$ is a probability density function given by the weighted sum of the GMM's M components:

$$p(x|\lambda) = \sum_{i=1}^{M} \omega_i \; g(x|\mu_i, \Sigma_i), \tag{7}$$

with each component $g(x|\mu_i, \Sigma_i)$ being a function of the form:

$$g(x|\mu_i, \Sigma_i) = \frac{1}{(2\pi)^{\frac{N}{2}} |\Sigma_i|^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2} \; (x - \mu_i)' \; \Sigma_i^{-1} \; (x - \mu_i) \right\}. \tag{8}$$

The issue, however, is how to define the likelihood of the alternative hypothesis. λ_P can be acquired using the training data, but $\lambda_{\overline{P}}$ must encompass the entire space of possible alternatives to person P. One could calculate the probability $\lambda_{\overline{P}} = F(p(x|\lambda_1), ..., p(x|\lambda_N))$, where F is a function such as average or maximum, for all the possible alternatives to person P but this is not suitable for applications with many alternatives. A Universal Background Model (UBM) pools samples from several different classes, resulting in one single model which represents all alternative hypotheses:

$$p(X|\lambda_{\overline{P}}) = p(x|\lambda_{UBM}). \tag{9}$$

Thus, given an instance $x \in X$, the method calculates the LR for each class using both the corresponding GMM and the UBM:

$$LR(x, y) = \frac{p(x|\lambda_y)}{p(x|\lambda_{UBM})}, \tag{10}$$

accepting x as belonging to the class y' with the highest LR if $LR(x, y') > \theta$, where θ is a provided threshold. The models of this algorithm are static, and hence they do not update. If all classes reject the instance, this method assigns the instance to one of the unknown classes using the first algorithm. However, after accumulating the minimum number of samples, it generates a new GMM fitted to the corresponding data.

3 Experiments

The expected scenario for these algorithms consists of small and closed environments (e.g., building entrance, corridor, room) where registered and unknown individuals should be recorded and recognized. As such, the experiments were designed to mimic these expected application settings. The data used was drawn from the the VGGFace 1 database [6], resulting in a subset comprising 30k images from 300 identities.

The experiments consist of two scenarios: closed-set and open-world. The first scenario validates whether the algorithms can recognize faces properly and consists of an increasing number of subjects from 50 to 100, 150, and 200. The second scenario validates whether the algorithms can handle the unknown subjects and learn new identities. It uses an increasing number of known and unknown subjects: 50 and 100 known identities, and 50, 100, 150, and 200 unknown identities.

The training and validation data sets comprise 70 and 30 images per individual, respectively. All these images were randomly shuffled before being fed to the algorithms. To detect and extract the faces, both algorithms employed the Facenet [8] library, a state-of-the-art method that converts each image into a 512-dimensional feature vector. All implementations and evaluations were performed using Python. Performance is evaluated through recognition accuracy: the fraction of test queries which are correctly assigned their true identity label (in case they are enrolled) or the unknown class.

4 Results

This section presents and discusses the results obtained in the experiments on closed-set and open-world scenarios. Figure 3 presents the results in the closed-set scenario. The first variant of the proposed method stands out for presenting

higher accuracy when compared to the alternatives, and the recognition performance attained by both proposed method variants remains relatively stable with growing sets of identities. In most cases, the proposed method outperformed the baseline algorithm: oNNO's performance increased significantly as the number of identities increased, surpassing the second variant and almost achieving the same accuracy as the first variant.

