Face Identification in Large Galleries

Rafael H. Vareto, Filipe Costa, William Robson Schwartz
Smart Surveillance Interest Group, Department of Computer Science
Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
rafaelvareto@dcc.ufmg.br, filipe.oc87@gmail.com, william@dcc.ufmg.br

***Abstract*—Face Recognition is one of the most studied problems in computer vision due to its importance to areas such as forensics, surveillance, neuroscience and psychology. This work comprises a preliminary study of the face identification (one of the tasks comprising face recognition) considering large galleries. We analyze how the face recognition rate behaves under the presence of an increasing number of individuals in the gallery set, referred to as distractors. Nowadays, many real-world applications demand face search and identification at colossal scale. Therefore, we direct our attention towards the 1:N face identification task, where N, the gallery size, is extremely large. The idea behind this work is to study and evaluate the performance of face identification for such large face galleries.**

***Keywords*-Distractors; Partial Least Squares; Face Hashing; Scalability; Face Identification;**

I. Introduction

Face Recognition (FR) is one of the most prominent computer vision problems, important to areas such as forensics, surveillance, neuroscience and psychology. Surveillance systems count on quiet and passive acquisition by taking the face image deprived of cooperation or knowledge of the person being framed.

In general, the FR task is dealt like a machine learning problem. The learning stage consists of a collection of face images labeled with their corresponding subject's identity, denominated *gallery set*, and the *probe set*, a collection of unlabeled face images from the same array of people (closed set), which constitutes the testing stage. Given a *gallery* and a *probe set*, most approaches explore all gallery samples, extracting a series of descriptive features from each face in order to learn a model. Then, a machine learning technique, applied as a classifier, assigns each probe image's feature vector the label that closest matches a subject from the gallery.

There are several works in the literature that address recognition accuracy enhancement. In particular, Huang et al. [1] and Neves et al. [2] designed unconstrained datasets primarily for this purpose. On the other hand, just a few of them delve into scalable approaches for large gallery sets. Hence, scalability in FR has not yet been appropriately addressed. Due to their "small" size, most existing FR datasets fail approach two problems: both (i) identification rate decreases and (ii) multi-class classification complexity increases as the dataset size expands.

When it comes to large galleries, a natural alternative is to replace objects and shapes by their features and, then, apply some sort of indexing and search strategy in the new "low-dimensional" space. We believe that building a solid gallery filtering approach is an essential procedure for face identification at scale. When it is not employed, searching for the right identity in a dataset containing countless individuals will be like finding a needle in a haystack. If a filtering step is not implemented, a large load of tests will be required to retrieve an individual's true identity. Contrarily, a stable refinement step casts aside individuals enrolled in the gallery that are improbable to correspond to the probe sample identity with low computational penalty.

Many real-world applications demand face search and identification at colossal scale. Therefore, we direct our attention towards the $1 : N$ face identification task, when 1 represents a probe set, and$N$, the gallery size, is extremely large. As N increases, the computational cost rises significantly. We evaluate how a simple recognition algorithm performs with a massive number of distractors, that is, subjects only used to populate the gallery set. Some researchers focus on the open-set face identification problem, but in this study, similarly to the majority of works in the literature, our performance evaluation is carried out in the closed-set gallery fashion.

In this work, we analyze how a face recognition system approach reacts with the addition of distractors, which are subjects enrolled in the gallery set if and only if they are not present in the query set. In general, the overall performance deteriorates with the increase of the gallery size. Consequently, we investigate both search performance and search time for large datasets – up to 10 thousand additional images are used in this preliminary study.

The remainder of this document is structured as follows. Section II encapsulates scalable face identification algorithms as well as a few large datasets. Section III presents a simple algorithm for feature extraction and the description of Partial Least Squares. Section IV exhibits the results of our initial experiments. Finally, in Section V, conclusions and future works are pointed out.

II. Related Works

There is a large number of works proposed to solve identification problems, either in unrestrained scenarios or in relatively "small" datasets [3]–[6]. It is needless to say that these studies accomplished substantial progress in the last 10 years; however, FR is far from being solved since many applications have failed to deliver in scenarios containing billions of individuals.

With scalability in mind, some datasets were proposed: Ng et al. [7] released a moderate-sized image database called

FaceScrub; Shlizerman et al. [8] announced the *MegaFace challenge*, and Wang et al. [9] proposed what seems to be the largest face search experiment up to now, conducted on a dataset with 80-million images downloaded from the Internet.

Some researchers have employed convolutional neural networks [10]–[13] whereas others focused either on clustering or hashing techniques [14]–[16] to get around the scalability problem. Grother et al. [17] analyzed leading algorithms on millions of images taking into account accuracy and speed metrics. The results were considerably satisfactory, attaining recognition result of 90% on a 1.6-million-subject dataset.

