AGGREGATING PARTIAL LEAST SQUARES MODELS FOR OPEN-SET FACE IDENTIFICATION

SAMIRA SANTOS DA SILVA

AGGREGATING PARTIAL LEAST SQUARES MODELS FOR OPEN-SET FACE IDENTIFICATION

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais como requisito parcial para a obtenção do grau de Mestre em Ciência da Computação.

Orientador: William Robson Schwartz
Coorientador: Filipe de Oliveira Costa

Belo Horizonte

Janeiro de 2018

SAMIRA SANTOS DA SILVA

AGGREGATING PARTIAL LEAST SQUARES MODELS FOR OPEN-SET FACE IDENTIFICATION

Dissertation presented to the Graduate Program in Ciência da Computação of the Universidade Federal de Minas Gerais in partial fulfillment of the requirements for the degree of Master in Ciência da Computação.

Advisor: William Robson Schwartz
Co-Advisor: Filipe de Oliveira Costa

Belo Horizonte
January 2018

Ficha catalográfica elaborada pela Biblioteca do ICEx - UFMG

Silva, Samira Santos da.

S586a Aggregating partial least squares models for open-set face identification. /Samira Santos da Silva. – Belo Horizonte, 2018.
xxiv, 55 f.: il.; 29 cm.

Dissertação (Mestrado) - Universidade Federal de Minas Gerais – Departamento de Ciência da Computação.

Orientador: William Robson Schwartz
Coorientador: Filipe Oliveira Costa

1. Computação - Teses. 2. Visão computacional 3. Reconhecimento de faces. 4. Partial Least Squares I. Orientador. II. Coorientador. III. Título.

CDU 519.6*82.10(043)

UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS EXATAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

FOLHA DE APROVAÇÃO

Aggregating Partial Least Squares Models for Open-set Face Identification

SAMIRA SANTOS DA SILVA

Dissertação defendida e aprovada pela banca examinadora constituída pelos Senhores:

PROF. WILLIAM ROBSON SCHWARTZ - Orientador
Departamento de Ciência da Computação - UFMG

DR. FILIPE DE OLIVEIRA COSTA - Coorientador
Pós-Doc/Departamento de Ciência da Computação - UFMG

PROF. DAVID MENOTTI GOMES
Departamento de Informática - UFPR

PROF. ERICKSON RANGEL DO NASCIMENTO
Departamento de Ciência da Computação - UFMG

Belo Horizonte, 16 de Janeiro de 2018.

Dedicado à memória de meu pai, Valter.

Acknowledgments

First, I thank God for giving me the courage to face adversities throughout these years.

I would like to thank my family for supporting and trusting me even when I was not able to. I also would like to thank my husband for saying motivation words in a lovely way when I had no hope.

I am deeply grateful to professor William Robson Schwartz for being patient, a good listener and excellent advisor during my graduate study. Thank you for not giving up on me.

I specially thank my colleagues Rafael Vareto and Filipe Costa who have profoundly collaborated to the conclusion of this work. It was a pleasure working with you, guys.

I also thank my friends from Computer Science Department (DCC) at UFMG: Susana Gordillo, Danilo Boechat, Elaine Muniz, Guilherme Vezula, Natália Machado, Jhielson Pimentel and Masoomeh Nezhadbiglari, for the good conversations in the corridors of DCC and sincere friendship.

I am also very grateful to all members of the SSIG team and the NPDI lab for being kind, receptive and supportive, specially to Antonio C. N. Junior, Artur J. L. Correia, Cassio E. S. Junior, Carlos A. C. Junior, Gabriel R. Gonçalves, Jessica S. Souza, Keiller Nogueira, Raphael F. C. Prates, Renan O. Reis, Ricardo B. Filho, Tiago M. H. C. Santana and Victor H. C. Melo.

I would like to thank the Coordination for the Brazilian National Research Council – CNPq (Grant #311053/2016-5), the Minas Gerais Research Foundation – FAPEMIG (Grants APQ-00567-14, RED-00042-16, PPM-00540-17) and the Coordination for the Improvement of Higher Education Personnel – CAPES (DeepEyes Project).

"If you can dream it,
you can do it."
(Walt Disney)

Resumo

Identificação de faces é uma tarefa importante em visão computacional com diversas aplicações, como em vigilância, forense e interação humano-computador. Nos últimos anos, diversos métodos têm sido propostos a fim de identificar faces considerando cenários com galeria fechada, ou, seja, métodos que desconsideram o fato de que faces no teste não necessariamente pertencem a um indivíduo da galeria. Entretanto, em aplicações mais realísticas, podemos querer determinar a identidade de faces em galerias abertas. Neste trabalho, propomos um novo método, simples mas eficiente, para identificação de faces em galeria aberta através da combinação de modelos *Partial Least Squares* usando o protocolo um-contra-todos. As saídas dos modelos são combinadas em um histograma de respostas que será balanceado caso a face de teste não pertença a nenhum indivíduo da galeria ou terá um pico caso contrário. Experimentos são executados em quatro datasets: FRGCv1, FG-NET, Pubfig e Pubfig83 e os resultados mostram significante melhoria quando comparados com abordagens estado-da-arte independente das dificuldades impostas pelos datasets.

Palavras-chave: Reconhecimento de Faces em Galerias Abertas, Identificação de Faces, Partial Least Squares.

Abstract

Face identification is an important task in computer vision and has a myriad of applications, such as in surveillance, forensics and human-computer interaction. In the past few years, several methods have been proposed to solve face identification task in closed-set scenarios, that is, methods that make assumption of all the probe images necessarily matching a gallery individual. However, in real-world applications, one might want to determine the identity of an unknown face in open-set scenarios. In this work, we propose a novel method to perform open-set face identification by aggregating Partial Least Squares models using the one-against-all protocol in a simple but fast way. The model outputs are combined into a response histogram which is balanced if the probe face belongs to a gallery individual or have a highlighted bin, otherwise. Evaluation is performed in four datasets: FRGCv1, FG-NET, Pubfig and Pubfig83. Results show significant improvement when compared to state-of-the art approaches regardless challenges posed by different datasets.

Keywords: Open-set Face Recognition, Face Identification, Partial Least Squares.

List of Figures

List of Tables

Contents

Chapter 1

Introduction

Face Recognition is a natural task performed daily and effortlessly by human being. However, the necessity of recognizing large sets of faces either in challenging scenarios or in a short time has led to the emergence of research and development of computer systems able to automatically recognize face images. As stated by Phillips et al. [2011], the first face recognition system was developed by Kanade [1974] in 1973, and since then it has numerous applications, such as surveillance, forensics, human-computer interaction and biometrics.

According to Chellappa et al. [2010], there are three different tasks in face recognition relying upon which scenario it will be required: verification, identification and watch list. The *verification* task consists in an *one-against-one* matching problem. Here, the goal is to compare two face images in order to determine whether or not they belong to the same person. It can be applied, for instance, in a system allowing only authenticated persons entering in a room. So, one may claim an identity and his/her face image should be matched against the claimed identity face image in order to allow the access or deny. Figure 1.1 illustrates the face verification process.

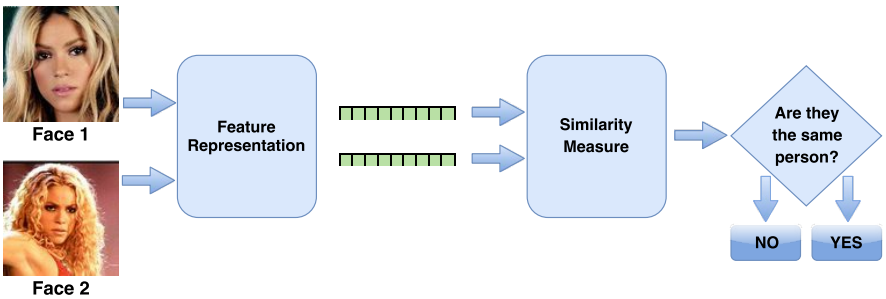


Figure 1.1: Overview of the face *verification* process. A similarity measure method is applied over feature vectors extracted from two face images. Then, a response value is thresholded to determine whether they belong to the same person.

In face *identification* task, we want to solve an *one-against-all* matching problem, comparing a query face against multiple faces in the enrollment database to associate the face identity to one of those gallery individuals. Figure 1.2 depicts the identification process.

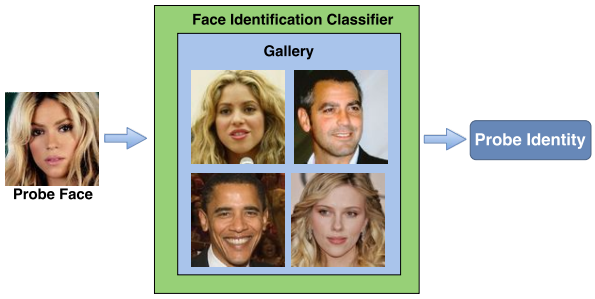


Figure 1.2: Overview of the face *identification* process. A probe face image is projected onto a classifier learned with gallery individual samples. It presents then the most likely identity of probe sample.

In the *watch list* task, also known as open-set recognition task [Wechsler, 2009], the goal is more than finding the most likely identity for a probe face. This task also considers the possibility of the query face does not belong to any individual enrolled in the gallery. In other words, the query face may belong to an unknown individual. Thus, although the goal is to determine the query face identity, a method for filtering faces which belong to gallery and non-gallery individuals is demanded to be applied first. Generally, the filtering method is based on confidence threshold which delimits the minimum similarity score necessary to consider a query as belonging to a known individual. Queries which do not attain that minimum value are then rejected (not identified). Figure 1.3 illustrates the watch list process.

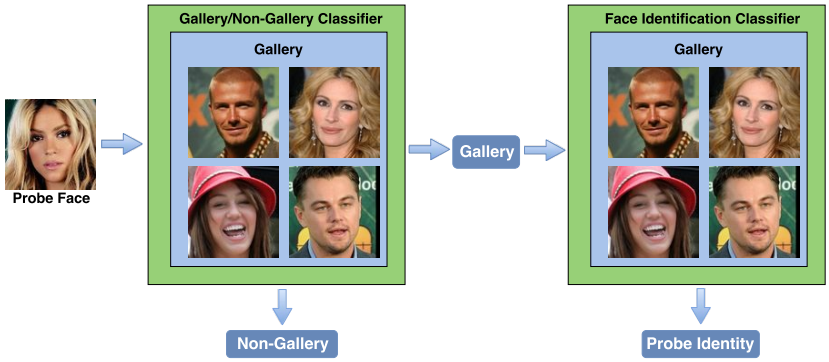


Figure 1.3: Overview of the watch list process. A probe face image is projected onto a classifier which detects whether it belongs to gallery or not. If so, its identity is predicted. Otherwise, the sample is rejected as unknown.

Numerous challenges are associated to face recognition, mainly in unconstrained scenarios. Face images can vary considering pose or illumination, occlusion, background conditions, scale/resolution, noise, blur and expression. These characteristics make recognition a more difficulty task since they lead to differences in face images from the same person. Thus, several methods to overcome these problems have been proposed [Hayat et al., 2017; Zhu et al., 2015; Tran et al., 2017; Hassner et al., 2015; Best-Rowden et al., 2014; Liao et al., 2014].

Besides these image acquisition aspects, specific traits regarding the scenario in which a system is being applied can also hamper face recognition task. Several literature approaches fail to identify a face in scenarios with a large number of individuals enrolled in the gallery, for example. Considering that traditional methods to determine the identity of a face image first train a model to each individual in the gallery and then compare a probe face against all individual models, the number of subjects may affect significantly the performance in terms of computational time. Other relevant aspect is the number of samples of a determined subject used in training stage. Usually, the more examples of a subject we have, the more accurate the approach is. Thus, scenarios with few subject samples available or a large number of subjects may hinder recognition task.

Most studies in the literature are focused on solving problems related to the aforementioned challenges considering a closed-set scenario. In identification task, for instance, they assume that all probe face images belong to a known individual. However, in a real-world application, this assumption may not be correct. When approaches encounter a probe face image belonging to a novel subject, one of the known classes/subjects would be chosen to be assigned to that image, while it should be rejected. On the other hand, in an open-set scenario, a probe face image may not belong to a subject in the gallery (watch list). In other words, approaches for watch list consider that a probe sample may be sometimes unknown. In fact, most probe images tend to be unknown in realistic scenarios.

Different from most approaches in the literature, we propose in this thesis approaches for combining a set of classification models trained in an *one-against-all* protocol, aiming at identifying face images in an open-set scenario. The models are based on Partial Least Squares (PLS) [Wold, 1985], which operates weighting features from images in order to discriminate throughout different classes, dealing satisfactorily with high-dimensional data and overcoming the frequent problem of lack of samples. We present an effective and simple technique for face recognition that handles unknown subjects, namely open-set face identification task or watch list. Specifically, we want to identify a face image from a gallery/known individual and reject if it belongs to a

non-gallery/unknown individual. We demonstrate that information produced by a set of simple *one-against-all* PLS models can be combined to reveal not only whether a sample is known but also its identity.

1.1 Motivation

The use of face as biometrics entails several advantages when compared to other biometric modalities, such as fingerprint and iris. In addition to its simplicity and non-intrusiveness, it can also be captured discreetly, in a reasonable distance from the target. According to the six biometric attributes considered by Heitmeyer [2000], facial features attained the highest compatibility in a Machine Readable Travel Documents (MRTD) system[1] which takes into account the number of evaluation factors, such as renewal, enrollment, public perception and machine requirements.

Advances in technology such as in surveillance systems and mobile phone cameras, the huge amount of face images on the Web, and the increasing demand for higher security have also enhanced face recognition relevance. Authentication of people to enter a restrict room can be cited as an example. It requires an accurate and straightforward approach giving access grant only to known individuals after identification, while rejecting unknown individuals, whose identity is not in the gallery subject list. A misidentification, in this case, may lead to access of intruders, unknown persons that have no permission to access the room, affecting significantly safety and security.

1.2 Objectives

This work aims at handling the watch list problem by proposing a novel and efficient technique which is able to deal with unknown subjects in the probe image set. The specific objectives here can be summarized into four. First, we want to propose a method with the goal of determining whether a face image is from a known individual and predicting its identity, or rejecting otherwise. Second, we intend to show that a set of binary *one-against-all* classifiers specialized in one class can provide effective results. Third, we intend to demonstrate that the use of the whole gallery subjects halved in positive set and negative set,as proposed in the literature, is unnecessary since we can achieve better results with models learned with a specific subject as positive and the remainder as negative. Finally, in the next chapters, we endeavor to show that a set

[1] http://www.icao.int/mrtd/overview/overview.cfm

of binary *one-against-all* classifiers is the best option to perform face identification in open-set scenarios.

1.3 Hypothesis

The main hypothesis in this work is that the vote-list (see Section 4.3) of a gallery probe image looks different from a vote-list of a non-gallery probe image. Specifically, we make the assumption that the position in the vote-list associated to the gallery probe identity would receive more votes (or have the highest response value) than the remainder. On the other hand, a vote-list for a non-gallery probe image would be similar to a balanced histogram, with no highlighted position.

Secondly, contradicting the work of Vareto et al. [2017b] which considers the entire gallery to train each model in a collection, splitting in half as positive set and half as negative set, we believe that a better option is to train *one-against-all* models specialized in one subject at a time. Thus, we train a set of Partial Least Squares classifiers using each subject as positive set and the remainder as negative, since it could bring more discrimination for the method, resulting in stronger models.

Thirdly, rather than building a preset number of models as proposed by Vareto et al. [2017b], we propose to fix the number of models in the number of gallery subjects which is generally a small number when compared to the number of unknown subjects. It provides better performance in terms of accuracy while reducing the training and testing computational time.

1.4 Scientific Contributions

Inspired by the work proposed by Vareto et al. [2017b] which provides an watch list approach using a determined number of models separately learned with all individuals of the gallery and does not perform face identification after rejecting or accepting a face image, this work proposes a complete pipeline to perform open-set face recognition setting the number of models to the number of gallery subjects. The reported results evidence the effectiveness of our approach in comparison with state-of-the-art works on well-known datasets. The main contributions of this work for watch list task are presented as follows.

- A complete face identification pipeline in open-set scenarios based on a set of one-against-all Partial Least Squares classifiers which suffice to discriminate gallery subjects.