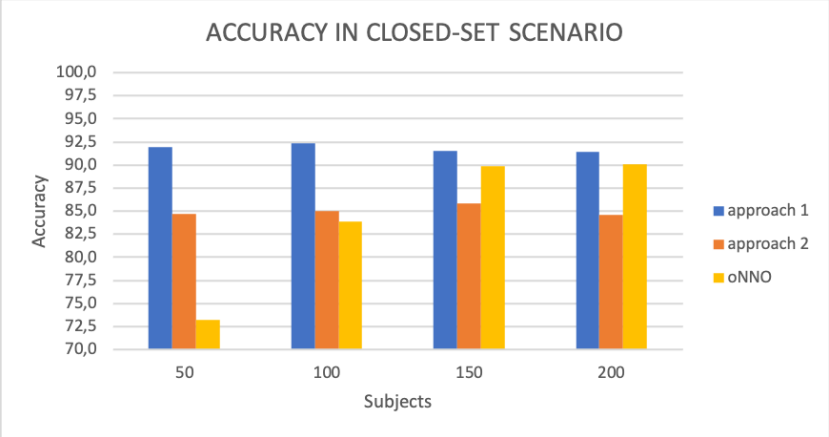
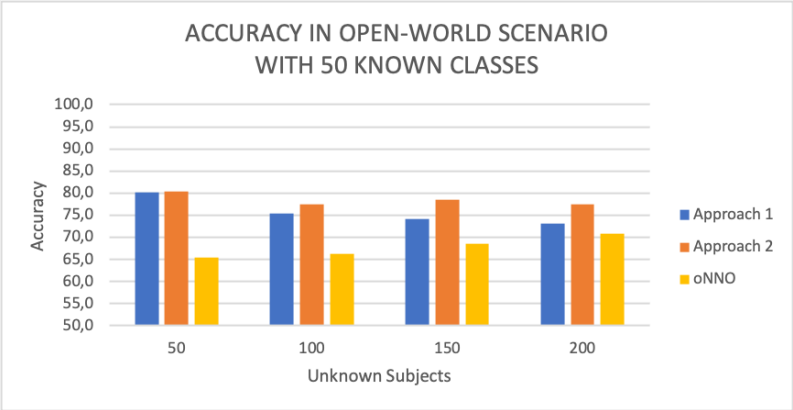


Fig. 3. Comparison of results on the closed-set scenario, with 50, 100, 150, and 200 subjects.

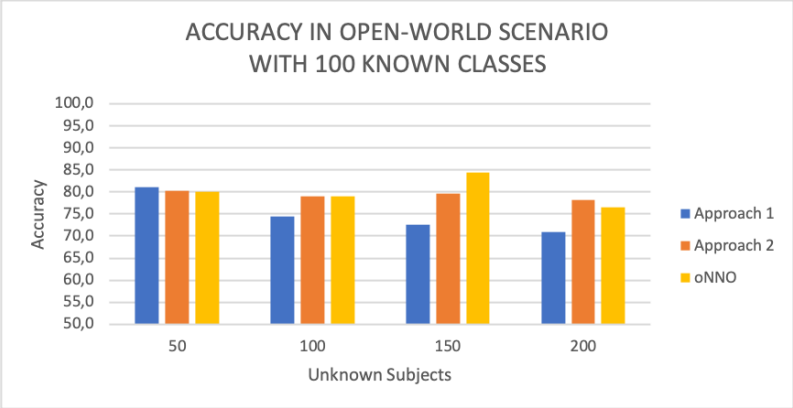
Figure 4 presents the results in the open-world scenario. In this scenario, with 50 known individuals, both proposed variants outperform the baseline by a considerable margin. However, as the number of unknown subjects increases, the accuracy of the first variant quickly degrades, unlike the baseline, whose accuracy generally increases with more unknown subjects. The second variant of the proposed method, in turn, offers more stability with growing sets of unknown identities, retaining high accuracy across all tests. Overall, the proposed method was able to outperform the baseline on this scenario with few known identities.

With 100 known identities, all methods have approximately the same performance with fewer unknown subjects. However, as the number of unfamiliar identities increased, the first variant presented once more a notable decrease in accuracy, performing worse than the baseline. On the other hand, the second variant of the proposed method retained competitive performance when compared to the baseline, presenting approximately the same accuracy regardless of the number of unknown identities.

Regardless of the scenario and number of identities employed, both variants of the proposed methodology for mixture-based open world recognition presented considerably faster execution times than the baseline, for both training and validation processes. This is largely due to the proposed method's simplic-



(a)



(b)

Fig. 4. Comparison of results on the open-world scenario obtained with (a) 50 and (b) 100 known individuals, and 50, 100, 150, and 200 unknown subjects.

ity and straightforwardness when compared with the alternative state-of-the-art approaches.