The approach proposed by Santos et al. [18] derives from Locality-Sensitive Hashing (LSH). Basically, they replaced LSH randomness by a PLS regression. The method projects a query image onto each learned PLS model in order to obtain a score value that is used either to increase or decrease weights of subjects in the positive set. Lastly, the list of subjects is rearranged and presented as a candidate list. They revealed at minimum 95% recognition rate on the Facial Recognition Technology (FERET) [19] dataset and up to 96.7% on the Face Recognition Grand Challenge (FRGC) [20] dataset.

Pham et al. [21] used a linked-node *m*-ary tree (LM-tree) for both Exact and Approximate Nearest Neighbor (ENN/ANN) searches. Their method narrows down the search space with pruning rules. They validated their methods on datasets composed of SURF [22], SIFT [23] and GIST [24] features to show it works fine for ENN and ANN searches in comparison to many state-of-the-art methods. Wang et al. [9] developed a face search system based on a cascade-like framework. In a nutshell, they normalize input images, generate templates and quickly select an approximate list of candidates. It employs a deep convolutional network combined with product quantization. The authors reported 98.23% accuracy rate on LFW standard protocol and 56.27% on BLUFR [25] using cosine similarity.

Up to now, the methodology employed in this work is essentially based on one of the methods proposed by Schwartz et al. [26]. The authors focus on reducing the time required to evaluate a massive volume of samples and identifying faces presenting a high dimensional feature descriptor using Partial Least Squares (PLS). Additionally, they combine low-level feature descriptors to find out those that best segregate different subjects. The method is evaluated on FRGC and FERET datasets.

III. Employed Approach

The evaluated FR approach comprises a simple feature extraction algorithm preceding a Partial Least Squares regression, which builds new predictor variables as linear combinations of the original variables. Consequently, this section firstly summarizes the *appearance-based* feature extraction process and then briefly outlines the One-Against-All PLS (OAA-PLS) regression.

After detecting, cropping and resizing the faces, each sample is decomposed into blocks. The low-level feature extraction is performed for all blocks from a cropped face. Then, these descriptors are concatenated into a high-dimensional feature vector so that it can describe the face.

A. Feature Extraction

An appropriate feature extraction methodology allows to efficiently represent relevant parts of an image as a compact feature vector. We employ image cropping and resizing before splitting images into several overlapping blocks, obtaining several feature descriptors from each window.

The feature extraction process starts with a face detection algorithm. Next, the image is clipped so that it can be placed into a grid that is comprised of several blocks. Then, instead of extracting features from the face image as a whole, the idea is to focus on the region within each block independently. By concatenating every block's histogram, we generate a local feature vector, which tends to be more robust to small variations when compared to changes in the whole face pattern.

So far, we utilized two feature descriptors: Circular Local Binary Pattern (CLBP) [27] and Patterns of Oriented Edge Magnitudes (POEM) [28]. CLBP defines a local neighborhood as a set of sampling points evenly spaced on a circle centered at the pixel yet to be labeled. Pixel values are bi-linearly interpolated whenever the sampling point is not in the center of a pixel. POEM is an oriented multi-resolution descriptor – capable of capturing variation-robust information – that meets three important principles: computational cost, robustness, and discriminative power.

B. Partial Least Squares

PLS is robust enough to manage multicollinearity in high-dimensional data. A great advantage of using PLS is that it attributes features weights according to their discriminatory capacity. Besides, it also works well when there are few face images to represent each subject in the dataset. In the considered approach, a PLS regression model is learned for each subject in the gallery following a one-against-all scheme.

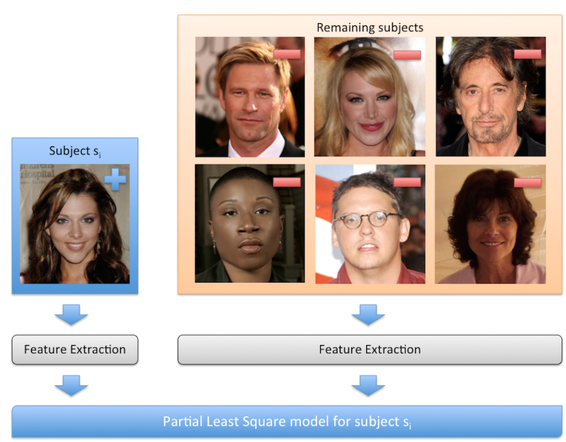

Fig. 1. Building PLS regression model for a subject in the gallery during training stage.

In the one-against-all approach, depicted in Figure 1, samples from the subject are learnt with response equal to $+1$ and samples from other subjects with response equal to -1. When the i-th individual is considered, all other subject's samples are used as counterexamples. In this case, the PLS regression model is learnt considering feature descriptors extracted from samples in the positive set with target values equal to $+1$ against samples in the negative set with target values equal to -1.

