Face Recognition Based on a Collection of Binary Classifiers

Rafael Henrique Vareto[1], Filipe de Oliveira Costa[2], William Robson Schwartz[1]
[1]Smart Sense Laboratory, Department of Computer Science, Universidade Federal de Minas Gerais, Brazil
[2]Image and Speech Processing Management, Centro de Pesquisa e Desenvolvimento em Telecomunicações, Brazil
{rafaelvareto, fcosta, william}@dcc.ufmg.br

***Abstract*—This work contemplates one of the most relevant Face Recognition tasks[1], the open-set face identification, which has to handle several unseen individuals and determine whether a given face image is associated with a subject registered in a gallery of known individuals or not. Different from closed-set face identification and face verification, the open-set face identification has much to improve since only few researchers have addressed the problem. In this work, we merge together hashing functions and classification methods to determine when probe samples are known (i.e., included in the gallery set). We carry out experiments with partial least squares, support vector machines and neural networks and show how the vote-list histograms tend to behave for known and unknown individuals whenever we test a probe sample. In addition, we obtain promising results as we conduct experiments on FRGCv1, PubFig83 and VGGFace to show that our method is effective regardless of the dataset difficulty.**

***Keywords*-open-set face identification, face recognition, surveillance, biometrics, ensemble, machine learning.**

I. Introduction

Face recognition has been one of the most important tasks in computer vision and biometrics during the last decades. Because of the wide range of face recognition applications in several environments – e.g., access control, forensics, law enforcement, social media, surveillance systems – and the accessibility of feasible recording and storage technologies in the last years, face recognition tasks received significant attention from the scientific community.

Conventional face recognition approaches traditionally extract image features that correspond to facial components. These methods would first look for shape of the eyes, mouth contour, facial hair, nose appearance, face silhouette to name a few and use them as discriminative features while exploring other images. Face recognition is a commonly employed literature term used to define mainly three tasks: *Face verification*, a 1:1 matching problem in which the goal is to determine whether a pair of images corresponds to the same subject. *Closed-set face identification*, a 1:N matching task, which assumes that every queried subject was previously catalogued, ensuring that the probe face holds a corresponding identity in the gallery set. *Open-set face identification*, commonly referred to as watch-list, similar to the closed-set task, but it does not guarantee that all query subjects are registered in the face gallery, making it a more challenging problem.

[1]This work corresponds to a MSc thesis [1].

There are many works on closed-set identification [2]–[7]. However, real-world applications cannot presume that every query image is known and, consequently, they are better pictured by the open-set face identification task since there is an incomplete knowledge of the world and countless unknown people. Such scenario comprises a classification model where only few classes are known at learning time, but many unknown classes might appear at test time [8]. Therefore, this work aims at providing efficient and straightforward techniques for the open-set face identification task.

Our main hypothesis for this work is that a vote-list histogram (a likelihood estimation of how much each subject in the gallery resembles a probe sample), proceeds differently whether we present probe face images whose identities are enrolled in the gallery set or whether we examine unseen individuals. Figure 1 illustrates two different queries: one corresponds to querying an enrolled individual and the other searches for an unknown subject.

The proposed approach combines Locality-Sensitive Hashing (LSH), Support Vector Machine (SVM), Partial Least Squares (PLS) and Artificial Neural Networks (ANN). LSH [9] was designed to solve near-neighbor search in high-dimensional spaces, mapping similar items to the same *bucket*. SVMs [10] are supervised learning methods that choose the hyperplane that maximizes the distance to nearest data points. PLS [11] weights feature descriptors to best discriminate throughout different classes, handling high-dimensional data and overcoming the problem of having just a few-sample-per-class. ANN [12] is a biologically-based programming paradigm that enables computer systems to learn from observational data. It is composed of numerous highly interconnected processing elements (neurons) working in harmony to solve specific problems. We replace each LSH random projection either by SVM, PLS or ANN to obtain better discrimination between positive and negative samples. A set of these learned classifiers is employed to find out whether a query sample is known, that is, it has a corresponding identity in the face gallery.