5 Conclusions

This paper addressed the open-world recognition problem by proposing a mixture-based methodology with two data representation variants. The first variant represents any N-dimensional vector as a curve in a bidimensional space and calculates a feature-by-feature distance between the curves. The second variant substitutes the curve representation by Gaussian Mixture Models to represent each class.

These approaches were compared with the online NNO state-of-the-art algorithm and evaluated across two experimental scenarios: close-set and open-world. For the first scenario, the first variant of the proposed approach outperformed the baseline, while the second variant presented a performance decay with increased number of identities. For the second scenario, with 50 known individuals, both proposed methods outperformed the baseline. On the other hand, with 100 known identities, the baseline outperformed the first variant and was, on average, as accurate as the second.

Overall, the results show that the proposed approach is competitive with the state-of-the-art for open-world face recognition. This is especially true for the expected application scenario, where the biometric system would only know (have enrollment data) of relatively few users. Additionally, regardless of the scenario and number of identities employed, both proposed variants have a considerably faster execution time than the baseline.

Future work will focus on dynamically fine-tuning the facenet representation for the application scenarios and adopting more advanced techniques such as Kalman filtering [10], Long Short-Term Memory (LSTM) [2], and reinforcement learning [11]. Other topics worth studying are video scenarios and scalability tests.

Acknowledgments

This work was financed by the ERDF – European Regional Development Fund through the Operational Programme for Competitiveness and Internationalization - COMPETE 2020 Programme and by National Funds through the Portuguese funding agency, FCT – Fundação para a Ciência e a Tecnologia within project "POCI-01-0145-FEDER-030707", and within the Ph.D. grant "SFRH/BD/137720/2018". The authors wish to acknowledge the creators of the VGG Face dataset (University of Oxford, UK), essential for this work.

References

1. Bendale, A., Boult, T.E.: Towards open world recognition. In: Conference on Computer Vision and Pattern Recognition (CVPR). (2015)
2. Corrêa, D., Salvadeo, D., Levada, A., Saito, J.: Using LSTM Network in Face Classification Problems. In: IV Workshop am Visão Computacional (WVC). (Nov. 2008)
3. Doan, T., Kalita, J.: Overcoming the challenge for text classification in the open world. In: 2017 IEEE 7th Annual Computing and Communication Workshop and Conference (CCWC). (2017) 1–7
4. Günther, M., Cruz, S., Rudd, E.M., Boult, T.E.: Toward open-set face recognition. In: Conference on Computer Vision and Pattern Recognition (CVPR). (2017)
5. Lonij, V.P.A., Rawat, A., Nicolae, M.: Open-world visual recognition using knowledge graphs. arXiv (2017) Available on: http://arxiv.org/abs/1708.08310.
6. Parkhi, O.M., Vedaldi, A., Zisserman, A.: Deep face recognition. British Machine Vision Association and Society for Pattern Recognition (2015) 41.1–41.12

7. Rosa, R.D., Mensink, T., Caputo, B.: Online open world recognition. arXiv (2016) Available on: http://arxiv.org/abs/1604.02275.
8. Schroff, F., Kalenichenko, D., Philbin, J.: Facenet: A unified embedding for face recognition and clustering. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR) (June 2015) 815–823
9. Shu, Y., Shi, Y., Wang, Y., Zou, Y., Yuan, Q., Tian, Y.: Odn: Opening the deep network for open-set action recognition. In: IEEE International Conference on Multimedia and Expo (ICME). (2018)
10. Wang, L.: Face Recognition Technology based on Kalman Filter. In: 2019 9th International Conference on Management and Computer Science (ICMCS). (2019) 11–18
11. Wang, P., Lin, W., Chao, K., Lo, C.: A face-recognition approach using deep reinforcement learning approach for user authentication. In: 2017 IEEE 14th International Conference on e-Business Engineering (ICEBE). (2017) 183–188
12. Xu, H., Liu, B., Shu, L., Yu, P.S.: Open-world learning and application to product classification. In: The Web Conference (WWW). (2019)