- A fast approach since the number of models is set to the number of gallery individuals (which in real-world applications is much smaller than the number of non-gallery ones).
- An easy to deploy algorithm in which only gallery samples are necessary in training stage (closer to realistic scenarios).
- A novel manner to compute the threshold based on the difference of bins with the highest values in vote-lists which improves literature results.
- An extensive experimental evaluation, discussion of the proposed algorithm and comparison with state-of-the-art approaches.

The following publications, whose we are co-authors, comprise approaches to watch list and face verification respectively, and are related to this work. The former one, specially, has been used as a baseline for this work. After adaptations, the results obtained by this thesis outperformed the published ones not only in terms of accuracy but also in terms of computational cost.

- Vareto, R., Silva, S., Costa, F., and Schwartz, W. R. (2017). Towards Open-Set Face Recognition using Hashing Functions [2]. In International Joint Conference on Biometrics (IJCB).
- Vareto, R., Silva, S., Costa, F., and Schwartz, W. R. (2017). Face Verification based on Relational Disparity Features and Partial Least Squares Models. In Conference on Graphics, Patterns and Images (SIBGRAPI).

1.5 Thesis Roadmap

The remainder of this work is organized as follows. Chapter 2 presents a brief review of literature, analyzing methods regarding two tasks in face recognition (Section 2.1) (face identification and watch list) and methods to solve open-set recognition (Section 2.2) in different scenarios (object, cameras, handwritten digit, etc). In Chapter 3, basic concepts for a better understanding of the proposed approach are explored. The proposed methodology is detailed in Chapter 4 and a vast number of experiments is reported in Chapter 5. Finally, we conclude in Chapter 6, presenting final remarks and future works.

[2] Best Paper Runner-up Award at the International Joint Conference on Biometrics, 2017.

Chapter 2

Literature Review

According to Google Scholar[1], almost a thousand of face recognition algorithms were published, 34 patents were filed, and dozens of face recognition startups were founded in 2016. It demonstrates that there is a significant interest by the community in terms of research and development in face recognition. This fact discloses that face recognition still is a problem not solved, mainly on a global scale (where millions, even billions of identities are to be distinguished) and on unconstrained scenarios.

This chapter provides a brief review on the literature concerning face recognition and open-set recognition. A vaste analysis reviewing literature works on face recognition can be found in the book titled *Hand-book of Face Recognition* [Maltoni et al., 2009]. The chapter is divided into two parts according to the main problem addressed: face recognition (Section 2.1) and open-set recognition (Section 2.2). The former comprises two tasks within face recognition: face identification and watch list. This latest is an instance of open-set classification for faces. The latter focuses on open-set recognition works in general not necessarily based on face recognition.

2.1 Face Recognition

Face recognition has been extensively explored during the past years [Schwartz et al., 2010; Barkan et al., 2013; Best-Rowden et al., 2014]. In this section, some of the main works regarding face identification and watch list are compiled.

[1]https://scholar.google.com

2.1.1 Face Identification

Face identification is the task of determining who a probe face image belongs to, considering that this identity is included in the gallery, that is, the image necessarily belongs to a known individual. One of the major challenges in face identification is dealing with large galleries due to the need of comparing a probe sample with all individuals enrolled in the gallery. A second problem is the obstacles generated by image acquisition in an uncontrolled environment, which results in images with variation in terms of illumination, pose or expression. Schwartz et al. [2010] tackle the aforementioned problems with an approach based on PLS to perform multi-channel feature weighting. More specifically, the problems associated with uncontrolled acquisition are reduced by using a combination of low-level feature descriptors carrying different traces. Then, PLS, which works well with very high-dimensional data and very few samples, weights the features. Finally, the traditional OAA (One-Against-All) classification scheme is employed to model the subjects in the gallery. This scheme is further adapted to a tree-based structure in order to handle large galleries. Each internal node of the tree is a binary classifier based on PLS regression and is used to guide the search for the subject matching. This structure not only reduces the number of comparisons when we want to determine the identity of a probe sample but also allows for addition of new individuals to the gallery without the need of rebuilding all PLS models.

de Paulo Carlos et al. [2013] also face the challenges associated with the enrollment of new subjects in large galleries using PLS. They intend to make the process of rebuilding the models to take into account the new individual more efficient, replacing the traditional one-against-all scheme by one-against-some, in which only a subset of all subject is used as counterexample. A set of PLS models trained with a random subset are build for each subject. In the test stage, instead of projecting a probe image onto all models generated for all subjects, a priority queue allows for a search only on subjects models with high chance of being the correct match. A priority queue helps to maintain a small number of projections when a probe sample is presented to the system, resulting in a scalable approach to enroll new subjects to the gallery.

Usually, methods to pose-invariant face identification rely on either performing face alignment, so the face front is completely visible, or learning a face representation regardless the original face pose. The idea of combining both techniques is proposed by Tran et al. [2017] aiming at them to leverage each other. Their novel approach called Disentangled Representation learning-Generative Adversarial Network (DR-

GAN) comprises three new characteristics. The first one is the possibility of not only learning a generative and discriminative representation but also synthesizing images through the encoder-decoder structure of the generator. The second contribution is the disentangling ability from other face variations due to the pose estimation in the discriminator and the pose code provided to the decoder. Finally, DR-GAN can receive one or several images and generate a single representation or synthesized images. They provide a discriminative approach since the other variations have been explicitly disentangled, while keeping generality with synthetic images able to be classified as the original identity. The disentangled representation is used as the identity features and cosine distance metric computes the similarity between a new feature vector and a determined identity.

Deep neural network has been a technique frequently used in face identification approaches. Dou et al. [2017] propose to reconstruct the monocular 3D facial shape from a single 2D face image using a deep neural network. Different from several approaches in which a reconstruction and refining of 3D are performed by using the RGB image and an initial 3D shape rendering iteratively, an end-to-end model avoids the 3D rendering process. Instead, a single forward operation is used to predict the optimal morphable parameters. Its input consists in only an image of facial region-of-interest (ROI) detected. An additional fusion convolutional neural network (CNN) and a multi-task loss function are incorporated in order to improve reconstruction. With both components, 3D face reconstruction is divided into two sub-tasks, the neutral 3D facial shape reconstruction and expressive 3D facial shape reconstruction, and different types of neural layers are trained not only for these tasks, but also to predict an identity. The division of the problem in two tasks makes reconstruction task simpler while keeping robustness.

Also taking advantage of the CNN technique, Wang et al. [2017] introduce the Discriminative Covariance oriented Representation Learning (DCRL), a framework for identification when both gallery and probe samples are image sets. In their work, the goal is to train a CNN to project the images onto a target feature space which maximizes the discriminative ability of the covariance matrices calculated in this target space. Two different loss functions are proposed to encode discriminative ability: the Graph Embedding scheme and the Softmax Regression scheme. They are used to optimize the network, leading to not only image representation mapping but also to the set model classification. The method's versatility allows for an integration with existing covariance modeling methods.

Hayat et al. [2017] utilize a single CNN to simultaneously register, which consists in adapting face to a canonical frontal view, and represent faces to face identification

in unconstrained scenarios, with extreme pose variation. Videos and images, which are generally called a media template, are kept intact in the sense that features are not extracted. A video/image representation is encoded by the output of one of the CNN layers. Instead, gallery templates are represented by SVM models trained in a one-vs-rest protocol. Next, a Bayesian strategy fuses decisions of all medias in a query template. Template based methods are pertinent in applications in which there exist multiple media from individuals.

2.1.2 Watch List

Watch list consists in first detecting whether a face image belongs to a gallery individual and then determining its identity. Frequently, classification methods face a small number of samples per subject in training stage, which leads to poor information to build models. With the purpose of overcoming this problem, Li and Wechsler [2005] propose the called "Open Set TCM - kNN" (Transduction Confidence Machine - k-Nearest Neighbors). It provides an estimation of likelihood needed for detection tasks, based on the relation between transduction and Kolmogorov complexity. This transduction-based approach consists in rejecting a test sample by using k-nearest neighbors to derive a distribution of credibility values for false classifications. Thereafter, test sample credibility is estimated by iteratively assigning it to every class in the k-neighborhood. If the highest credibility value does not attain a certain threshold, defined in the training stage, face is rejected as unknown. Otherwise, sample is identified. Despite its non-parametric implementation and automatic threshold selection, the requirement of computing kNN between probe sample and all gallery samples makes this approach unfeasible since it results in a high computational cost.

Stallkamp et al. [2007] assume a more realistic scenario for determining the identity of people entering through the door of a laboratory, in which the individuals are not supposed to cooperate. The unconstrained environment results in problems associated with illumination, pose, expression, and occlusion. Since camera is recording continuously, input sequences are first detected, tracked and registered considering the approach described by Stallkamp [2006], which is based on local feature vectors, computed using block-based Discrete Cosine Transform (DCT) and fused to image classification. Feature extraction is based on DCT following the approach of Ekenel and Stiefelhagen [2005], which employs block-based DCT to non-overlapping blocks, leading to a compact representation and fast computation. Then k-nearest neighbor and Gaussian mixtures are evaluated in order to generate individual frame scores. Confidence scores from each classification are progressively combined to perform video-

based open-set face recognition of the entire sequence through a threshold defined in the training stage. Aiming at giving different weights to each frame score, three measures to evaluate its contribution are proposed: distance-to-model (DTM), distance-to-second-closest (DT2ND), and their combination. DTM takes into account how similar a test sample is to the representatives of the training set. DT2ND looks for reducing the impact of frames which provide ambiguous classification results. The idea is that reliable matching requires the best match to be significantly better then the second-best match. Finally, they state that determining whether a person is known or unknown is a more challenging task than determining who the person is. In addition to the probably high cost of kNN computation, the modeling with Gaussian mixtures can also result in a computationally expensive and unstable training.

The most natural manner to perform open-set face recognition may be to consider this task as a multiple verification problem. In other words, if a probe face image does not match any of the gallery individuals in verification task, this sample should be reject as impostor. Ekenel et al. [2009] formulate the open-set face recognition problem as a multiple verification task. A real-time system is designed to act as a visitor interface. A person first looks at a monitor to read a welcome message before knocking on the door. Meanwhile, images from a webcam embedded in the monitor are captured to further person identification without explicit cooperation. To detect faces, a Haar-like features based cascade of classifiers [Jones and Viola, 2003] is utilized. Then, faces are aligned and rescaled. Sample augmentation are also applied to mitigate the effects of improper registration. The representation of faces consists in a global feature vector based on DCT (more details are given in [Ekenel and Stiefelhagen, 2006]). The classification stage which consists of multiple verifications (one per subject) employs the use of Support Vector Machines (SVMs), as shown in studies by Schölkopf et al. [1999], combined with a 2-degree polynomial kernel. Their experiments demonstrate that an increase in the gallery size can impact negatively the approach performance. Besides, classification with SVM may not perform effectively when only a low amount of training data is available.

[Wright et al., 2009] propose a method to recognize faces using frontal images with variation in expression and illumination in conjunction with occlusion and disguise. Different from several approaches based on traditional dictionaries, they tackle the problem of recognizing face images by representing them as a combination of training sample images. A linear combination can be employed if a sufficient number of samples are available in the training set for the correct testing class. This representation is called sparse since it involves only a small part of the training data and can be effectively obtained via a $l^1 - minimization$. Determining the sparsest representation for a test

sample automatically results in deciding its class. Besides face identification, they also face the watch list problem using sparse representation. Test samples which activate samples belonging to different classes are rejected since they tend to belong to unknown subjects. Otherwise, they are identified according to the samples identity. Since their approach relies greatly on the training samples to represent a probe face image, it may fail in scenarios with few samples per class or with a small number of poses described in training data. The proposed approach, on the other hand, can handle effectively situations with few samples per class in training due to the employment of Partial Least Squares. Besides, the VGGFace CNN descriptor can represent accurately the main features of a face image despite the challenges posed by pose or occlusion, for example.

The way human beings reject or identify a person can be an inspiration for automatic face recognition. With this purpose, Kamgar-Parsi et al. [2011] have designed an approach to identify faces basing on the human perceptual ability. Decision regions in the face space of target persons are identified by generating two large sets of borderline images, images morphed toward to become acceptable borderlines and unacceptable borderlines, projecting just inside and outside of the decision region. With these sets in hands, a feed-forward network classifier is trained for each person belonging to the watch list in order to identify the region corresponding to each individual in face space. In the test phase, if the probe image is projected inside the acceptance region of a classifier, it will be identified as the target person. If none of the classifier accepts a probe image, it likely belongs to an unknown individual, and thus it is rejected. Although their work provides an effective approach, the type of networks used requires a huge amount of data. Otherwise, the classifier would overfit.

Focusing only on the task of determining whether a probe sample is known (enrolled in the gallery), dos Santos Junior and Schwartz [2014] propose five approaches. Four approaches consider responses from face identification, and one consists in comparisons between known samples and a background set. The major difference between approaches is the feature explored in data. Among the four approaches considering responses from face identification, one is based on the distribution of response considering the Chebyshev inequality, and three are based on the separation margin between identification responses, one considers a single SVM classifier for all known subjects, other considers a SVM for each known subject, and the last is based on least squares. The Chebyshev inequality-based method presents the best performance when gallery has a large number of individuals, since the remainder presents a description of all enrolled individual, which results in difficulty to find similar common attributes when the number of subjects is large. Identification responses are

obtained by applying the method described by Schwartz et al. [2012], based on Partial Least Squares (PLS), as it presents good accuracy, scalability and is publicly available. A drawback associated with their methods consists on a huge decrease of accuracy as the gallery size increases. This aspect makes their approaches not suitable for real-world scenarios.

While most of approaches in open-set face recognition consider a single media (i.e. image) as probe in order to find a person of interest in a gallery, Best-Rowden et al. [2014] examine a media collection (image, video, 3D model, demographic data, and sketch). According to them, different media can carry complementary information and mitigate challenges associated with unconstrained scenarios in the building of candidate list for a person of interest. Each media of a probe is matched to the gallery and scores are integrated using different schemes of score level fusion. Besides, quality measures for fusion and video frame selection are utilized to determine when the fusion is helping to boost accuracy recognition. Their approach can be useful when images for face matching have low quality which makes identification difficult. In addition, the use of videos can improve face recognition since it provides diversity in face pose samples. However, their work relies on a good face quality measure and 3D model construction.

The study of Liao et al. [2014] proposes a new benchmark protocol under LFW face dataset concerning large-scale unconstrained face recognition evaluation under not only open-set identification but also face verification with focus on low False Acceptance Rates (FARs) and high number of impostor probes in test phase. Twenty one face recognition approaches are evaluated based on the combination of three different features and seven learning algorithms. By using this protocol, authors have shown that the large-scale unconstrained face recognition problem is still unsolved and efforts to advance research in open-set identification concerning face representation and learning procedure are necessary.

Ortiz and Becker [2014] address the problem of open-set recognition in web-scale datasets. In their study, face images are totally unconstrained since they are collected from Facebook social network. Their objective is recognizing specific individuals in a picture and reject everyone else, which consists in an application of open-set face recognition. They propose a Linearly Approximated Sparse Representation-based Classification (LASRC) algorithm. SRC presents a principle that a given test image can be represented by a linear combination of a subset of images from a large dictionary of faces. It performs slowly, needing seconds per face, due to l^1-minimization. The LASRC proposes an approximation of l^1-minimization, using a linear regression, technique also employed in this dissertation. They take advantage of the speed of least-squares

and robustness of sparse solutions such as SRC. However, since they use local feature descriptors, a preprocessing stage is necessary before feature extraction. Therefore, performance also relies on an accurate preprocessing procedure.

Vareto et al. [2017b] propose the Partial Least Squares Hashing (PLSH), a scheme combining hashing functions and classification methods in order to predict when probe samples belong to gallery set based on the methods proposed by [dos Santos et al., 2015, 2016]. A preset number of PLS or Fully Connected Network (FCN) models trained with gallery face samples makes the probe prediction. In training stage, the set of gallery individuals is randomly splited into two parts, the positive and negative sets, that are used to build a new classification model. The number of models accords with the number of times we split the gallery set. In testing stage, when a probe is projected onto a model, a vote-list, containing a position per gallery individual, is incremented in the positions corresponding to the positive set. After all models projection, if the vote-list has a position with value much greater than the remainder, sample is predicted as known. Otherwise, sample is predicted as unknown. The idea behind this is the assumption that known probe may be similar to only one individual (the correct identity) in the vote-list. Therefore, the vote-list would be a histogram with a highlighted bin. This assumption seems to be correct, since the results show its effectiveness. However, they only evaluate the former step of watch list in which known/unknown samples are detected. The identification of samples predicted as known is not performed and evaluated. Furthermore, the fact that each models is trained with half of gallery in each set, positive and negative, builds models less discriminative than using the one-against-all protocol, as in the proposed work. Figure 2.1 depicts the approach of [Vareto et al., 2017b].