To the best of our knowledge, this is the first combination of binary classifiers and LSH for the open-set identification task. The original and extended versions of this paper are described by Vareto et al. [1], [13], [14], which received a significant award from a renowned international conference on biometrics (IJCB 2017).

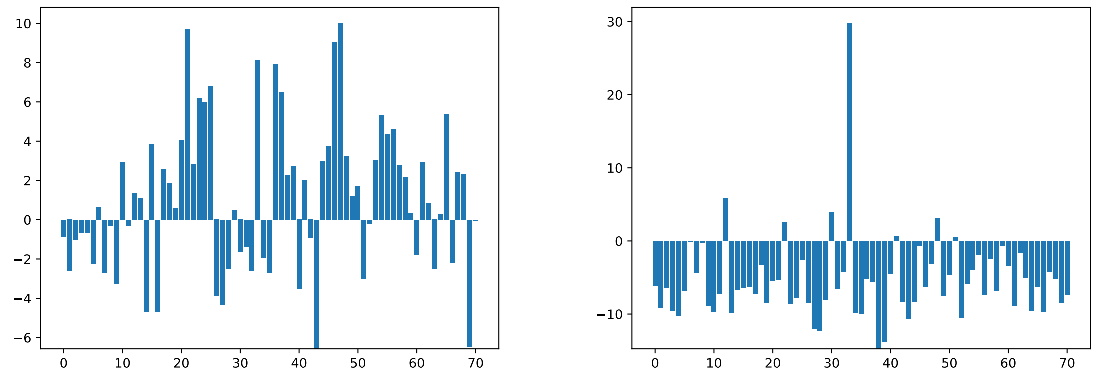

Fig. 1. VOTE-LIST HISTOGRAMS – Two queries for the same individual of the FRGCv1 dataset when the searched subject is either an unknown person *(left)* or a gallery-registered individual *(right)*. Each bin of the histogram corresponds to a registered individual. Note that on the left vote-list histogram, a considerable number of subjects from the gallery set turns into candidates when there is no clue what the identity for the query image is. This is a common behavior when the queried subject is not known during training time. On the other hand, on the right vote-list, a single bin stands out from the others, indicating the correct identity for the queried subject.

The predominant contribution of this work for face recognition are: *(i)* an adaptation of locality-sensitive hashing linked with different binary classifiers in a supervised learning setting, *(ii)* easy-to-implement and effective algorithm with few parameters to be estimated; *(iii)* a fast approach that is capable of handling the combination of diverse feature descriptors, and *(iv)* an extensive experimental evaluation and discussion of the proposed algorithm.

II. PROPOSED APPROACH

Open-set identification determines whether a given face image is associated with a subject registered in a set of known individuals, called *gallery set*. As illustrated in Figure 1, our premise is that vote-lists behave differently when we present probe face images whose identities are enrolled in the gallery set. That is, we presume that when a probe sample is known, most classifiers would *vote* for the correct identity or otherwise distribute the votes among distinct individuals that have already been registered in the gallery.

The algorithm presented here employs an embedding approach in conjunction with binary classifiers, also referred as *hashing functions*. It converts the original feature data into a metric space where a Hamming distance seems to represent well the similarity between gallery and probe images. Instead of having LSH splitting the feature space by establishing random regressions, we appraise either partial least squares, support vector machines or artificial neural networks for discriminability enhancement and classification. Our method offers a balance between simplicity and learning speed on one hand and accuracy and flexibility of the learned similarity concept on the other. In the following subsections, we clarify how these classifiers can be incorporated into LSH in favor of generating independent hashing functions.

To determine whether a face image is enrolled in the gallery of individuals, visual feature descriptors are extracted from a query face image. Then, the descriptors are presented to each hashing function to avoid comparing the probe feature vector to all gallery subjects, which makes the method scalable (i.e., the number of tests does not depend on the number of subjects in the gallery). A vote-list histogram is hence set up with size in accordance with the number of individuals enrolled in the gallery set during training time. If the algorithm establishes that a probe image corresponds to an enrolled identity, it sorts the vote-list histogram in descending order so that it turns into a list of candidates.