PLS regression models are used to determine whether a test sample belongs to the positive or negative subset. During the testing stage, probe samples are presented to each PLS model and associated to the identity related to the model that returns the highest score.

IV. Preliminary Experiments

The employed approach is evaluated on a recent non-constrained dataset, which is part of the Megaface Challenge [8] and on FRGC [20] (check it for experiments description). Initially, we test our method on datasets containing no additional individuals (Section IV-A); then, we add distractors (Section IV-B).

For the CLBP feature extraction, we used 24×24 pixel sliding windows and a 14-pixel stride. Moreover, we ended up setting the radius parameter to 9. POEM parameters: 3 bins, 7-pixel radius, 8×8-pixel cell, 32×32-pixel block, 16-pixel stride. This setup was implemented to both datasets: FRGC v1.0 and FaceScrub.

Metrics: Cumulative Match Characteristic (CMC) curves evaluate the capability of biometric systems that return a ranked list of candidates. Another curve, Rank-1 Identification Rate, takes into consideration only the CMC Rank-1 value. It describes a biometric system as a function of the growing number of subjects or distractors in the gallery.

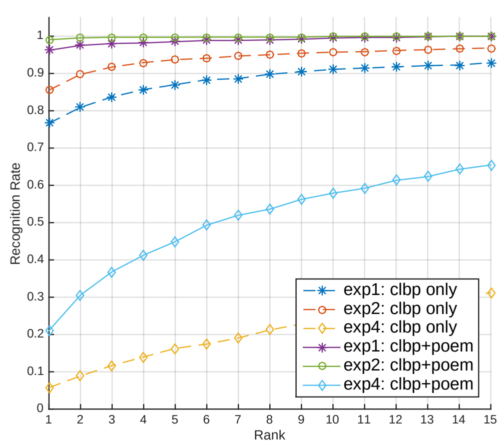

Fig. 2. CMC curves: Identification rates obtained for all considered FRGC experiments using CLBP alone or combined with POEM.

A. Standard Gallery

Figure 2 indicates low recognition rates in FRGC experiment four, even with the combination of feature descriptors. We understand that experiment four is quite challenging because it consists of a gallery set having just one controlled face image for each subject and a probe set composed of uncontrolled images only. Experiments one and two do not involve uncontrolled images and, consequently, attained satisfactory results. We illustrate how the number of training samples per subject influences recognition capability in Figure 3.

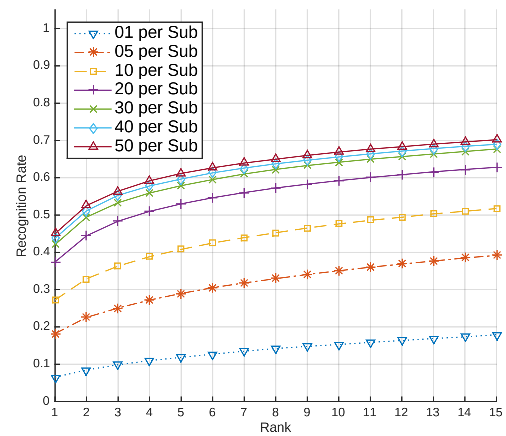

Fig. 3. CMC curves: Average recognition rates obtained for FaceScrub using CLBP+POEM with a variable number of samples per subject.

B. Addition of Distractors

Now, we analyze the performance of the employed approach method considering the addition of new individuals into the gallery. We present Rank-1 identification rate for both datasets in Figure 4. The experiments hold the maximum number of samples per subject: four and fifty images per individual for the FRGC (CLBP descriptor only) and FaceScrub, respectively. Note that these values do not apply to the number of samples per distractor, whose number of samples is fixed in one. Although we only add up to ten thousand distractors, the system accuracy loss is notable.

Just as we assumed in the beginning, algorithms efficiency deteriorates in increasingly galleries even when the probe set size does not vary. We believe that the significant fall observed on the FRGC curve can be explained by the low number of samples per subject, only four, when set side by side with FaceScrub's outcome, which involves fifty images for each individual. It is worth noting, based on these experiments, that algorithms trained on larger gallery sets – numerous samples per class – tend to perform better at scale regardless of the selected dataset, which brings out into open what once were subtle differences across many methods that do not cope with scalability.