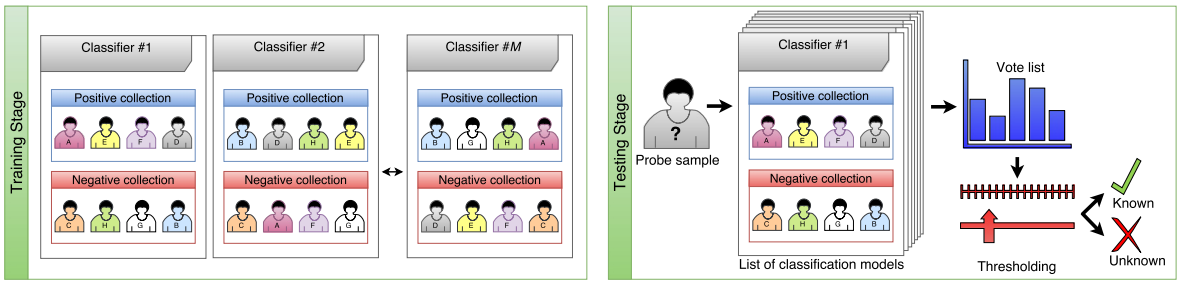


Figure 2.1: A depiction of binary classifiers embedded into hashing functions: *Training:* Gallery samples are randomly partitioned into positive and negative sets M times, and different models are learned with different subjects in each subset. *Testing:* A query face image is projected onto all classification models and their responses are used to increment a vote list. Finally, if vote-list has a position with value much greater than the remainder, query image is considered as being from a known individual. Figure extracted from [Vareto et al., 2017b].

Liu et al. [2017] state the problem of deep face recognition in open-set scenario as a problem in which face features have smaller maximal intra-class distance than minimal inter-class distance considering a determined metric. Since few works have effectively dealt with this problem, they propose the angular softmax (A-Softmax) loss which allows for CNNs to learn angularly discriminative features (SphereFace), replacing the traditional softmax loss function. A-Softmax loss imposes discriminative constraints on a hypersphere manifold whose angular margin size is a parameter to be set. Besides, it takes into account the prior that faces lie on non-linear manifold. Considering that CNN has already been trained using A-Softmax in a different dataset, the testing stage consists in first presenting a probe image and its flipped images to the network, extracting deep features (SphereFace) from the FC1 layer, and concatenating them. Then, the score (metric) is computed by the cosine distance of two features, the one extracted from the testing sample and other from the training set. The nearest neighbor classifier is used to face identification, presenting a high computational cost, since score is needed to be computed between probe image and all gallery subjects.

In this section, we presented works related to the watch list task. Some of them are based on the nearest neighbor classifier to compute the distance between a probe image and the gallery training. This computation may cause an expensive computational cost. In our approach, the evaluation consists in performing a single dot product between feature vector and each PLS regression vector which is simple to compute. Concerning the training time, our approach is also fast since we train an one-against-all PLS classifier per subject. Different from approaches which are highly impacted by an increase in gallery size, we demonstrate that this does not affect significantly our approach. Besides, different from approaches such as those based on SVM or neural networks that require large amounts of training data, our method can also handle low amount of data since PLS can handle few samples. Furthermore, the proposed approach does not require a preprocessing procedure before feature extraction nor depend on the accuracy of other approaches. Finally, different from aforementioned approaches, the number of models we use is limited to the number of gallery individuals, and we perform the complete watch list procedure, that is, after we predict a probe as gallery, we also determine its identity.

2.2 Open-set Recognition

Aside from face identification, open-set recognition, in general, has been studied and employed in many applications. The work of Scheirer et al. [2013] presents

a formalization for the open-set recognition problem as a constrained minimization problem. With base on the traditional 1-class and binary SVMs, the "1-vs-set machine" approach is presented, aiming at sculpting a decision space from the marginal distances of a 1-class or binary SVM with a linear kernel. Thus, instead of one plane marking the decision boundary, as in the traditional SVM, two planes are defined to minimize a determined equation in a greedy way and then a refining step is applied. Experiments are performed not only in object recognition but also in face verification. The binary version presents the best accurate results. Despite their good performance, SVMs require not only a high computational time but also a considerable number of samples per subject in order to attain good accuracy which is not always possible. Figure 2.2 illustrates the approach proposed by Scheirer et al. [2013].

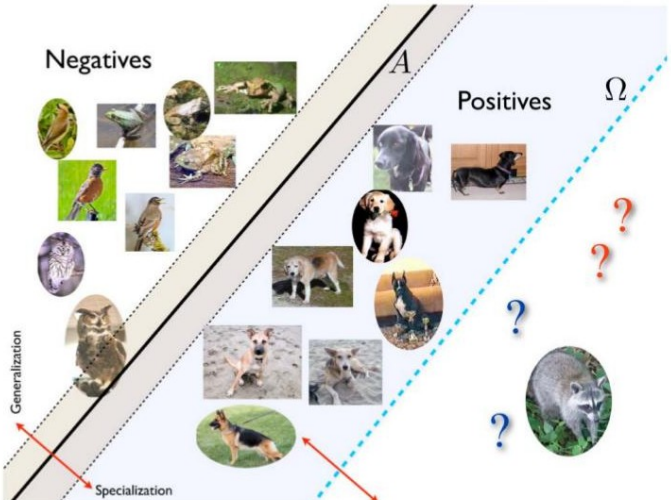


Figure 2.2: The two hyperplanes generated by 1-vs-set machine. Plane A maximizes the SVM margin making the "dog" class a half-space, which is mostly open space. The second plane Ω is added by 1-vs-set machine, and an optimization adjusts A and Ω in order to balance empirical and open space risk. Figure extracted from [Scheirer et al., 2013].

As an extension of the aforementioned work which proposes the term "open space risk" to account for the space beyond the reasonable support of known classes, the work proposed by Scheirer et al. [2014] applies this concept to accommodate non-linear classifiers in a multi-class scenario. A new open-set recognition model named Compact Abating Probability (CAP) is presented, in which the likelihood of a class membership is decreasing with the increasing of the distance between a sample and the known data toward open space. Despite the CAP formulation can be applied along with multiple algorithms in literature, they also propose the Weibull-calibrated SVM (W-SVM) algorithm, that combines properties of statistical extreme value theory for

determining scores in both one-class and binary support vector machines (SVMs).

Regarding the problem of open-set source camera attribution, de O. Costa et al. [2014] propose a work relying on attributing the most likely camera used to take a determined photo. Different from most works in literature, they do not make assumption all the cameras are known, that is, the photo could have been taken by an unknown camera. To solve this problem, a decision space is modeled taking advantage of the few known cameras and a Decision Boundary Carving technique, which moves the decision hyperplane during training, is employed. Their method can be seen as a variation of the approach proposed by Scheirer et al. [2013] which considers a non-linear kernel and moves a single plane in order to have a more accurate region.

Different from the majority of methods in open-set recognition, Bendale and Boult [2015] handle the possibility of at testing time seeing a sample from an unknown class and adding it as a new class to the model. This problem is introduced and formalized by them as "Open World Recognition". Their contribution relies not only on proving that thresholding sums of monotonically decreasing functions of distances in linearly transformed feature space can balance the "open space risk" and empirical risk, but also on presenting the Nearest Non-Outlier (NNO) algorithm which allows for efficient model evolving, while detecting outliers, managing open space risk and adding new object classes incrementally. Moreover, they also combine existing algorithms with their theory.

Visual recognition problems have taken advantage of deep networks in many aspects [Simonyan and Zisserman, 2014; He et al., 2014; Ng et al., 2015]. They can be mostly useful in applications where human being are not effective in recognizing since they consider parts in images which are unmeaningful for humans. The development of an open-set deep network was only first approached in the work of Bendale and Boult [2016]. The idea is to adapt deep networks by introducing a new model layer, which replaces the traditional Softmax, called OpenMax, to compute the likelihood of an input being from an unknown class. Scores from the penultimate layer of deep networks (the fully connected layer before Softmax) are used to estimate if the input is "far" from the known training data. Basically, the OpenMax algorithm measures distance between an activation vector for a probe image and the model vector for the top few classes estimating a probability of being unknown. Pre-trained networks from the Caffe Model-zoo on ImageNet 2012 validation data are used to evaluation.

Zhang and Patel [2016] perform the recognition of digits, faces and objects in an open-set scenario using the Sparse Representation-based Classification (SRC) method. This approach identifies a test sample by seeking its sparsest representation in terms of training data. However, its natural formulation does not take into account the

open world assumption. Thus, in their work, they extended the traditional SRC to the perform open-set recognition. The idea is to summarize the open-set recognition problem into a set of hypothesis testing problems by modeling the tail distributions of the matched and non-matched reconstruction errors using the Extreme Value Theory (EVT). The confidence scores associated to the tail distributions of a probe sample are fused in order to determine its identity.

The approaches addressed in this section intend to solve open-set recognition using an "open-risk space", the SVM classifier or neural networks. As mentioned before, SVM presents several limitations, mainly associated to the amount of data or even with the computational time. In addition, SVM also presents difficulty in handling problems with unbalanced amount of samples per class. Neural networks provide impressive accuracy but also present drawbacks related to few samples. Training a neural network with few samples may cause overfitting in the data. Other aspect with respect to neural network training is the also expensive computational cost.Different from these approaches, we employ the Partial Least Squares classifier which can handle few and unbalanced amount of data while also keeping a low computational cost.

Chapter 3

Background Concepts

In this chapter, we present and describe techniques and concepts associated to the proposed work. Section 3.1 presents the Partial Least Squares (PLS) technique employed in this work. Then, Section 3.2 describes the One-Against-All (OAA) protocol used along with PLS.

3.1 Partial Least Squares

Partial Least Squares (PLS) [Wold, 1985] is a widely explored regression method that models the relationship between features by means of latent variables. Several works in literature have employed PLS in different tasks, such as face identification [de Paulo Carlos et al., 2013], face verification [Vareto et al., 2017a], and open-set face recognition [Vareto et al., 2017b]. The popularity of this approach is associated to several advantages related to its characteristics. PLS presents robustness in cases in which the feature vector is a combination of different feature descriptors, it handles effectively with high dimensional feature vectors, it is robust to unbalanced classes and able to train with few samples. In this work, we employ PLS as a classification approach, which produces high accuracy even in few-samples scenarios with low computational cost for projecting samples (a dot product between the feature vector and regression coefficients provides the response) onto the models and also training a model.

PLS regression can be formally defined as follows. Let $X_{n\times d}$ be the zero mean matrix representing n samples in a d-dimensional space of features and Y_n be a zero mean matrix which represents the label class of data. First, the decomposition of $X_{n\times d}$

and Y_n estimates a p-dimensional latent subspace by

$$\begin{aligned} X &= TP^T + E, \\ Y &= UQ + F, \end{aligned} \tag{3.1}$$

where T and U are both $n \times p$ matrices and denote latent variables from feature vectors and response values, respectively. Matrix P has dimension $d \times p$, vector Q is p-dimensional and both represent loadings. Residuals of transformation are represented by a matrix E, with dimension $n \times d$, and F, a n-dimensional vector. We can find P and Q computing the maximum covariance between U and T [Rosipal and Krämer, 2006]. For this purpose, we used in this work the Nonlinear Iterative PLS algorithm (NIPALS) [Wold, 1985] (described in Algorithm 1), which estimates the maximum covariance between the latent variables $T = \{t_1, ..., t_p\}$ and $U = \{u_1, ..., u_p\}$ using the matrix $W_{d \times p} = \{w_1, ..., w_p\}$ which satisfies

$$arg \max_{(t_i, u_i)} [cov(t_i, u_i)]^2 = arg \max_{w_i} [cov(Xw_i, Y)]^2, \tag{3.2}$$

such that $|w_i| = 1$.

Algorithm 1 : NIPALS($X_{n \times d}$,Y_n,p)

for $i \leftarrow 1$ to p **do**
  start u_i randomly or with some column of X
  repeat
    $w_i \leftarrow X^T u_i / ||X^T u_i||$
    $t_i \leftarrow X w_i$
    $q_i \leftarrow Y^T t_i / ||Y^T t_i||$
    $u_i \leftarrow Y q_i$
  until *convergence*
  $b_i \leftarrow u_i t_i / ||t_i||$
  $p_i \leftarrow X^T t_i / ||t_i||$
  $X \leftarrow X - t_i p_i^T$
  $Y \leftarrow Y - b_i (t_i p_i^T)$
return T, P, U, Q, B

Up to this point, PLS has performed a dimensionality reduction to the data representation of X. However, it can also be used for regression [Guo et al., 2011] by applying the matrix of weight W, obtained in the previous Equation 3.2, on the feature vector v_i. For this purpose, we first estimate the regression coefficients vector β:

$$\beta = W(P^T W)^{-1} (T^T T)^{-1} T^T Y. \tag{3.3}$$

Then, the regression response y_{v_i} for a zero-mean probe feature vector v_i can be computed by

$$y_{v_i} = \bar{y} + \beta^T v_i, \tag{3.4}$$

where $\bar{y}$ denotes the average of y. Constructing a PLS model can be defined as estimating the necessary variables to y_{v_i} prediction which are β, v_i and $\bar{y}$. Figure 3.1 illustrates the separation achieved by PLS for a particular feature space. The left plot shows the initial arrangement of data in three dimensions. The figure on the right illustrates the data after the use of PLS, which reduces the dimension to two, while keeping a good discrimination.

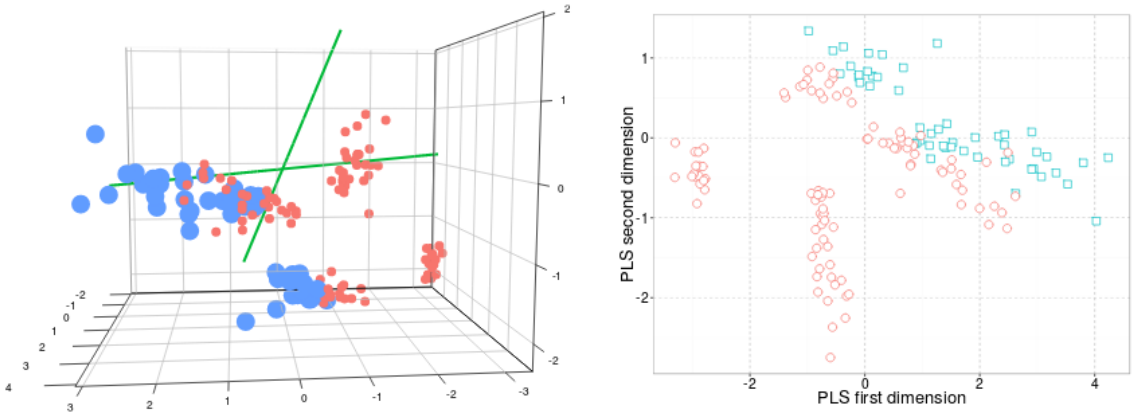


Figure 3.1: Distribution of data before and after PLS projection. The figure on the left depicts the initial 3-dimension data from two different classes, represented in blue and red.In green, we see possible manners to separate the data. Figure on the right depicts data after PLS projection. PLS provides a reduction in the dimension number while keeping a good separation of data. Figure extracted and adapted from [dos Santos Junior, 2015].

3.2 The One-Against-All Protocol

The one-against-all (OAA) protocol [Bishop, 2006] is a strategy for multi-class classification. It consists in training a single classifier - or model - per class. In other words, each model is built by considering samples of a single class at a time as positive and all the remaining samples as negative. This technique requires from the classifier a real-valued confidence score for the decision, rather than just a discrete label since it can lead to ambiguities. Besides, a posterior strategy is also necessary to combine the scores in order to determine the correct label. A natural strategy would be simply

assuming that the model which produces the highest response value determines the predicted label. This strategy is used in the identification stage of our approach.

The OAA Partial Least Squares (PLS) consists in increasing discriminability between classes by modeling each class against the remainder as counterexamples. PLS predicts the regression coefficients based on the discrimination ability of the features. This protocol implies in a PLS model per class. As a drawback, we can mention the necessity of rebuild all models if a new subject is enrolled. To perform a fair comparison with the literature, the OAA PLS is employed in the face identification phase of all approaches tested in this work, including the proposed one. Figure 3.2 illustrates the training data separation using an one-against-all protocol.