Similar to most supervised learning problems, our algorithm is based on two canonical steps: training and testing. The proposed approach analyzes feature vectors and their corresponding identities to learn an inferred function for every single hashing model, which are used to generate vote-list histograms whenever a query is requested.

A. Training Stage

As illustrated in Figure 2, the training stage starts randomly partitioning all subjects of the gallery set into two disjoint collections, positive and negative sets, multiple times. In pursuance of a balanced division, samples are drawn from a binomial distribution in the interest of associating a gallery-enrolled subject with the positive class when the distribution value gets closer to one or with the negative class, alternatively.

Just as we split all subjects into the positive or negative collection, we guarantee that each subset contains approximately the same number of individuals (fifty-fifty partitions). Besides, we ensure that all samples belonging to a given individual will be in the same class. The approach runs a learning algorithm so that distinct classification models $c_i \in C$ are created – a model for each pair of positive and negative collections in a binary fashion. The binary model training is repeated several times and, therefore, each classification model contains different individuals belonging to positive and negative collections. Feature descriptors are extracted from samples in the positive and negative sets and are combined with their corresponding target values ($+1$ to samples in the positive set and -1 to samples in the negative set), so that classification models can be properly learned.

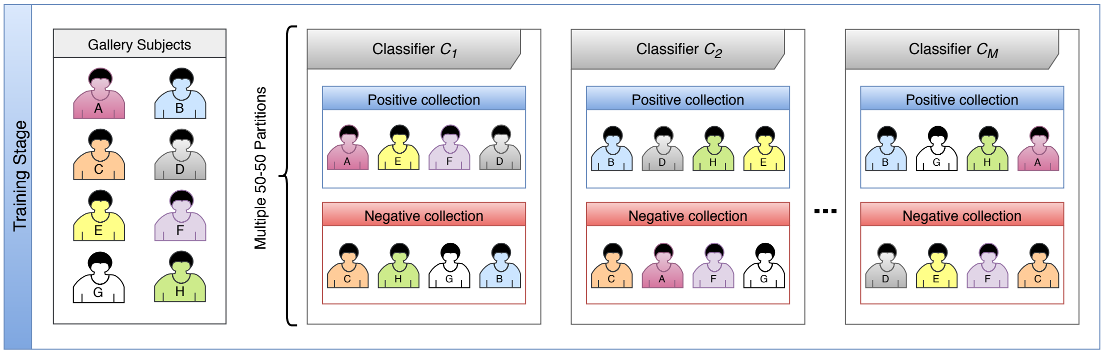

Fig. 2. TRAINING STAGE – Feature descriptors are extracted for all subjects' samples before partitioning them into positive and negative sets. Then, different homogeneous classification models are generated containing distinct individuals in each collection. Note that in this example, each classifier shares the same eight individuals; however, their distributions among positive and negative sets are unequal to provide a unique *binary signature*.

The binomial random partitioning assigns binary values to each known subject M times so that different classification models hold distinct subjects in every single collection. This way, a *unique* binary signature identifies one subject from a group of similar individuals since any chance of repeatedly having two or more subjects in the same partition reduces as the number of classification models, M, increases, providing to each subject in the gallery an exclusive binary signature during training time. Then, presenting the query sample to each classifier, outputs its corresponding probe binary signature and the vote-list histogram. Gallery subjects containing signatures similar to the probe signature are more likely to match its true identity in case of known subjects since it has a firsthand impact on how vote histograms probability estimates are distributed.

The proposed approach does not learn an embedding containing heterogeneous machine learning models. More precisely, when a list of classifiers C is generated, all its classifiers $c_i \in C$ share a homogeneous nature in terms of learning methods since only one out of SVM, PLS and FCN is adopted (uniform). The method is an adaptation of Bootstrap Aggregating [15], an ensemble meta-algorithm for machine learning.

B. Testing Stage

At the testing stage, the vote-list histogram is hence set up with size in accordance with the number of subjects enrolled in the gallery set during training time. The method extracts the same feature descriptors employed during the training step to extract features from the query face image. Each classification model has a record of which individuals are randomly categorized as belonging to the positive and negative collection. Figure 3 illustrates the process the query feature vector is presented to each one of the classifiers $c_i \in C$ in exchange for the corresponding response value r_i.