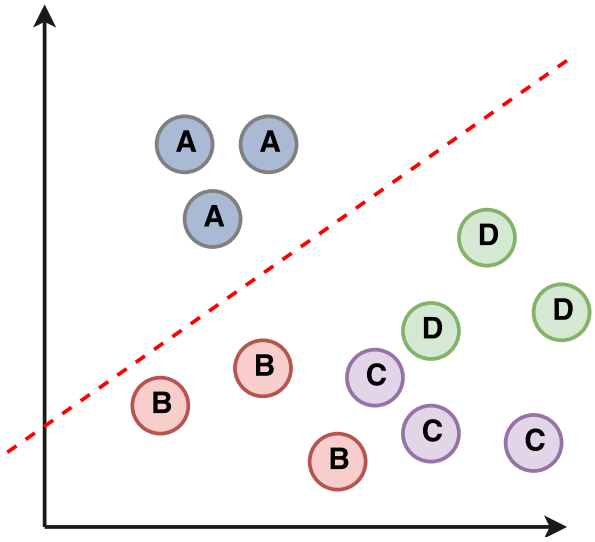


Figure 3.2: The training phase using an one-against-all scheme. This protocol consists in training a model to discriminate between a specific class and the remainder. In other words, in this example, we want to estimate the model that best separates class A from the remaining classes (B, C and D).

Chapter 4

Methodology

This chapter describes the approach proposed in this work with the purpose of performing watch list. A set of PLS models is used to better discriminate and enhance overall performance. PLS presents numerous advantages when compared to other traditional classification approaches. The fact that it maps features from the original space to a more compact space makes classification more efficient, since usually features hold redundant information. PLS also preserves effectiveness when we have a low and unbalanced number of training samples. Furthermore, evaluation with PLS is fast due to the need of only a single dot product evaluation.

Advantages of the proposed approach when compared to other methods in literature include the fact that we do not make a direct comparison with all the gallery samples. Instead, we learn models which vote for the likely identity of probes using a score, resulting in a faster performance. Additionally, the combination of binary PLS classifiers in our approach attains very competitive performance on the detection of gallery subjects, offering simplicity while keeping good discriminability. Finally, we also benefit from a straightforward approach which attains good accuracy compared to literature state of the art.

This method is inspired by the approach developed by Vareto et al. [2017b], which also proposes an embedding of PLS classifiers. However, we present modifications in many aspects to promote improvements in terms of accuracy, while requiring a lower number of models. Instead of training all models with all gallery subject samples split into two sets, we train models containing one subject against the remainder. The positive set contains one subject and the negative set the remaining subjects. Thus, we learn models specialized in one subject. Another difference is that our approach sets the number of models to the number of subjects in the gallery. On the other hand, in their work, the number of models is a parameter to be set and they present effectiveness

when this number is much higher than the number of subjects in the gallery. Therefore, we reduce both the computational time and the memory required for store the models.

Different from several works in the literature, we present herein the complete procedure for identifying a person in an open-set scenario. It first checks whether a probe face image belongs to a known subject, and if so, it determines its identity, which is also a difference compared to the work proposed in [Vareto et al., 2017b]. We divide the procedures into four different stages. We extract features from all samples in the dataset and partition data in order to have gallery training and gallery and non-gallery testing data. Second, we train a set of PLS models with gallery training data. Then, we project probe samples onto the trained models to obtain a vote-list and determine whether a probe sample belongs to a known subject. Samples which correctly meet this requirement are projected onto One-Against-All PLS classifiers in order to determine their identity which is associated to the classifier with maximum response.

The complete procedure is presented and detailed in the next sections. Section 4.1 describes the partitioning and representation of data. Sections 4.2 and 4.3 address the training and the test phases, respectively. The final step, which is the identity prediction for images correctly matched as gallery using OAA PLS, is tackled in Section 4.4.

4.1 Representation and Partitioning Stage

Our method is based on four canonical procedures: feature extraction, training, testing and identification. Initially, subjects of dataset are randomly split into two balanced sets: known (gallery) subjects and unknown (non-gallery) subjects. Samples from known subjects are also equally partitioned into training and testing data. Features are then extracted for all dataset samples using the VGGFace CNN descriptor [Parkhi et al., 2015], method detailed in Section 5.1.2. Thus, the input of the training stage (described in the next section) consists in features extracted only from known subject samples. On the other hand, the test stage receives features extracted from both known and unknown subject samples.

4.2 Training Stage

Let gallery subjects be represented by $S=\{s_1, s_2, s_3, s_4, ..., s_{n-1}, s_n\}$, where n is the number of subjects in gallery. We train a PLS (method described in Section 3.1) model per subject s_i with features extracted from gallery samples in training data as

follows. A PLS_i model is a classifier trained with samples of subject s_i in the positive set and samples of the remaining subjects $\{s_1, s_2, .., s_{i-1}, s_{i+1}, s_{i+2}, ..., s_{n-1}, s_n\}$ in the negative set. In other words, each PLS_i classifier models a specific gallery subject s_i against the remaining gallery subjects. We perform the training of PLS_i models until we have one model per subject s_i. The idea is to train specific models able to discriminate between a specific subject, the positive set, and the remainder, the negative set. Different from the work proposed by Vareto et al. [2017b], we do not create a set of models, with a preset size, trained with the whole gallery partitioned equally. Alternatively, each subject is trained against the remainder in a model. Thus, the number of models learned is equal to the number of gallery subjects. Figure 4.1 depicts the training stage of the proposed approach.

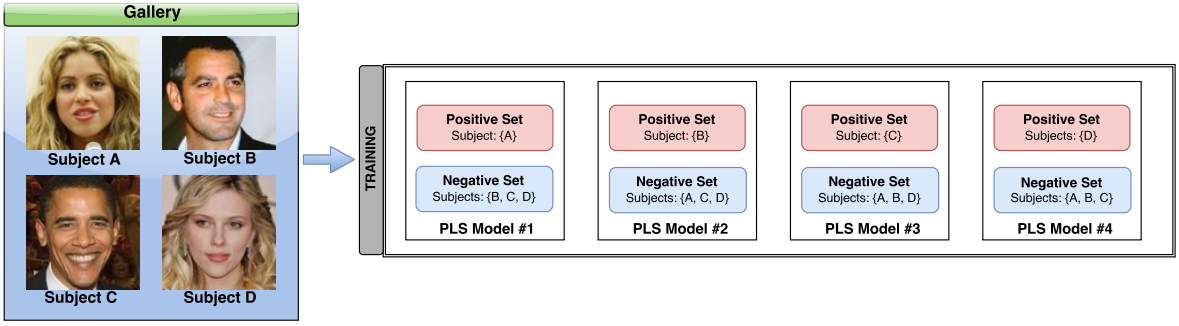


Figure 4.1: The training stage of the proposed approach. Face images of gallery subjects are used to learn a set of PLS classifiers. Each model utilizes a different subject as positive set against a negative set constituted by the remaining subjects. Models generation is finished in the moment we have a model per gallery subject.

4.3 Test Stage

We start the test stage by creating a vote-list v with length equals to n, the gallery size, replete of zeros. This list is similar to a histogram in which each bin represents the response values associated to a gallery subject. After projecting a probe image onto all PLS_i classifiers, if the vote list has a highlighted bin, probe probably belongs to the subject corresponding to it. Otherwise, probe does not match any individual, being then predicted as unknown.

Higher responses are provided by models trained with the probe identity in the positive set. Lower responses are obtained by models trained with the non-matching identities in the positive set. If an unknown probe is presented to the models, the responses obtained would not follow any rule. Probe can be similar to any gallery subject, but none in particular. Thus, the vote-list would be balanced, that is, probe

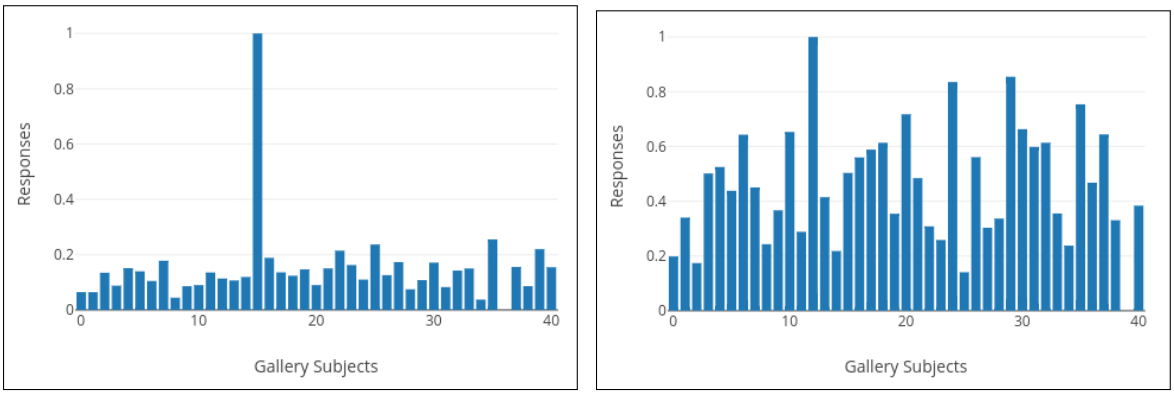
would always be similar to a different subject when projected onto each model. If a known probe is presented to the models, only the model using its identity in the positive set would provide high response.

Considering all PLS_i models have already been trained, a query face image is presented to each of them and their response values are added to the position i of the vote-list v_i which corresponds to the identity in the positive set of the corresponding model. In other words, each model provides a score that encodes the similarity between the probe and the subject in the positive set used to train the model. It can be high if probe matches the positive set or low if probe does not match the positive set.

Aiming at having values between 0 and 1, we normalize the vote-list. So, the threshold will not be influenced by very high and low response values and we can determine more precisely a vote-list threshold that allows for accepting known probe images and rejecting unknown. Equation 4.1 portrays how to obtain each position $\mathbf{z}_i$ of the normalized vote-list $\mathbf{z}$.

$$\mathbf{z}_i = \frac{\mathbf{v}_i - min(\mathbf{v})}{max(\mathbf{v}) - min(\mathbf{v})}, \tag{4.1}$$

where $\mathbf{v}_i$ is the i-th position of the original vote-list $\mathbf{v}$, for $i = 1, 2, ..., n$, and $min(\mathbf{v})$ and $max(\mathbf{v})$ are the minimum and maximum value in $\mathbf{v}$, respectively. Figure 4.2 exemplifies the normalized vote-list for both cases, known *(a)* and unknown *(b)* probe samples.



(a) Gallery probe (b) Non-Gallery probe

Figure 4.2: The vote-list for known *(a)* and unknown *(b)* probe images from FG-NET dataset. When the probe image is from a subject registered in the gallery set *(a)*, vote-list presents a position with value much higher than the remainder. On the other hand, when the probe image is from an unknown subject *(b)*, we obtain a vote-list with balanced values since probe image is not similar to any gallery subject in particular.

Finally, we threshold the normalized vote-list with the purpose of identifying whether it presents a position with a highlighted value, and therefore, probe is predicted

as known. Otherwise, probe is rejected as unknown. The method we use to perform this procedure is described in Section 5.2.3. Figure 4.3 summarizes the test stage of the proposed approach.

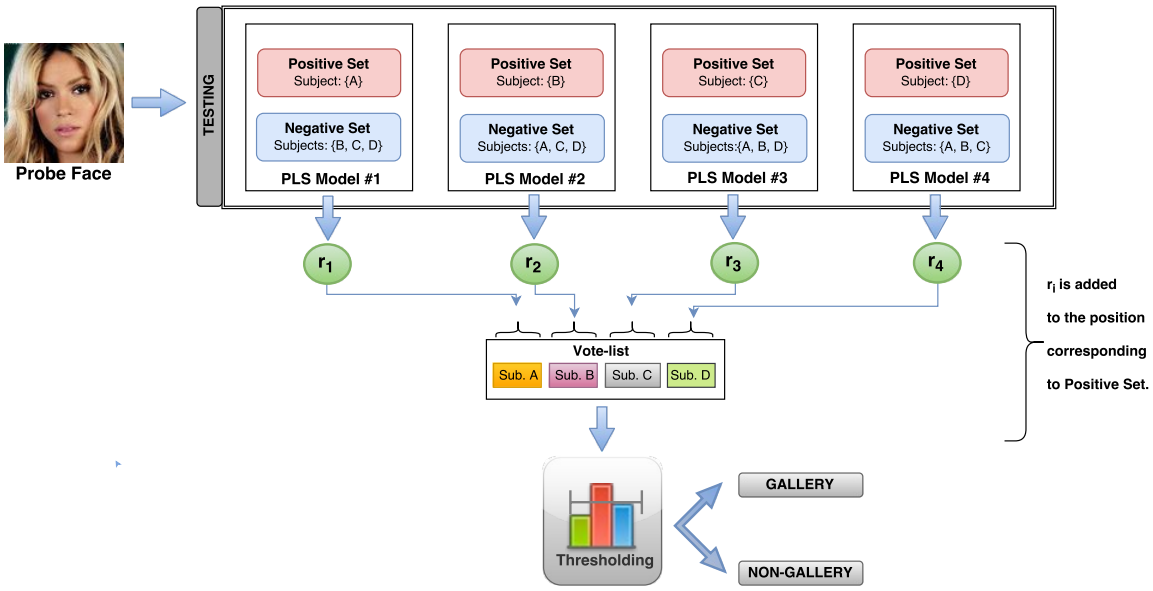


Figure 4.3: The test stage of the proposed approach. We first project a test sample onto the set of PLS models learned in training stage. Each model provides a response r_i related to its similarity to the corresponding positive set. Then, we add the response r_i to the vote-list position corresponding to the positive set of model i. We repeat this proceeding for all models. Finally, a threshold applied in the normalized vote-list predicts probe image as known or unknown.

4.4 Face Identification

The PLS with the One-Against-All protocol (detailed in Section 3.2) is employed in the face identification phase for all approaches compared in the Chapter 5, since for a fair comparison this stage may be fixed and only the stage of rejecting or accepting a face image may be variated. Taking into account a probe image is predicted as belonging to a gallery subject, one may want to determine its identity. The same training set used in the training stage (Section 4.2) is now used to learn OAA PLS models. We learn a PLS model per gallery subject, considering the remainder as negative set. The test phase consists in projecting gallery probe images onto OAA PLS models, aiming at determining which model provides the highest response value. The PLS model that provides the maximum response has the predicted identity in the positive set. In other words, probe image is labeled according to the positive set of the PLS model providing

the highest response. Figure 4.4 depicts the procedure of face identification using OAA PLS models.

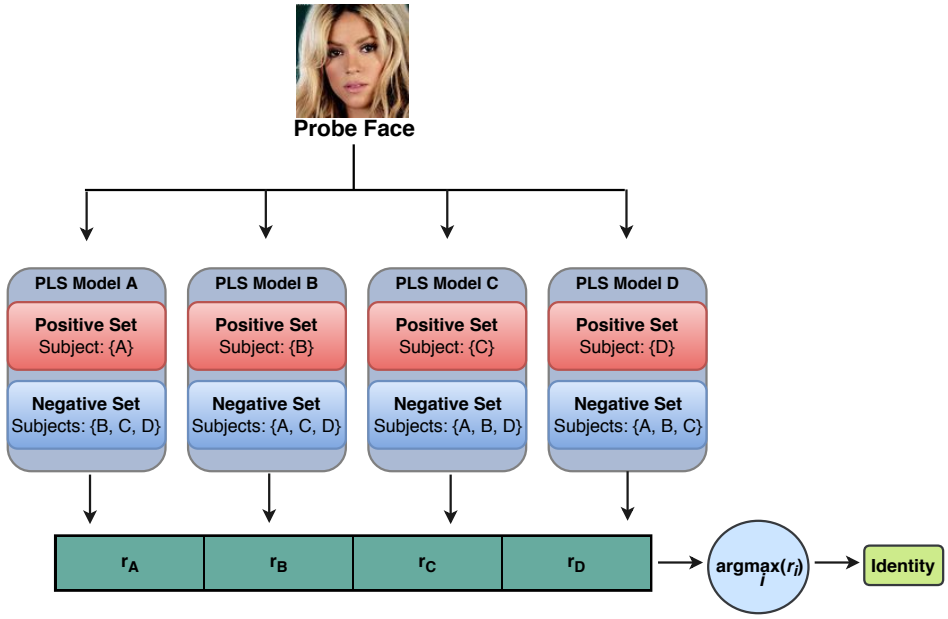


Figure 4.4: Face identification of a known probe image. PLS models trained for each subject in a OAA scheme are used to obtain response values in the projection of a known probe image. The model presenting the maximum response value corresponds to the predicted label.