When a classifier c_i's response score r_i is closer to $+1$, it indicates that the query image sample is very much alike the subjects in the positive collection. The algorithm votes for individuals from c_i's positive class as it only increases their bins in the vote list. Additionally, if classifier c_i's score is closer to -1, then the query image sample probably resembles subjects in the negative set, which results in a subtraction of the vote list bins that correspond to c_i's positive subjects.

While we compare a probe sample to all classification models, we increment each model's response value r_i on the vote list for those subjects belonging to the positive set only. In the end, we sort the vote list in decreasing order in behalf of keeping individuals with higher probability of matching the probe sample on the top of the vote list ranking. In essence, we want to find out a threshold that indicates how much the top scorer TS_1, namely the leading subject of the sorted vote list, stands out from all other individuals.

It is important to note that the number of probe projections does not depend on the gallery set size (it depends only on the number of classifiers), implying that the computational cost at testing time remains the same as the number of enrolled subjects increases. This characteristics makes the proposed method scalable.

C. Artificial Neural Network

In pursuance of better recognition results, we replace the standard partial least squares and support vector machine classifiers with artificial neural networks classifiers [12]. As depicted in Figure 4, we propose a small network architecture with three layers: input, hidden and output layer. Each node is a neuron with a nonlinear activation function that is connected to every neuron in the previous layer. The hidden layer is set up with rectification non-linearity (ReLU) and the last layer is equipped with a soft-max function. This was the chosen architecture because it reported the best results in an exploratory experiment with several other architectures, considering different numbers of neurons and layers.

III. EXPERIMENTAL RESULTS

In this section, we evaluate the approach described in Section II. A thorough experimental information is detailed in the Master Thesis for the proposed approach [1]. All experiments are performed on a Intel Xeon E5-2630 CPU with 2.30 GHz and 16GB of RAM using Ubuntu 14.04 LTS. Code

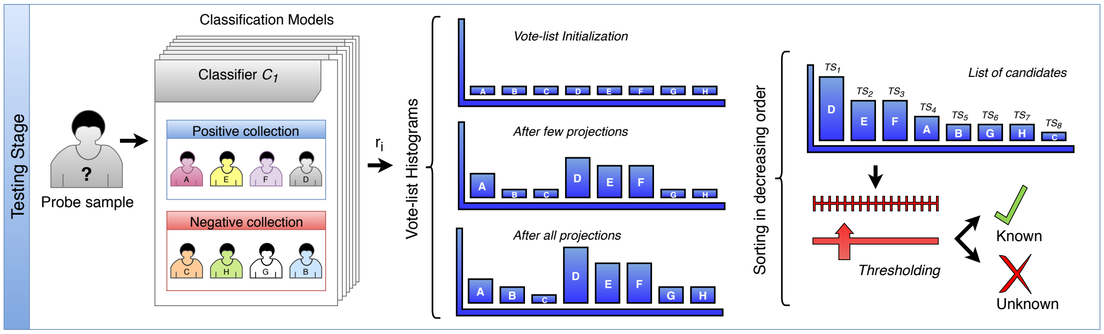

Fig. 3. TESTING STAGE – A vote list histogram is initialized containing all individuals enrolled in the gallery set during training time. The probe feature is presented to all classification models and their response values are used to increment the vote list. The vote list is then sorted in decreasing order in the interest of computing the ratio of the top scorer TS_1 to the remaining individuals. If the ratio of the highest score to the remaining subjects satisfies a threshold-based decision criterion, the subject is considered as known (i.e., belonging to the gallery).

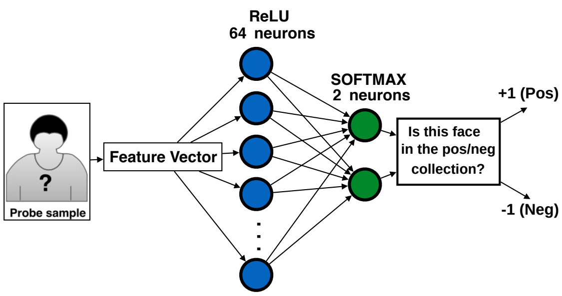

Fig. 4. The three-layer artificial neural network designed to be in place of each PLS or SVM model. The network is fed up with feature vectors to learn weights that will determine whether the probe is closer to the positive or negative collection.

TABLE I
THRESHOLD SELECTION – AVERAGE AUC FOR THE FRGCV1 DATASET ON EXPERIMENT ONE WITH DIFFERENT THRESHOLDS.

Theshold	AUC ± STD
$\tau_1 = \frac{\mathcal{H}_{TS_1}}{mean(\mathcal{H}_{TS_2}+\mathcal{H}_{TS_3})}$	**0.96 ± 0.02**
$\tau_2 = \frac{\mathcal{H}_{TS_1}}{\mathcal{H}_{TS_2}}$	0.96 ± 0.03
$\tau_3 = \frac{\mathcal{H}_{TS_1}}{mean(\mathcal{H}_{TS_2}+\ldots+\mathcal{H}_{TS_p})}, p = \lceil 0.15 \times \lvert\mathcal{H}\rvert \rceil$	0.92 ± 0.07

and experimental data are available on our GitHub repository[2]. From now on, we refer to the combination of locality-sensitive hashing and support vector machine as HSVM . Equivalently, HPLS turns into the association of hashing methods with partial least squares, and HANN represents the embedding of artificial neural networks.

a) Feature Descriptors: We consider two feature descriptors in this work: Histogram of Oriented Gradients (HOG) [16] and features extracted from the VGGFace network [17]. The former was designed for object detection whereas the latter is based on convolutional neural networks (CNN) for face detection and recognition.

b) Datasets: To demonstrate the effectiveness of the proposed approach, the selected datasets contain different characteristics, ranging from frontal cropped images taken under controlled scenarios to images with lighting and pose variations. The FRGCv1 [18] consists of 152 subjects and samples that include 2D images and 3D models. We evaluate our algorithm on experiments *one, two* and *four*, consisting of two-dimensional images. The PubFig83 [19] is a fragment of the original Public Figures dataset [20] composed of 83 individuals with at least 100 samples each, comprising several uncontrolled images with pose and expression variations. The VGGFace dataset [21] contains about 2.6 million samples of more than 2,600 celebrities and public figures collected from the web. Due to its massive size and high training time required, we randomly selected a portion of the original VGGFace containing 1,000 subjects with 15 samples each.

c) Evaluation Metrics: We consider both extensively employed Receiver Operating Characteristic (ROC) curves and its Area Under Curve (AUC) for all datasets. In addition, plotting detection and identification rate (DIR) vs. false alarm rate (FAR) produces a chart known as Open-set ROC, a metric generally used to evaluate approaches composed by filtering and identification steps [22].

d) Protocols: There is not a worldwide consensus when it comes to protocols for open-set face recognition. For comparison reasons, we evaluate PubFig83 on a protocol exploited by few researchers [19], [23], [24], formalized herein as *literature* protocol. We also propose a new protocol for the experiments carried out with FRGCv1 and VGGFace, which partitions the entire dataset, varying the known individuals set size in 10%, 50% and 90%. All the remaining individuals become unseen classes during training time. For each subject in the known subset, 50% of the samples are randomly selected for training and the remaining are left for testing.

A. Parameter Evaluation

We carry out the evaluation of three different thresholds to finding out the best gallery-member decision criterion. Table I shows the mean area under the ROC curve for each threshold. Basically, they are based on the ratio of the vote-list histogram $\mathcal{H}$'s top scorer TS_1 to the average of the succeeding subjects. Better results achieved with τ_1 and τ_2 demonstrate that the

[2] https://github.com/rafaelvareto/HPLS-HFCN-openset

TABLE II
NUMBER OF CLASSIFICATIONS MODELS – EMPLOYING VGGFACE FEATURES AND VARYING THE NUMBER OF HASHING MODELS ON PUBFIG83 DATASET.