Chapter 5

Experiments

This chapter provides an analysis of the experiments performed with the proposed approach and a comparison with literature methods. First, Section 5.1 details the datasets, the feature descriptor to represent each image and evaluation metrics used for face identification. Second, Section 5.2 describes the evaluation protocol, baseline approaches, the use of background set and the threshold selection. Third, Section 5.3 presents the results using Receiver Operating Characteristic (ROC) and open-set ROC curves. Fourth, Section 5.4 compares our method with the work we are based on, the PLSH [Vareto et al., 2017b]. Finally, Sections 5.5 and 5.6 reports a comparison with literature approaches using lower false positive rate (FPR) and false alarm rate (FAR) and an evaluation on the variation of the known set size, respectively.

5.1 Experimental Setup

This section presents details on the datasets: FRGCv1, PubFig83, FG-Net Aging and Pubfig, the VGGFace CNN feature descriptor and the ROC and open-set ROC evaluation metrics that we use to evaluate the approach proposed in this work.

5.1.1 Datasets

To validate the effectiveness of our method, we select datasets with different characteristics mostly associated to the acquisition process. They vary from frontal images taken in total controlled scenarios to images in the wild collected from Internet, presenting variations including illumination, pose and expression. We evaluate our method on four datasets: FRGCv1, Pubfig, Pubfig83 and FG-Net Aging. The first is a

well-known face recognition dataset and the others are more recent and unconstrained datasets.

Face Recognition Grand Challenge v1.0 (FRGCv1) [Phillips et al., 2005]: It consists in more than five thousand images of 152 subjects, presenting variation in two facial expressions (smiling and neutral), distributed in six experiments. We only evaluated the proposed method on two of them (*experiment two* and *experiment four*) and use samples from *experiment one* as background set. *Experiment four* considers a gallery with one controlled still picture for each subject a probe set containing multiple uncontrolled images. *Experiments one* and *two* only contain controlled images. *Experiments three*, *five*, and *six* are not utilized since they do not correspond to 2D face recognition. Figure 5.1 portrays samples of face images in FRGCv1.

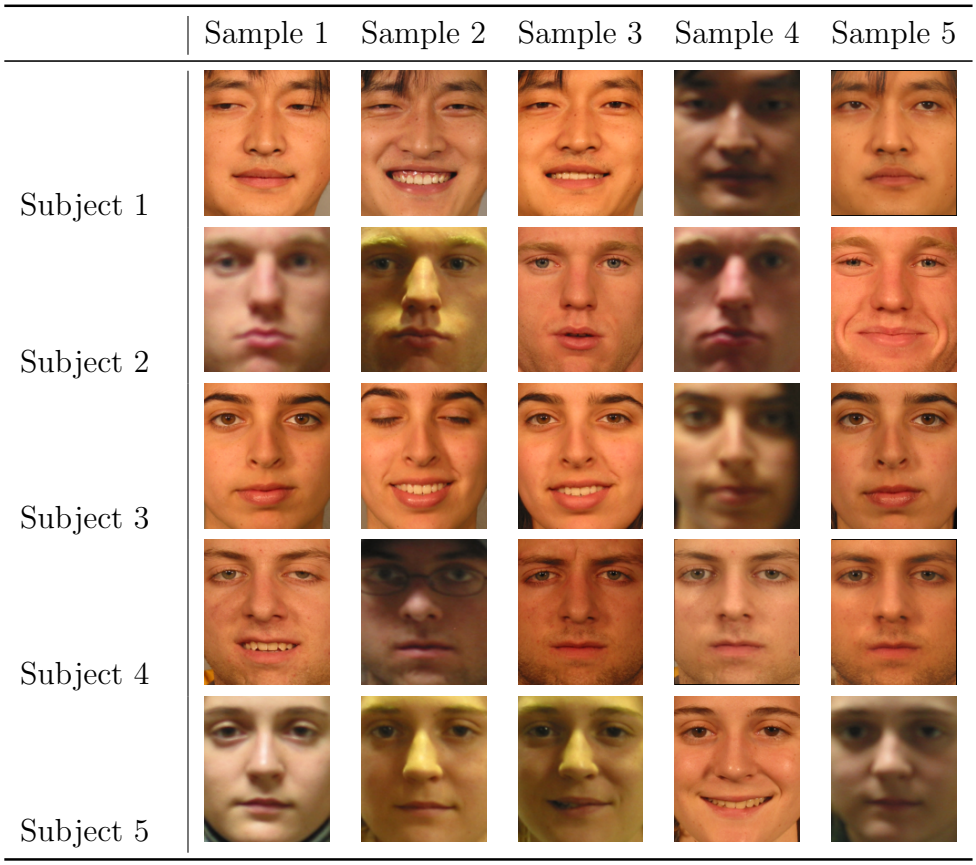


Figure 5.1: Examples of faces of five subjects from the FRGCv1 Dataset.

Public Figures (Pubfig) [Kumar et al., 2009]: This is a large and real-world face dataset comprising 58,797 images of 200 subjects collected from the Internet. Currently, only 26,787 images remain available due to copyright issues. Images are taken in totally uncontrolled situations with non-cooperative subjects with different

age, gender and ethnicity, resulting in variations in pose, lighting, expression, scene, camera and imaging conditions. The Pubfig dataset is partitioned into two units, the development set containing 16,336 images of 60 subjects and the evaluation set with 42,461 images of the remaining 140 subjects. There is no overlap between the two sets. Figure 5.2 shows examples of subjects in Pubfig.

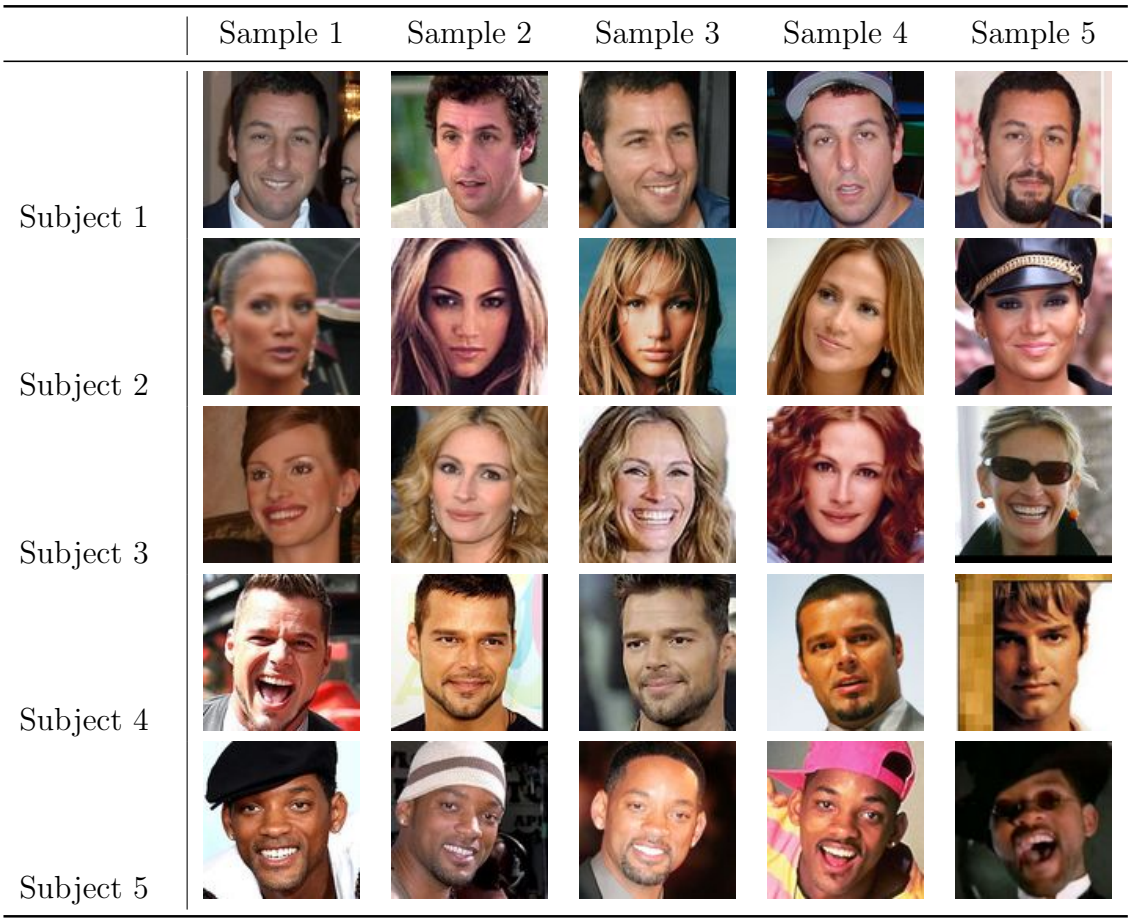


Figure 5.2: Examples of faces of five subjects from the Pubfig83 and consequently the Pubfig datasets.

Public Figures 83 (PubFig83) [Pinto et al., 2011]: This dataset is a subset of the original Public Figures dataset. It contains images of 83 subjects with at least 100 samples. With a total of 8,300 cropped facial images, it presents variations in pose and expression due to the natural uncontrolled scenarios of images collected from the Internet. Figure 5.2 depicts samples of face images in Pubfig83.

FG-Net Aging [Panis et al., 2016]: According to Panis et al. [2016], FG-Net Aging Database is part of the Project FG-NET (Face and Gesture Recognition Network), a project funded by the 5th Framework Programme, Information Society Technologies

(IST). The goal was to encourage research and development of techniques in the area of face and gestures. FG-Net induces studies mainly in facial appearance changes associated to aging. Consisting of 1,002 images of 82 subjects, the dataset comprises multiple races, ages ranging from 0 to 69 years and a large variation of illumination, pose and expression. Although its original website is no longer available, alternative ones make possible the download and execution of experiments with this challenging dataset. Figure 5.3 exemplifies face images of subjects in FG-Net.

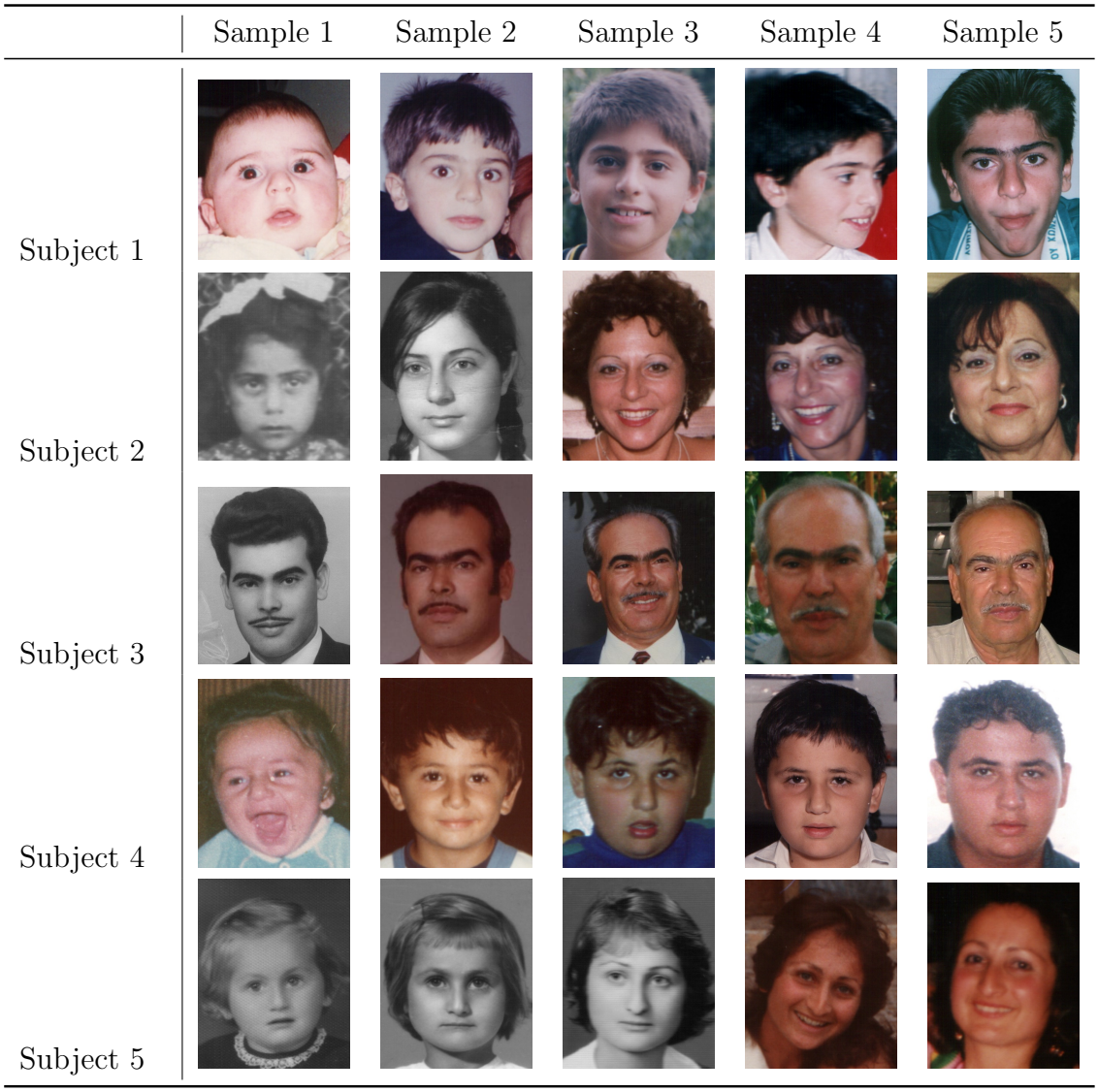


Figure 5.3: Examples of faces of five subjects from the FG-Net Aging Dataset.

5.1.2 Feature Descriptor

Feature extraction is useful to represent face images rather than using the raw pixel values which would result in variance in many aspects in images of the same subject.

For this reason, discriminant features are extracted and face images are represented by a feature vector.

In this work, we examine the results provided by a well-known descriptor, the VGGFace Convolutional Neural Network (CNN) Descriptor. The VGGFace CNN descriptor is based on the VGG-Very-Deep-16 CNN architecture [Simonyan and Zisserman, 2014] which is implemented by the computation described by Parkhi et al. [2015] using the VGGFace dataset. In this network, the first layers consist in a stack of 3×3 convolutional filters followed by three fully-connected and a soft-max layer. The rectification non-linearity (ReLU) is found in all hidden layers.

We consider the standard training weights used by the authors. Thus, the network is not fine tuned towards the selected datasets mentioned in Section 5.1.1. We choose this descriptor since it can extract robust and discriminative features from uncontrolled images due to the unconstrained dataset used in its training. Its high dimensionality ($4,096$ features) is not a problem since PLS projects features onto a lower-dimensionality feature space, keeping only relevant features. Figure 5.4 shows samples of the VGGFace dataset, used by Parkhi et al. [2015] to learn the network weights.

Figure 5.4: Examples of face images of VGGFace dataset used in the training of the VGGFace Convolutional Neural Network, model used to extract features in this work. Figure extracted from the presentation[1] of [Parkhi et al., 2015].

[1] http://www.robots.ox.ac.uk/~vgg/software/vgg_face/

5.1.3 Evaluation Metrics

When it comes to metrics for open-set face recognition, there is not a worldwide consensus. Thus, in this work, we choose two metrics to evaluate different aspects in recognition: the Receiver Operating Characteristic (ROC) and the Open-set Receiver Operating Characteristic (Open-set ROC). The former captures information on how accurate an approach is in terms only of detecting a gallery probe sample, i.e., determining whether a probe sample belongs to the gallery. The latter evaluates both the detection of a gallery sample and its correct identification.