#Models		10	30	50	100	300	500
HSVM	AUC	0.683	0.881	0.908	0.940	0.966	0.972
	STD	0.028	0.018	0.013	0.010	0.007	0.005
HPLS	AUC	**0.743**	0.885	0.932	0.940	0.960	0.968
	STD	**0.030**	0.029	0.021	0.020	0.006	0.004
HANN	AUC	0.385	**0.921**	**0.959**	**0.973**	**0.977**	**0.981**
	STD	0.059	**0.016**	**0.009**	**0.004**	**0.003**	**0.003**

TABLE III
PERCENTAGE OF KNOWN SUBJECTS – EMPLOYING 100 HASHING MODELS FOR HSVM, HPLS AND HANN WITH VGGFACE DESCRIPTOR.

	Known individuals		10%	50%	90%
FRGCv1 4	HANN	AUC	**0.900**	0.867	0.868
		STD	**0.045**	0.026	0.014
	HPLS	AUC	0.848	0.863	0.839
		STD	0.059	0.020	0.024
	HSVM	AUC	0.877	**0.871**	**0.869**
		STD	0.021	**0.016**	**0.011**
	WSVM [25]	AUC	0.866	0.862	0.848
		STD	0.035	0.015	0.019
VGGFace	HANN	AUC	**0.987**	**0.976**	**0.965**
		STD	**0.003**	**0.004**	**0.006**
	HPLS	AUC	0.978	0.961	0.926
		STD	0.005	0.003	0.005
	HSVM	AUC	0.967	0.943	0.725
		STD	0.014	0.006	0.004
	WSVM [25]	AUC	0.841	0.839	0.835
		STD	0.013	0.007	0.007

addition of more bins to the threshold estimation (τ_3) worsens the algorithm decision performance. To the remainder of the open-set identification experiments, we opt for threshold τ_1.

To verify how the method responds to the parameter adjustments, we analyze the behavior of the approaches by varying the number of hashing models for the PubFig83 dataset and alternating the size of the subset of known individuals for both VGGFace and FRGCv1. For the following experiments, we also run the one-class WSVM algorithm proposed by Scheirer et al. [25] as a baseline.

Table II shows great improvement in the initial classifier augmentation when varying from 10 to 50 binary classifiers and considering 75 randomly chosen subjects out of 83 in the known set. No significant accuracy improvement is noticed when more than 100 classifiers are established. The little raise in AUC for PubFig83 with increasingly hashing models may be justified by the fact that algorithms trained with multiple-sample-per-class gallery sets are inclined to remain stable regardless of the number of hashing functions.

Generally, the accuracy of a recognition system tends to reduce when more subjects enroll in the gallery set. Table III shows that our methods does not deteriorate with the enrollment of new subjects from the FRGCv1 dataset since having more samples increase the discriminability of classifiers when there are only few samples per subject. On the other hand, the VGGFace dataset offers more challenging experiments as it is composed of 1,000 subjects. According to the results, there was a sudden accuracy drop running HSVM on VGGFace as the 100-model SVM embedding could not separate the training data linearly when assigning 90% of all VGGFace individuals as the known set.

TABLE IV
SINGLE CLASSIFIER EVALUATION ON FRGCV1 DATASET EXPERIMENT FOUR WITH FEATURES EXTRACTED USING VGGFACE FEATURES.

Values	SVM (%)	PLS (%)	ANN (%)
AVG	71.379	73.559	**77.026**
STD	02.382	01.943	**01.648**
MIN	66.419	70.614	**73.245**
MAX	76.217	77.631	**80.592**

For the purpose of analyzing each classifier's behavior individually, Table IV exposes via *hit rate* how well a single classifier correctly matches the class a probe sample belongs to. Hit rate measures the proportion of positives that are correctly identified. The goal behind this experiment is not to determine whether a subject is enrolled the gallery set, but when a subject $s \in S$ is randomly assigned to the positive collection, it is expected that the classifier outputs a positive response. Otherwise, it is likely to output a negative response value. Results show that the ANN classifier provides better results than a PLS or SVM model.

B. *State-of-the-art Comparison*

Table V presents the outcome for the proposed identification approach containing 100 binary classification models for all three datasets. It points out the performance of each feature descriptor independently. Although desirable, it was not possible to obtain the results for all datasets using HOG, resulting in some blank cells in Table V.

TABLE V
LITERATURE COMPARISON – VGGFACE, PUBFIG83 AND EXPERIMENT FOUR OF FRGCV1 DATASET.

	AUC ± STD		
Datasets	FRGCv1	VGGFace	PubFig83
Least Squares [24]	0.869	-	-
SVM-Single [24]	0.853	-	-
WSVM-VGG [25]	0.862 ± 0.014	0.839 ± 0.007	*Failed*
WSVM-HOG [25]	0.515 ± 0.027	-	-
HSVM-VGG	0.871 ± 0.016	0.943 ± 0.006	0.951 ± 0.014
HSVM-HOG	0.902 ± 0.015	-	-
HPLS-VGG	0.863 ± 0.020	0.961 ± 0.003	0.957 ± 0.006
HPLS-HOG	**0.910 ± 0.022**	-	-
HANN-VGG	0.867 ± 0.026	**0.976 ± 0.004**	**0.973 ± 0.006**
HANN-HOG	0.613 ± 0.105	-	-

According the the results, the proposed approach is capable of improving baselines with considerable margin in most cases, indicating the robustness of the proposed approaches regardless of the selected dataset. Note that even though HOG is a low-level feature descriptor, it performed well with HPLS. We believe that it can be explained by the HOG structure and FRGCv1's predominant characteristics since it encompasses high-resolution images acquired under partial controlled conditions and without pose variation. VGGFace descriptor is designed to obtain good representations from uncontrolled near-frontal face images. As a consequence,

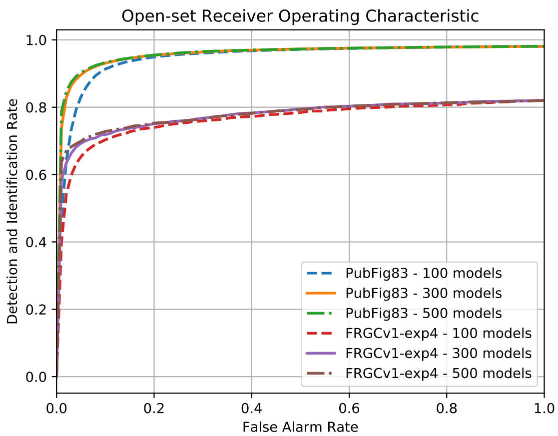

Fig. 5. FACE IDENTIFICATION RESULTS – Open-set ROC curves for PubFig83 and FRGCv1 dataset on experiment four achieved with the combination of VGGFace features, HPLS and OAA PLS.

results obtained with VGGFace descriptors on unconstrained datasets (PubFig83 and VGGFace) are considerably better than the ones from FRGCv1 under comparable parameter configuration.

C. Identification Evaluation

On the contrary of the previous experiments that only notify whether individuals are known (i.e., belong or not to the gallery), Figure 5 assesses the complete identification pipeline. Particularly, we couple the algorithm proposed in Section II with another PLS for regression so that a single model can be learnt for each subject following an *one-against-all (OAA)* classification scheme, implemented in the work of Schwartz et al. [26].

In this experiment, we employ HPLS as the trigger to OAA PLS since the latter is only executed when HPLS considers a subject as known. OAA PLS learns the same training samples employed in HPLS. Figure 5 demonstrates that our method achieved very good results for PubFig83. In this scenario, FRGCv1 turns out to be more challenging than PubFig83 due to the fact the latter has several samples per subject whereas the former holds only a single image per subject for training.

IV. CONCLUSIONS

The proposed method showed to be promising in solving a task not frequently considered in the literature, namely, open-set face recognition. We were inspired by the potential of simple binary classifiers and how locality-sensitive hashing splits the feature space. We decided to take advantage of their speed and low computational cost to determine when probe face samples are known. One of the main advantages of the proposed method is their simplicity and practical deployment since only one key parameter deeply influence performance: the number of hashing functions. Experiments have shown that VGGFace CNN descriptor contains more valuable information than HOG in the unrestrained open-set face recognition task. In addition, a comparison with the literature shows high accuracy on challenging datasets.