Receiver Operating Characteristic (ROC): It is an extensively employed graphical plot which depicts the distinguishing competence of a binary classifier system as its discrimination threshold is varied. The ROC curve is the plotting of true positive rate (TPR) as a function of the false positive rate (FPR). TPR is also named sensitivity, recall or detection probability. On the other hand, FPR is also known as the fall-out or probability of false alarm and can be computed as 1−specificity. Therefore, this curve is the sensitivity as a function of fall-out.

The ROC popularity may be attributed mainly to its ability to provide a cost/benefit analysis in the diagnostic decision. In the comparison of approaches using ROCs plotting, the largest Area Under Curve (AUC), which ranges from 0.5 to 1, indicates the most worthwhile. The optimal point in a curve is the nearest to the top left corner.

The Receiver Operating Characteristic curve is employed in this work with the purpose of making a comparison between the proposed approach and literature approaches considering only the gallery detection stage. Figure 5.5 illustrates an example of a ROC curve.

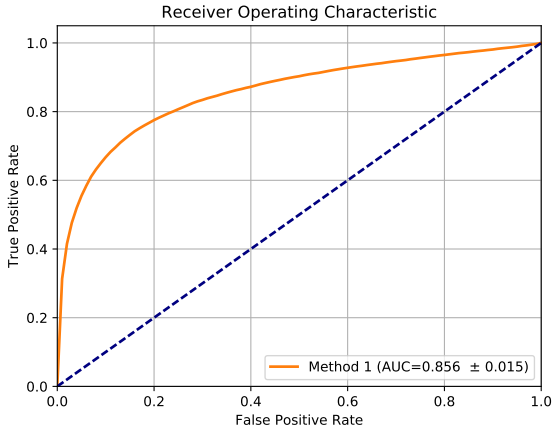


Figure 5.5: Example of a Receiver Operating Characteristic plot.

Open-set Receiver Operating Characteristic: In the evaluation of open-set identification systems, it is necessary to consider that only samples with score higher than a threshold are identified. In other words, only if a gallery probe sample is detected (i.e., the probe sample is predicted as belonging to the gallery), the identification is performed. One can lower this threshold to increase the candidate list size, so more samples would be processed. However, it also would increase the number of false alarms (number of probe samples from unknown subjects which is accepted as known). Therefore, an appropriate mode of evaluating approaches may combine detection and identification rates. In [Marcel, 2013], the combination of two metrics is used with the purpose to evaluate open-set identification systems:

- Detection and Identification Rate (DIR): estimation of the likelihood associated to a known/gallery subject being detected.

- False Alarm Rate (FAR): estimation of the likelihood of an unknown subject is incorrectly predicted as known.

Plotting the DIR as a function of FAR curve generates a chart also known as open-set ROC. The watch list task consists in first determining whether a probe sample corresponds to a subject in gallery. If so, identification is performed. The former task can be considered detection whereas the latter identification. DIR can be seen as the sensitivity to the system in correctly identifying the subject. DIR and FAR can be computed by the following equations:

$$DIR(\Delta) = \frac{|\text{probes with score} > \Delta|}{|\text{gallery probes correctly detected}|}, \tag{5.1}$$

$$FAR(\Delta) = \frac{|\text{non-gallery probes incorrectly detected with score} > \Delta|}{|\text{non-gallery probes}|}, \tag{5.2}$$

where Δ is the current threshold, which ranges in order to build a curve. In this work, we consider the use of Open-set ROC as one of the manners of evaluating the proposed approach since it provides a more realistic scenario in which one does not only want to detect a gallery sample but also determining its identity. Figure 5.6 illustrates an example of an Open-set ROC curve.

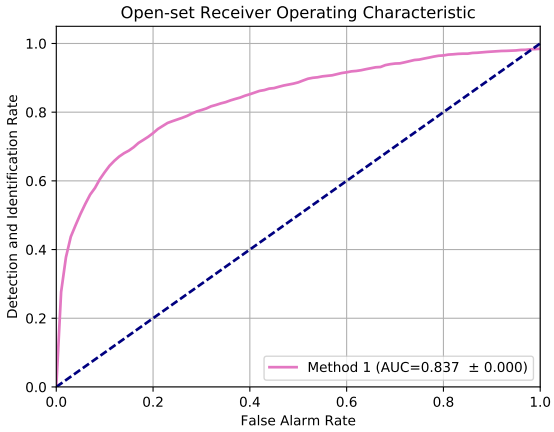


Figure 5.6: Example of an Open-set Receiver Operating Characteristic plot.

5.2 Evaluation Protocol

This section presents the evaluation protocol of the experiments reported in Section 5.3. We detail the baseline approaches, the use of a background set and the threshold selection.

The proposed approach will be compared with five different approaches: 1-class and Binary 1-vs-set M. [Scheirer et al., 2013], the 1-class and Binary SVM baselines, and PLSH [Vareto et al., 2017b]. The experiments with the 1-vs-set M. approaches were performed on a Intel Xeon E5-2687W CPU with 3.10GHz and 24GB of RAM using Ubuntu 16.04 LTS operating system, but no more than 4GB of RAM was required. The remaining experiments were performed on two Intel Xeon E5-2640 CPUs with 2.30GHz and 64GB of RAM using Windows 7 Professional system, but no more than 10GB of RAM was required.

For all experiments, except for the one varying the number of known subjects, we divide the whole dataset into two equal parts: the known and unknown subject sets. Then, we randomly choose 50% of known samples for training and the remainder for testing. In the Binary SVM experiments we also need unknown samples in training. For this experiment, specifically, we also divide the unknown samples into two sets, 50% of them to train and the remainder to test. All reported results consist in averages of evaluation results obtained in the performing repeated ten times.

The PLSH [Vareto et al., 2017b] experiments take the number of models to build as a parameter. In our experiments, we report the results using two different numbers: 500, which is the maximum value evaluated in their paper, and the number of known subjects, which is the same number of models of the proposed approach. We call them PLSH-500 and PLSH-Subj, respectively.

5.2.1 Baselines: Binary and 1-Class SVMs

The traditional support vector machine (SVM) [Schölkopf et al., 2001] is a popular classification technique aiming at defining the hyperplanes that better separates all data points of one class from those of other classes. The best hyperplane is the one with the largest margin between classes. Thus, when a probe data is projected onto the trained model, the distance between its data point and the hyperplanes determines the decision value.

The binary SVM version consists in a problem in which the number of classes is equal to two. Therefore, a single hyperplane is generated to separate classes. On the other hand, instead of defining a hyperplane, the 1-class SVM generates a spherical boundary around the data in the feature space, since in training stage only one class is considered. Adding most data into the hyper-sphere makes it becoming an optimization problem. For the open-set recognition problem, 1-class SVM seems to be a more natural solution since only known subject data are available in training stage.

Internally, SVM uses a kernel function which measures the similarity between two data points to determine the best hyper-plane/sphere. There are several kernel functions which can be applied for this purpose. In this work, we chose to use the linear kernel to make a fair comparison with other linear-SVM based approaches. Besides, we use the default values for all parameters and a single binary or 1-class SVM model is built considering all known subjects as positive set against unknown subjects as negative (in binary version) in the training.

5.2.2 Background Set

Whenever we train binary classifiers, we may want to add some extra data to the negative set, since a large amount of data helps to reinforce the differences between negative and positive sets making the method more robust. Taking this into account, we propose a version of our approach, named "Ours-bk", in which we add some extra data from subjects which have no overlap neither with the known set nor with the unknown set to the negative set, the background set. The background set contributes to increase the discriminability of a model if its images have acquisition characteristics similar to the dataset considered.

In addition to the proposed approach, the methods proposed by Scheirer et al. [2013], the 1-class and Binary 1-vs-set M., also employ the background set. However, they do not add the background set to the negative set. Instead, in each model, the background set data exclusively is used as negative set. The use of the negative set in the 1-class version may not be intuitive, but it helps to refine the final model.

In most evaluations, we choose the *experiment one* of FRGCv1 as the background set. In *experiment four*, we use samples from *experiment one* with removed overlapping samples as the background set. *Experiment two* uses the *experiment four* samples as background set since all samples in *experiment one* are contained in *experiment two*. Finally, Pubfig employs the Pubfig evaluation data (Pubfig-eval) as the "main" dataset and the background set is constituted by 10% of samples of 50% of subjects in the Pubfig development data (Pubfig-dev) since they carry similar acquisition characteristics. We sub-sample data since the dataset is large in terms of samples what would result in a long-time training, mainly in approaches based on SVM.

5.2.3 Threshold Selection

The threshold determines whether a probe face image belongs to a gallery subject or not. It is employed in the vote-list described in Section 4.3. We provide an evaluation of four different thresholds to find out the one that better impacts our algorithm. The threshold varies in:

$$\tau_0 = \mathcal{Z}_{S_1} - AVG(\mathcal{Z}_{S_2}, ..., \mathcal{Z}_{S_{p+1}}), p = \lceil 0.10 \times |\mathcal{Z}| \rceil, \tag{5.3}$$

$$\tau_1 = \frac{\mathcal{V}_{S_1}}{AVG(\mathcal{V}_{S_2}, \mathcal{V}_{S_3})}, \tag{5.4}$$

$$\tau_2 = \frac{\mathcal{V}_{S_1}}{\mathcal{V}_{S_2}}, \tag{5.5}$$

$$\tau_3 = \frac{\mathcal{V}_{S_1}}{AVG(\mathcal{V}_{S_2}, ..., \mathcal{V}_{S_r})}, r = \lceil 0.15 \times |\mathcal{V}| \rceil, \tag{5.6}$$

where $V = \{\mathcal{V}_{S_1}, ..., \mathcal{V}_{S_n}\}$ and $Z = \{\mathcal{Z}_{S_1}, ..., \mathcal{Z}_{S_n}\}$ are the original and normalized vote-lists ($\mathbf{v}$ and $\mathbf{z}$, respectively, described in Section 4.3) for a probe image, sorted in ascending order. The values of p and r define the proportion of Z and V considered, respectively. The function $AVG(X)$ computes the average of the values in X. The idea is to relate the top scorer s_1 with the succeeding subjects.

In this work, we propose the threshold τ_0, detailed in Equation 5.3, which is based on the difference between the normalized top scorer and the succeeding subjects. The idea is to capture the relation between the subject in vote-list with the highest value and the average of the p succeeding.p is set to 10% of the total of known subjects since it presents the best results. If the probe image belongs to a known subject, the difference between them would be high, since probe has similar characteristics with

only one subject. Otherwise, the difference would be low since probe presents similar characteristics to various subjects. The remaining equations are based on the ratio between the top scorer and the succeeding subjects, and are proposed by Vareto et al. [2017b]. According to their experiments, the best threshold is τ_3.

If the value of τ_0, τ_1, τ_2 or τ_3 attains a minimum value, probe is predicted as known, and we can determine its identity using OAA PLS. Otherwise, probe is rejected as unknown. Figure 5.7 displays the ROC curve using the proposed approach on FG-Net dataset for each threshold aforementioned. As depicted, the threshold described in Equation 5.3 provides better performance than the remainder, confirming our assumption that the difference between the top scorer value and the averaged remainder in vote-list better separates known subjects from unknown when compared to approaches based on the ratio. Therefore, we choose the threshold τ_0.

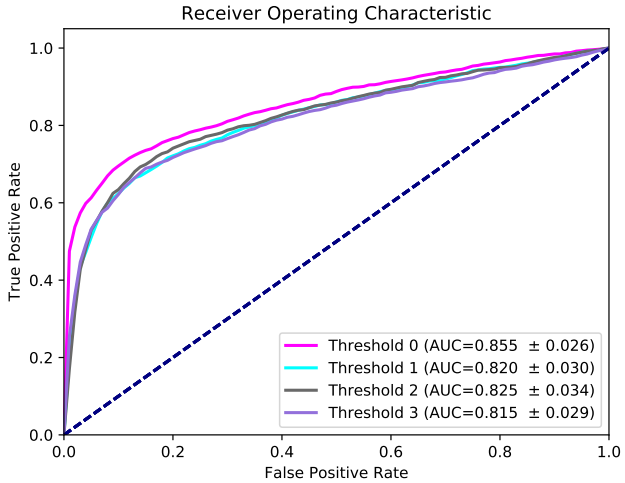


Figure 5.7: Average ROC curves for the proposed approach performed on the FG-NET dataset with different thresholds. We repeated this experiment ten times, fixing variable p to 10% and variable r to 15% of all known subjects.

5.3 Detection and Identification Evaluation

In this section we present the experimental results reported with ROC and Open-set ROC curves (described in Section 5.1.3) in different datasets. While the former measures the performance of approaches in the detection of gallery probe face images, the latter measures the performance of approaches in terms of the complete procedure of identification, that is, considering not only the detection performance but also the identification of probe samples predicted as known. The AUC (Area Under Curve)

attained by different approaches is provided to make a comparison regardless the dataset difficulty.

Tables 5.1 and 5.2 summarize the results obtained by the proposed approach and literature methods using AUC of ROC and Open-set ROC curves, respectively. In the complete procedure, presented in Table 5.2, we can note that the proposed approach outperforms all approaches in literature using all datasets except for the Binary-SVM, that presents similar result but utilizes unknown subject data in training stage, what we can not assume that is available.

Table 5.1: Summarization of AUC (± standard deviation) of ROC curve for different open-set recognition approaches and datasets.

Dataset	FRGC-SET2	FRGC-SET4	FG-NET	Pubfig83	Pubfig-Eval
1-Class-1-vs-set-M.	0.817 ± 0.044	0.610 ± 0.037	0.630 ± 0.029	0.856 ± 0.015	0.719 ± 0.014
1-Class-SVM	0.525 ± 0.049	0.511 ± 0.028	0.520 ± 0.064	0.550 ± 0.047	0.523 ± 0.026
Binary-1-vs-set-M.	0.970 ± 0.005	0.873 ± 0.022	0.499 ± 0.029	0.833 ± 0.025	0.899 ± 0.018
Binary-SVM	0.994 ± 0.000	**0.913**± 0.023	0.831 ± 0.029	0.973 ± 0.003	0.915 ± 0.005
PLSH-500	0.990 ± 0.003	0.867 ± 0.019	0.820 ± 0.018	0.966 ± 0.003	0.926 ± 0.006
Ours	0.993 ± 0.002	0.882 ± 0.022	**0.855** ± 0.026	**0.975** ± 0.003	0.942 ± 0.003
Ours-Bk	**0.994** ± 0.001	0.905 ± 0.019	0.851 ± 0.019	0.973 ± 0.003	**0.943** ± 0.004

Table 5.2: Summarization of AUC (± standard deviation) of Open-set ROC curve for different open-set recognition approaches and datasets.

Dataset	FRGC-SET2	FRGC-SET4	FGNET	Pubfig83	Pubfig-Eval
1-Class-1-vs-set-M.	0.814 ± 0.044	0.551 ± 0.038	0.540 ± 0.030	0.849 ± 0.014	0.704 ± 0.013
1-Class-SVM	0.524 ± 0.049	0.454 ± 0.031	0.423 ± 0.050	0.543 ± 0.045	0.510 ± 0.025
Binary-1-vs-set-M.	0.966 ± 0.007	0.781 ± 0.031	0.412 ± 0.031	0.824 ± 0.024	0.845 ± 0.017
Binary-SVM	**0.992** ± 0.003	0.777 ± 0.034	0.675 ± 0.031	0.957 ± 0.004	0.858 ± 0.006
PLSH-500	0.988 ± 0.004	0.784 ± 0.028	0.728 ± 0.026	0.957 ± 0.005	0.886 ± 0.008
Ours	**0.992** ± 0.003	0.787 ± 0.037	**0.761** ± 0.032	**0.965** ± 0.004	**0.899** ± 0.005
Ours-Bk	0.991 ± 0.005	**0.788** ± 0.026	0.739 ± 0.027	0.962 ± 0.005	0.895 ± 0.005

5.3.1 Results on the FRGC Dataset

Figure 5.8 shows the ROC and Open-set ROC curves obtained using the experiment *two* of FRGCv1. *Experiment two* comprises images acquired in a controlled scenario becoming the simplest dataset among all used in this work. Thus, in the experiments, all tested methods demonstrate high performance, except by 1-class SVM, which performs poorly in almost every experiment since it creates a single model with all the gallery training data. We achieve the highest AUC using the background set embedded in the negative set. Both 1-class and binary 1-vs-set-M. [Scheirer et al., 2013] methods

also use the background set (as explained in Section 5.2.2). The former makes use of *experiment one* in the refining step to adjust the boundaries and the latter employs it in the negative set. Considering the open-set ROC, the proposed approach outperforms the remainder by a slightly difference. Since this is an easy experiment, except for the 1-class SVM, all methods present high AUC using both metrics.