There are two publications concerning the proposed approach [13], [14]. The latter was honored with the *Best Paper Runner-up Award* at the IARP/IEEE International Joint Conference on Biometrics, 2017.

ACKNOWLEDGMENTS

The authors would like to thank the Brazilian National Research Council – CNPq (Grant #311053/2016-5), the Minas Gerais Research Foundation – FAPEMIG (Grants APQ-00567-14 and PPM-00540-17) and the Coordination for the Improvement of Higher Education Personnel – CAPES (DeepEyes Project).

REFERENCES

[1] R. H. Vareto, F. O. Costa, and W. R. Schwartz, "Face recognition based on a collection of binary classifiers," Master's thesis, Department of Computer Science, UFMG – Belo Horizonte, Brazil, 2017.
[2] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *PAMI*, 2006.
[3] B. F. Klare and A. K. Jain, "Heterogeneous face recognition using kernel prototype similarities," *PAMI*, 2013.
[4] J. Lu, Y.-P. Tan, and G. Wang, "Discriminative manifold analysis for face recognition from a single training sample per person," *PAMI*, 2013.
[5] D. Yi, Z. Lei, and S. Z. Li, "Towards pose robust face recognition," in *CVPR*. IEEE, 2013.
[6] Y. Lei, M. Bennamoun, M. Hayat, and Y. Guo, "An efficient 3d face recognition approach using local geometrical signatures," *PR*, 2014.
[7] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *CVPR*. IEEE, 2015.
[8] F. Li and H. Wechsler, "Open-set face recognition using transduction," *PAMI*, 2005.
[9] J. Wang, S. Kumar, and S.-F. Chang, "Semi-supervised hashing for large-scale search," *PAMI*, 2012.
[10] I. Steinwart and A. Christmann, *Support vector machines*. Springer Science & Business Media, 2008.
[11] R. Rosipal and N. Krämer, "Overview and recent advances in pls," in *Subspace, latent structure and feature selection*. Springer, 2006.
[12] B. Yegnanarayana, *Artificial neural networks*. PHI Learning Pvt. Ltd., 2009.
[13] R. H. Vareto, S. S. da Silva, F. O. Costa, and W. R. Schwartz, "Face verification based on relational disparity features and partial least squares models," in *SIBGRAPI*. IEEE, 2017.
[14] ——, "Towards open-set face recognition using hashing functions," in *IJCB*. IEEE, 2017.
[15] L. Breiman, "Bagging predictors," *Machine learning*, vol. 24, no. 2, pp. 123–140, 1996.
[16] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *CVPR*. IEEE, 2005.
[17] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *ICLR*, 2015.
[18] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge," in *CVPR*. IEEE, 2005.
[19] N. Pinto, Z. Stone, T. Zickler, and D. Cox, "Scaling up biologically-inspired computer vision: A case study in unconstrained face recognition on facebook," in *CVPRW*. IEEE, 2011.
[20] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar, "Attribute and simile classifiers for face verification," in *ICCV*. IEEE, 2009.
[21] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *BMVC*. IEEE, 2015.
[22] N. Poh, C. Chan, J. Kittler, J. Fierrez, and J. Galbally, "Description of metrics for the evaluation of biometric performance," Tech. Rep., 2012.
[23] G. Carlos, H. Pedrini, and W. R. Schwartz, "Fast and scalable enrollment for face identification based on pls," in *AFGR*. IEEE, 2013.
[24] C. S. Junior and W. R. Schwartz, "Extending face identification to open-set face recognition," in *SIBGRAPI*. IEEE, 2014.
[25] W. J. Scheirer, L. P. Jain, and T. E. Boult, "Probability models for open set recognition," *PAMI*, 2014.
[26] W. R. Schwartz, H. Guo, J. Choi, and L. S. Davis, "Face identification using large feature sets," *TIP*, 2012.