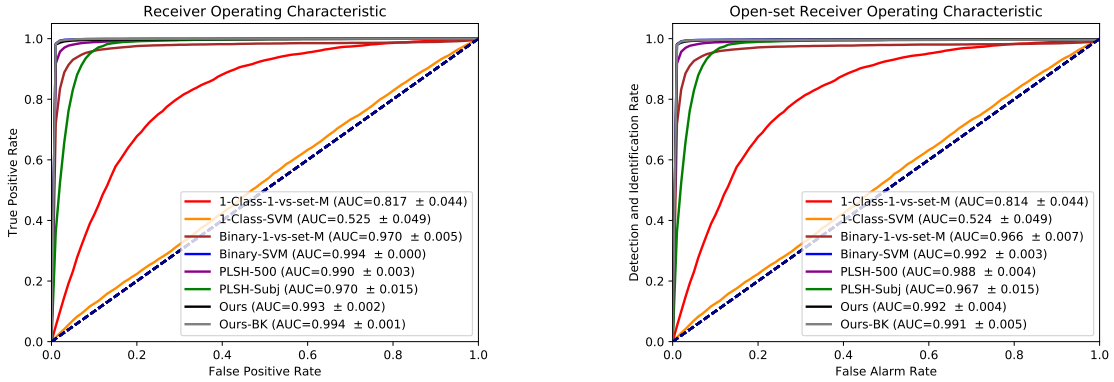


Figure 5.8: Average ROC and Open-set ROC curves for experiments conducted on *experiment two* of FRGCv1 dataset repeated ten times to compare six literature approaches and the two versions of the proposed method.

We report in Figure 5.9 the ROC and Open-set ROC curves obtained using the experiment *four* of FRGCv1. *Experiment four* considers uncontrolled images, representing therefore a more challenging scenario. In the experiments with this dataset, the background set also contributes for a better discrimination between the known subjects in positive set and the remainder. The best AUC in ROC is achieved by the Binary-SVM method with a slightly difference of 0.008 from the proposed approach using the background set. The binary-SVM presents the drawbacks mentioned before, mainly associated to the need of using unknown data.

When compared to the PLSH [Vareto et al., 2017b] with the maximum number of models considered in their paper (PLSH-500), our approach with or without background set is more effective. In the open-set ROC plot, we attain the best AUC, even outperforming the results with Binary-1-vs-set-M., which also benefits from the background set embedded in the negative set. Finally, both PLSH versions, PLSH-Subj and PLSH-500, attain a lower AUC. The fact that PLSH-500 contains 6 times more models than PLSH-Subj do not result in significant improvement.

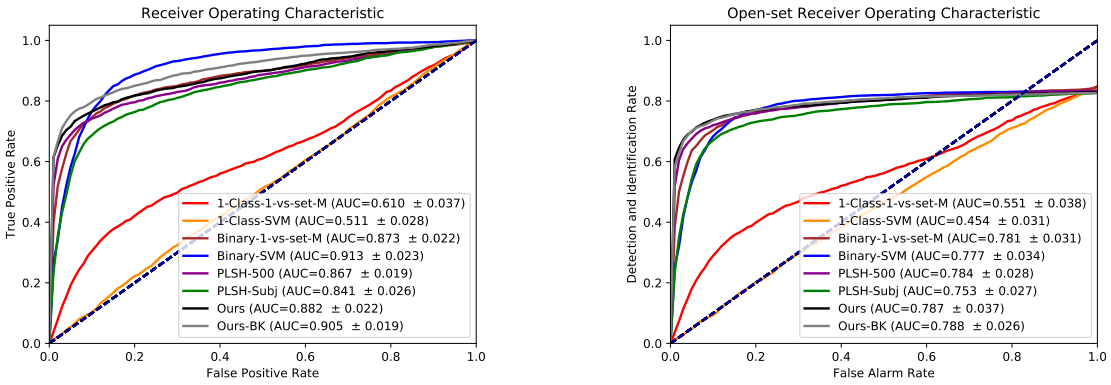


Figure 5.9: Average ROC and Open-set ROC curves for experiments conducted on *experiment four* of FRGCv1 dataset repeated ten times to compare six literature approaches and the two versions of the proposed approach.

5.3.2 Results on the FG-Net Dataset

In Figure 5.10, we report the ROC and Open-set ROC curves obtained using the FG-Net dataset. Since it comprises face images from subjects at different ages and in different acquisition scenarios, the difficulty in detecting known probe is high. In this experiment, we show that the proposed approach is robust to these changes, while other approaches fail, such as Binary-1-vs-set-M. We can notice that in this case, the background set does not help the recognition since it does not present similar characteristics with FG-Net. However, the AUC keeps the same in the experiments with or without a background set using our approach. This experiment also highlights the fact that the Binary-1-vs-set-M. shows great dependence on a similar background set. Besides, the results confirm that SVM-based approaches also require a large amount of data to achieve accurate results.

When comparing with PLSH, we outperformed both versions, with 500 models and with 41 models, demonstrating the effectiveness of the proposed approach while maintaining a low number of models trained. According to the Open-set ROC plot, the best AUC is achieved by the proposed approach with no use of background set, since it has no characteristics in common with FG-Net as mentioned before. The approaches based on SVM fail since they require a large amount of data and a few samples are available in this dataset. Comparing with the PLSH method, we achieved higher AUC while using fewer PLS models.

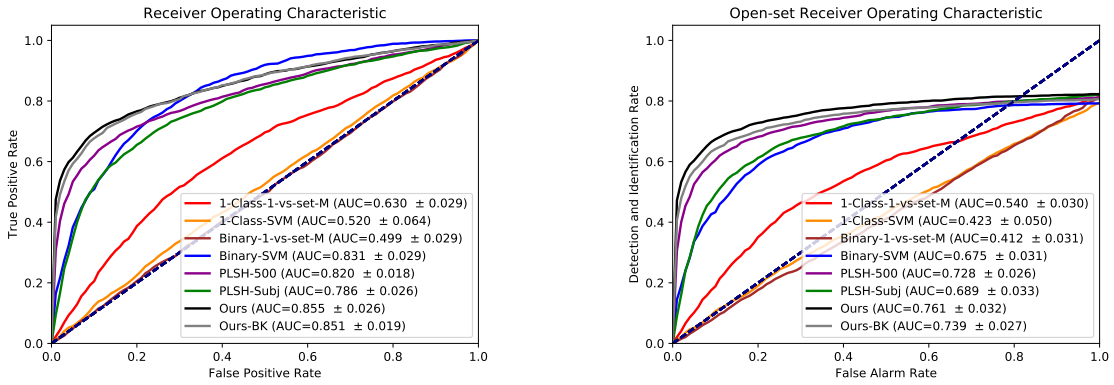


Figure 5.10: Average ROC and Open-set ROC curves for experiments conducted on FG-Net dataset repeated ten times to compare six literature approaches and the two versions of the proposed approach.

5.3.3 Results on the Pubfig83 Dataset

Figure 5.11 shows the ROC and Open-set ROC curves obtained using the Pubfig83 dataset, which consists of subjects in the original Pubfig with more than 100 images collected from Internet, making it an "in-the-wild" dataset. Different from most of datasets considered, Pubfig83 contains several samples of each subject which helps the learning of more discriminative models. Despite our approach attains the best performance, the remaining SVM-based approaches also achieve high AUC in the ROC curve, since SVM usually needs a large number of samples to build accurate models. We can note that background set does not influence the performance of the proposed approach since it does not have common characteristics with Pubfig83. When comparing with PLSH, we also outperform both versions, with 500 models and with 41 models.

The Open-set ROC plot also shows that the proposed approach without background set presents the best AUC since background does not have characteristics in common with Pubfig83. In this case, when an unknown probe is projected onto a model, it would be similar to any data (including the known data) but the background set, due to the extreme dissimilarity in acquisition and, as a consequence, in the feature vectors of this set. The presence of background set would not affect the recognition rate.

In comparison to our approach, the binary-SVM and PLSH-500 present a near AUC. However, the binary-SVM uses unknown data in the training phase, that is, subjects from unknown subject are also used to train. In realistic scenarios, the assumption of knowledge of non-gallery subjects can not be made. On the contrary, our

method reaches high performance with no unknown training data. With regard to the PLSH method, it learns roughly 13 times more models than ours since our approach learns 41 models and PLS-500 learns 500 models, taking a higher computational cost.

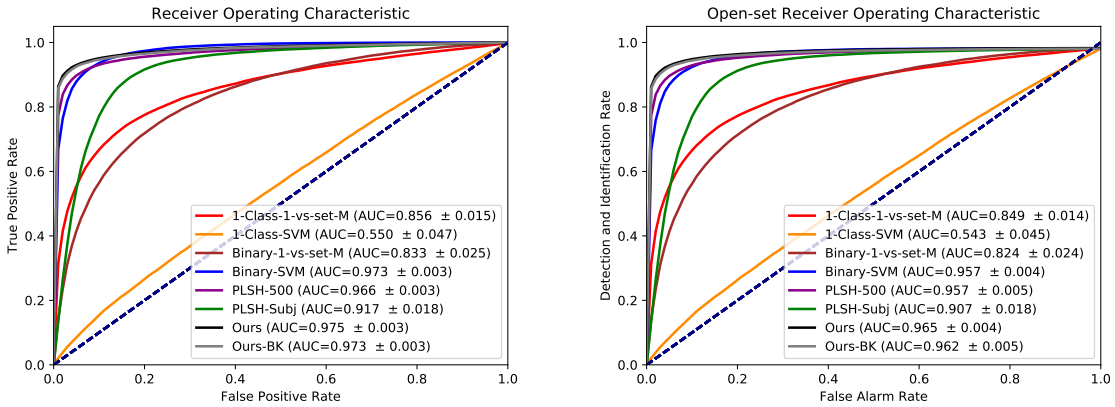


Figure 5.11: Average ROC and Open-set ROC curves for experiments conducted on Pubfig83 dataset repeated ten times to compare six literature approaches and the two versions of the proposed approach.

5.3.4 Results on the Pubfig Dataset

In Figure 5.12, we report the ROC and Open-set ROC curves obtained using the Pubfig dataset. Since Pubfig83 is part of Pubfig, the latter carries the same characteristics as the former. This experiment aims at demonstrating the proposed work ability of handling a large amount of data. Although Pubfig does not comprise a high number of subjects, it contains a large number of samples, mainly, when compared to the previously mentioned datasets. The curve plots demonstrate that even in large amount of data scenarios, when SVM-based methods present the most satisfactory effectiveness, our approach achieves better performance. When contrasted with PLSH, our approach, using the background set, attains the best AUC values, either in ROC or in open-set ROC plots.

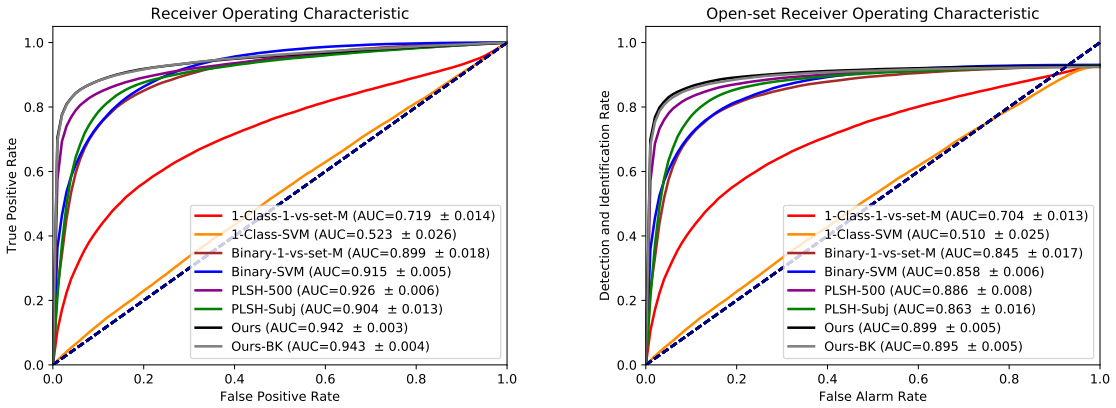


Figure 5.12: Average ROC and Open-set ROC curves for experiments conducted on Pubfig dataset repeated ten times comparing six literature approaches and the two versions of the proposed approach.

5.4 Evaluation with the Same Number of Models in PLSH Approach

Considering that our work is an extension of the approach proposed by Vareto et al. [2017b], this section presents a direct comparison between our work and PLSH. In order to make a fair comparison, we reduce the number of models in PLSH to the same number of models we build. Since the proposed approach learns an one-against-all model per subject, the number of models we use is equal to the number of gallery subjects.

Tables 5.3 and 5.4 report the results contrasting our method with PLSH-Subj using ROC and Open-set ROC curves, respectively. The number of models for "FRGC-SET2" and "FRGC-SET4" is equal to 76. For "Pubfig-eval", this number is equal to 70. In the experiments with "FG-Net" and "Pubfig83", we learn 41 models. These numbers represent half of the total subjects in the dataset. As we can notice, the proposed approach outperforms PLSH regardless the metric or dataset difficulty. This fact can be explained since modeling a single subject against the remainder is simpler since samples in the positive set present similar features. On the other hand, PLSH trains models considering half of the gallery subject as positive and the remainder as negative, that is, the positive set contains images with different features since they come from different subjects. In this case, building a single model to discriminate the positive set may be a harder task.

Table 5.3: Comparison between AUC (± standard deviation) of ROC of the proposed approach and PLSH containing the same number of models we use.

Dataset	FRGC-SET2	FRGC-SET4	FG-NET	Pubfig83	Pubfig-Eval
PLSH-500	0.990 ± 0.003	0.867 ± 0.019	0.820 ± 0.018	0.966 ± 0.003	0.926 ± 0.006
PLSH-Subj	0.970 ± 0.015	0.841 ± 0.026	0.786 ± 0.026	0.917 ± 0.018	0.904 ± 0.013
Ours	0.993 ± 0.002	0.882 ± 0.022	**0.855** ± 0.026	**0.975** ± 0.003	0.942 ± 0.003
Ours-Bk	**0.994** ± 0.001	**0.905** ± 0.019	0.851 ± 0.019	0.973 ± 0.003	**0.943** ± 0.004

Table 5.4: Comparison between AUC (± standard deviation)of Open-set ROC of the proposed approach and PLSH containing the same number of models we use.

Dataset	FRGC-SET2	FRGC-SET4	FGNET	Pubfig83	Pubfig-Eval
PLSH-500	0.988 ± 0.004	0.784 ± 0.028	0.728 ± 0.026	0.957 ± 0.005	0.886 ± 0.008
PLSH-Subj	0.967 ± 0.015	0.753 ± 0.027	0.689 ± 0.033	0.907 ± 0.018	0.863 ± 0.016
Ours	**0.992** ± 0.003	0.787 ± 0.037	**0.761** ± 0.032	**0.965** ± 0.004	**0.899** ± 0.005
Ours-Bk	0.991 ± 0.005	**0.788** ± 0.026	0.739 ± 0.027	0.962 ± 0.005	0.895 ± 0.005

5.5 Comparison in a Lower FPR and FAR

As described in Section 5.1.3, in both curves ROC and Open-set ROC, we desire an approach that reaches a high value in y axis keeping a low value in x axis. In ROC curves it means that with a low False Positive Rate (FPR) we intend to obtain a high True Positive Rate (TPR). Concerning the open-set ROC metric, with a low False Alarm Rate (FAR) we aim at achieving a high Detection and Identification Rate (DIR). Thus, in order to evaluate how approaches perform in a low FPR and FAR, we report the ROC and open-set ROC in Figure 5.13 using the Pubfig83 dataset. This dataset is chosen since it presents the best results for all approaches when we compare datasets with many samples per subject. Thus, it would be interesting to evaluate profoundly the behavior of approaches in this dataset and the separation between curves when we delimitate the interval of x axes. We range the values in x axes from 0.01 to 0.10. The plots of ROC and open-set ROC allows for a better analysis not only related to the value of TPR and DIR with low FPR and FAR but also related to a more evident separation between curves.

Our approach outperforms the literature approaches with no use of background set since the background set for this dataset is the *experiment one*. Considering the ROC curve, the best AUC values are achieved by the proposed approach in both versions, the Binary-SVM and PLSH-500. However, our curve keeps above the remainder throughout the variation of FPR, indicating a better behavior. In respect to the open-set plot, the same behavior remains. An interesting point to observe is the

significant difference in AUC presented by PLSH-Subj and PLSH-500 in both plots. These results indicate that PLSH fails with a lower FAR and FPR if the number of models built is the same as ours, that is, it is equal to the number of known subjects. In comparison with binary-SVM, which sees samples from unknown subjects in training, we can notice that with the lowest value in FPR and FAR, our approach presents a larger value in TPR and DIR using only samples from known subjects in the training phase.

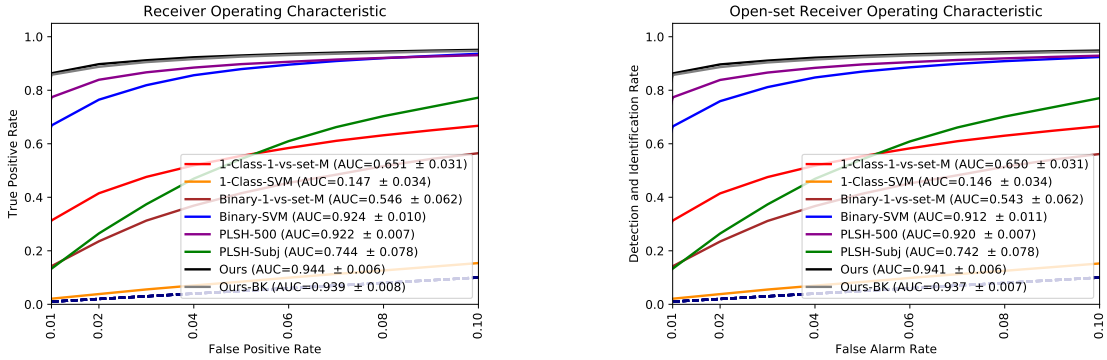


Figure 5.13: Average ROC and Open-set ROC curves for experiments conducted on Pubfig83 repeated ten times with FPR and FAR ranging from 0.01 to 0.10.

5.6 Evaluation on the Number of Known Subjects

In this section, we present an evaluation on how approaches behave when the number of subjects enrolled in gallery is equal to 10%, 50% and 90% of the total subjects in the dataset. In other words, we want to analyze the approaches robustness to variation in the known set size. Figure 5.14 depicts the AUC of ROC curve when the known set size is 15, 76 and 136 for the *experiment four* of FRGCv1 dataset. The best behavior is a constant value in AUC despite the variation on the known set size. The proposed approach maintains the AUC almost constant when the known set size varies, becoming the nearest curve to the Binary SVM, approach which sees unknown data in training. The remaining approaches present the same behavior, except for the 1-class SVM which builds a single model for a large number of subjects and the 1-class 1-vs-set M., which builds a single model per class but achieves the maximum response of one of this large number of models. While the former is harmed by the necessity of building a single model which separates many different subjects from the remainder, the latter

presents a drop in performance since if we have more models, the chance of getting the maximum response from the correct model is reduced.

It is important to notice that a variation in the number of known data also affects the number of unknown data. For instance, when we use 90% of subjects as known data, 10% of them are used as unknown data. Since our approach, PLSH, and the approaches based on 1-vs-set M. only train with known data, the AUC values on known size equal to 50% tends to present better results.

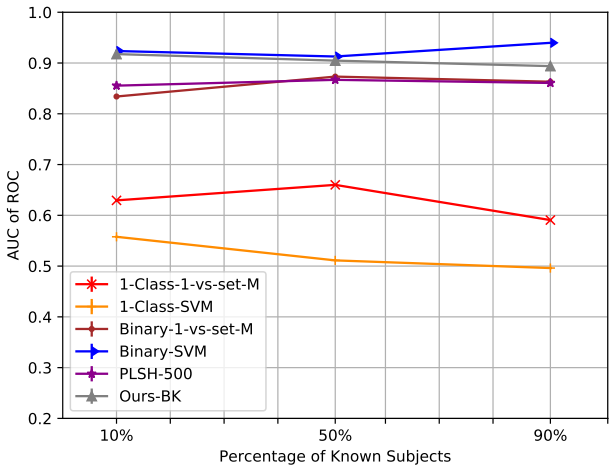


Figure 5.14: Average AUC of ROC using 10%, 50% and 90% of total subjects as known set size with the *experiment four* of FRGCv1 dataset and repeating ten times.

Chapter 6

Conclusions and Future Works

In this work, we proposed and evaluated an approach to perform open-set face identification using a combination of Partial Least Squares models trained in one-against-all protocol. Responses from the probe projection onto the models are aggregated in a vote-list and used to discriminate between known and unknown subjects. PLS presents many advantages when compared to other techniques, such as SVM. It handles feature vectors with high dimension, applications with few and unbalanced samples per class, and presents robust results while keeping a low computational cost.

Experiments were carried out in a diversity of datasets attaining satisfactory results regardless the challenges they present. Improvements in performance were achieved by embedding a background set, data presenting similar aspects to the main dataset, in the negative set of each PLS model. Different from some traditional approaches, we do not make assumption that unknown training data is available. Instead, we provide a robust method using only training samples from known subject that outperforms literature approaches in most of the experiments.

Considering the AUC of ROC curve as the metric, the proposed approach presented the best results when compared to literature approaches in experiments using all considered datasets, except by the experiment four of FRGC in which the binary SVM obtains a better result. However, different from the proposed approach, binary SVM sees unknown data in training stage. In the experiments using the AUC of Open-set ROC curve as metric, that is, evaluating the complete pipeline of watch list, our approach reaches the best results using all datasets, but the experiment two of FRGC dataset. In this experiment, our approach presents the same result as binary SVM.

The adoption of a large number of models, as used in the PLSH algorithm, is

unnecessary since training only one model per subject in an one-against-all protocol can provide more discriminative models impacting the results in the detection of gallery probes, while keeping a lower number of models. Learning a model with features from only one subject in the positive set allows for more consistent classifiers since the feature vectors present similar behavior. On the other hand, training models with half of the training subjects in the positive set, as considered in PLSH, leads to feature vectors with different characteristics in the positive and negative set, hampering the building of accurate models. We also contribute with a novel manner to compute the existence of a highlighted position in vote-list based on the difference between normalized values which attains a more accurate performance when compared to other literature techniques. Different from PLSH, we propose a complete pipeline to perform open-set face identification considering not only the stage of detecting a gallery probe but also determining its identity.

Although the majority of methods to perform face identification do not take into account the possibility of an unknown probe in the test phase, the development of techniques in this regard is an important task in Computer Vision since more applications would be encompassed. In this work, we show that the watch list task can be solved with a simple but effective approach. As a negative point, we presume that huge galleries would affect significantly the computational cost of the proposed approach since the test consists in projecting the probe onto all models. However, in real-world applications, usually, the number of known subjects is significantly smaller than the number of unknown subjects.

In future works, we intend to consider an evaluation of the proposed method on datasets with a huge amount of subjects in order to evaluate the impact in the computational cost. Besides, the method used to build each model, PLS, can be replaced by other binary classifiers such as neural-network-based techniques and other classifiers with accurate performance. We could also study more the influence of adding the background set to the negative set of a model. Lastly, changes in the partitioning of positive and negative sets in a model could be made in order to not use the whole gallery.

Bibliography

Barkan, O., Weill, J., Wolf, L., and Aronowitz, H. (2013). Fast high dimensional vector multiplication face recognition. In *IEEE International Conference on Computer Vision*, pages 1960--1967.

Bendale, A. and Boult, T. (2015). Towards open world recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1893--1902.

Bendale, A. and Boult, T. E. (2016). Towards open set deep networks. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1563--1572.

Best-Rowden, L., Han, H., Otto, C., Klare, B. F., and Jain, A. K. (2014). Unconstrained face recognition: Identifying a person of interest from a media collection. *IEEE Transactions on Information Forensics and Security*, 9(12):2144--2157.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.

Chellappa, R., Sinha, P., and Phillips, P. J. (2010). Face recognition by computers and humans. *IEEE Computer*, 43(2):46--55.

de O. Costa, F., Silva, E., Eckmann, M., Scheirer, W. J., and Rocha, A. (2014). Open set source camera attribution and device linking. *Pattern Recognition Letters*, 39:92--101.

de Paulo Carlos, G., Pedrini, H., and Schwartz, W. R. (2013). Fast and scalable enrollment for face identification based on partial least squares. In *IEEE International Conference and Workshops on Automatic Face and Gesture Recognition*, pages 1--8.

dos Santos, C., Kijak, E., Gravier, G., and Schwartz, W. (2015). Learning to hash faces using large feature vectors. In *International Workshop on Content-Based Multimedia Indexing*, pages 1--6.

dos Santos, C., Kijak, E., Gravier, G., and Schwartz, W. (2016). Partial least squares for face hashing. *Neurocomputing*, 213:34–47.

dos Santos Junior, C. E. (2015). Partial least squares for face hashing. Master's thesis, Federal University of Minas Gerais.

dos Santos Junior, C. E. and Schwartz, W. R. (2014). Extending face identification to open-set face recognition. In *Conference on Graphics, Patterns and Images*, pages 188--195.

Dou, P., Shah, S. K., and Kakadiaris, I. A. (2017). End-to-end 3d face reconstruction with deep neural networks. *arXiv preprint arXiv:1704.05020*.

Ekenel, H. K. and Stiefelhagen, R. (2005). Local appearance based face recognition using discrete cosine transform. In *European Signal Processing Conference*, pages 1--5.

Ekenel, H. K. and Stiefelhagen, R. (2006). Analysis of local appearance-based face recognition: Effects of feature selection and feature normalization. In *IEEE Conference on Computer Vision and Pattern Recognition Workshop*, pages 34--34.

Ekenel, H. K., Szasz-Toth, L., and Stiefelhagen, R. (2009). Open-set face recognition-based visitor interface system. In *International Conference on Computer Vision Systems*, pages 43–52.

Guo, H., Schwartz, W. R., and Davis, L. S. (2011). Face verification using large feature sets and one shot similarity. In *International Joint Conference on Biometrics*, pages 1--8.

Hassner, T., Harel, S., Paz, E., and Enbar, R. (2015). Effective face frontalization in unconstrained images. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 4295--4304.

Hayat, M., Khan, S. H., Werghi, N., and Goecke, R. (2017). Joint registration and representation learning for unconstrained face identification. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages --.

He, K., Zhang, X., Ren, S., and Sun, J. (2014). Spatial pyramid pooling in deep convolutional networks for visual recognition. In *European Conference on Computer Vision*, pages 346--361.

Heitmeyer, R. (2000). Biometric identification promises fast and secure processing of airline passengers. *International Civil Aviation Organization journal*, 55(9):10--11.

Jones, M. and Viola, P. (2003). Fast multi-view face detection. In *IEEE Computer Vision and Pattern Recognition*.

Kamgar-Parsi, B., Lawson, W., and Kamgar-Parsi, B. (2011). Toward development of a face recognition system for watchlist surveillance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(10):1925--1937.

Kanade, T. (1974). Picture processing system by computer complex and recognition of human faces. *Kyoto University*.

Kumar, N., Berg, A. C., Belhumeur, P. N., and Nayar, S. K. (2009). Attribute and simile classifiers for face verification. In *IEEE International Conference on Computer Vision*, pages 365--372.

Li, F. and Wechsler, H. (2005). Open-set face recognition using transduction. *IEEE Pattern Analysis and Machine Intelligence*, 27(11):1686--1697.

Liao, S., Lei, Z., Yi, D., and Li, S. Z. (2014). A benchmark study of large-scale unconstrained face recognition. In *IEEE International Joint Conference on Biometrics*, pages 1--8.

Liu, W., Wen, Y., Yu, Z., Li, M., Raj, B., and Song, L. (2017). Sphereface: Deep hypersphere embedding for face recognition. *arXiv preprint arXiv:1704.08063*.

Maltoni, D., Maio, D., Jain, A., and Prabhakar, S. (2009). *Handbook of Fingerprint Recognition*. Springer Science & Business Media.

Marcel, S. (2013). Beat–biometrics evaluation and testing. *Biometric Technology Today*, 2013(1):5--7.

Ng, J. Y.-H., Hausknecht, M., Vijayanarasimhan, S., Vinyals, O., Monga, R., and Toderici, G. (2015). Beyond short snippets: Deep networks for video classification. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 4694--4702.

Ortiz, E. G. and Becker, B. C. (2014). Face recognition for web-scale datasets. *Computer Vision and Image Understanding*, 118:153--170.

Panis, G., Lanitis, A., Tsapatsoulis, N., and Cootes, T. F. (2016). Overview of research on facial ageing using the fg-net ageing database. *IET Biometrics*, 5(2):37--46.

Parkhi, O. M., Vedaldi, A., Zisserman, A., et al. (2015). Deep face recognition. In *British Machine Vision Conference*, volume 1, page 6.

Phillips, P. J., Flynn, P. J., Scruggs, T., Bowyer, K. W., Chang, J., Hoffman, K., Marques, J., Min, J., and Worek, W. (2005). Overview of the face recognition grand challenge. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 947–954.

Phillips, P. J., Grother, P., and Micheals, R. (2011). *Evaluation Methods in Face Recognition*. Springer.

Pinto, N., Stone, Z., Zickler, T., and Cox, D. (2011). Scaling up biologically-inspired computer vision: A case study in unconstrained face recognition on facebook. In *IEEE Conference Computer Vision and Pattern Recognition Workshops*, pages 35--42.

Rosipal, R. and Krämer, N. (2006). Overview and recent advances in partial least squares. *Lecture Notes in Computer Science*, 3940:34.

Scheirer, W. J., de Rezende Rocha, A., Sapkota, A., and Boult, T. E. (2013). Toward open set recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(7):1757--1772.

Scheirer, W. J., Jain, L. P., and Boult, T. E. (2014). Probability models for open set recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(11):2317--2324.

Schölkopf, B., Burges, C. J., and Smola, A. J. (1999). *Advances in Kernel Methods: Support Vector Learning*. MIT press.

Schölkopf, B., Platt, J. C., Shawe-Taylor, J., Smola, A. J., and Williamson, R. C. (2001). Estimating the support of a high-dimensional distribution. *Neural Computation*, 13(7):1443--1471.

Schwartz, W. R., Guo, H., Choi, J., and Davis, L. S. (2012). Face identification using large feature sets. *IEEE Transactions on Image Processing*, 21(4):2245--2255.

Schwartz, W. R., Guo, H., and Davis, L. S. (2010). A robust and scalable approach to face identification. In *IEEE European Conference on Computer Vision*, pages 476--489.

Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Stallkamp, J. (2006). Video-based face recognition using local appearance-based models. Master's thesis, University of Karlsruhe.

Stallkamp, J., Ekenel, H. K., and Stiefelhagen, R. (2007). Video-based face recognition on real-world data. In *IEEE International Conference on Computer Vision*, pages 1--8.

Tran, L., Yin, X., and Liu, X. (2017). Disentangled representation learning gan for pose-invariant face recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 4, page 7.

Vareto, R., Silva, S., Costa, F., and Schwartz, W. R. (2017a). Face verification based on relational disparity features and partial least squares models. In *Conference on Graphics, Patterns and Images*, pages 209--215.

Vareto, R., Silva, S., Costa, F., and Schwartz, W. R. (2017b). Towards open-set face recognition using hashing functions. In *International Joint Conference on Biometrics*, pages 1--8.

Wang, W., Wang, R., Shan, S., and Chen, X. (2017). Discriminative covariance oriented representation learning for face recognition with image sets. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 5599--5608.

Wechsler, H. (2009). *Reliable Face Recognition Methods: System Design, Implementation and Evaluation*, volume 7. Springer Science & Business Media.

Wold, H. (1985). Partial least squares. *Encyclopedia of Statistical Sciences*, pages 581--591.

Wright, J., Yang, A. Y., Ganesh, A., Sastry, S. S., and Ma, Y. (2009). Robust face recognition via sparse representation. *IEEE transactions on pattern analysis and machine intelligence*, 31(2):210--227.

Zhang, H. and Patel, V. (2016). Sparse representation-based open set recognition. *IEEE Pattern Analysis and Machine Intelligence*, 39(8):1690--1696.

Zhu, X., Lei, Z., Yan, J., Yi, D., and Li, S. Z. (2015). High-fidelity pose and expression normalization for face recognition in the wild. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 787--796.