V. Future Works

In this paper, we have detailed the challenges involving large gallery sets. Massive galleries contain numerous subjects,

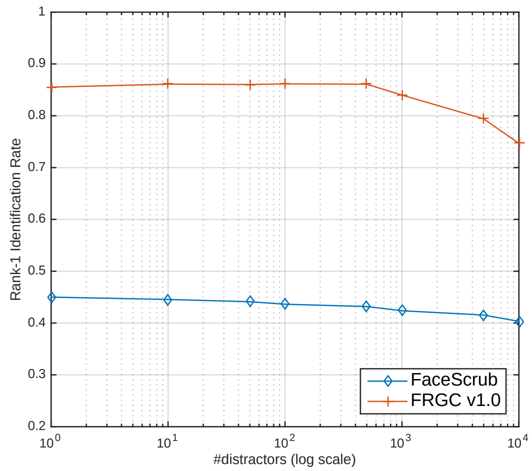

Fig. 4. Identification rate at Rank-1 obtained for FRGC (CLBP descriptor only) and FaceScrub with the addition of new individuals.

resulting in a large number of comparisons, turning the OAA-PLS approach very inefficient. Therefore, new methods must be studied to better engage in the scalable recognition problem.

We will investigate Space Partitioning (SP) since it iteratively splits the feature space into subsets of the same size. The space is randomly divided considering the axis holding the highest variance value in the original data. We are aware that tree-based space partitioning may take completely "wrong turns". Then, we have to boost it in case we opt for SP.

If we employ any Clustering Algorithm (CA), it will execute until the number of elements in every cluster drops below a fixed value. Some researchers describe clustering as a black hole since there is no generic way for identifying the K number of clusters [29]. We know that K-means is sensitive to outliers; so, even if a sample is far from any cluster centroid, it would be erroneously moved into its closest cluster.

We believe that LSH is a promising approach since it splits the entire data by setting up lots of random direction vectors. If two data points are close by, they are likely of being inserted to the same bucket. Some works limit LSH to the Hamming space [30]. Then, we may need to modify the LSH method to the l_2 norm – inserting l_2 space into the l_1 space and after that introducing l_1 space into the Hamming space.

ACKNOWLEDGMENTS

The authors would like to thank the Brazilian National Research Council CNPq (Grant #477457/2013-4), the Minas Gerais Research Foundation FAPEMIG (Grants APQ-00567-14 and PPM-00025-15) and the Coordination for the Improvement of Higher Education Personnel CAPES (DeepEyes Project).

REFERENCES

[1] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," University of Massachusetts, Tech. Rep., 2007.

[2] J. C. Neves, G. Santos, S. Filipe, E. Grancho, S. Barra, F. Narducci, and H. Proença, "Quis-campi: Extending in the wild biometric recognition to surveillance environments," in *ICIAP*. Springer, 2015.

[3] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *TIP*, 2010.

[4] D. Yi, Z. Lei, and S. Z. Li, "Towards pose robust face recognition," in *CVPR*, June 2013.

[5] X. Zhu, J. Yan, D. Yi, Z. Lei, and S. Z. Li, "Discriminative 3d morphable model fitting," in *AFGR*. IEEE, 2015.

[6] B. F. Klare, B. Klein, E. Taborsky, A. Blanton, J. Cheney, K. Allen, P. Grother, A. Mah, M. Burge, and A. K. Jain, "Pushing the frontiers of unconstrained face detection and recognition: Iarpa janus benchmark a," in *CVPR*. IEEE, 2015.

[7] H.-W. Ng and S. Winkler, "A data-driven approach to cleaning large face datasets," in *ICIP*. IEEE, 2014.

[8] I. Kemelmacher-Shlizerman, S. M. Seitz, D. Miller, and E. Brossard, "The megaface benchmark: 1 million faces for recognition at scale," in *CVPR*, 2016.

[9] D. Wang, C. Otto, and A. K. Jain, "Face search at scale: 80 million gallery," *arXiv*, 2015.

[10] Z. Zeng, T. Fang, S. Shah, and I. A. Kakadiaris, "Local feature hashing for face recognition," in *BTAS*. IEEE, 2009.

[11] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *CVPR*, June 2015.

[12] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *BMVC*, 2015.

[13] C. Ding and D. Tao, "Robust face recognition via multimodal deep face representation," *IEEE Transactions on Multimedia*, 2015.

[14] B.-C. Chen, Y.-Y. Chen, Y.-H. Kuo, and W. H. Hsu, "Scalable face image retrieval using attribute-enhanced sparse codewords," 2013.

[15] C. Otto, D. Wang, and A. K. Jain, "Clustering millions of faces by identity," *arXiv*, 2016.

[16] C. E. dos Santos, E. Kijak, G. Gravier, and W. R. Schwartz, "Partial least squares for face hashing," *ICN*, 2016.

[17] P. J. Grother, G. W. Quinn, and P. J. Phillips, "Report on the evaluation of 2d still-image face recognition algorithms," NIST, Tech. Rep., 2010.

[18] C. E. Dos Santos Jr, E. Kijak, G. Gravier, and W. R. Schwartz, "Learning to hash faces using large feature vectors," in *CBMI*. IEEE, 2015.

[19] P. J. Phillips, H. Moon, S. Rizvi, P. J. Rauss *et al.*, "The feret evaluation methodology for face-recognition algorithms," *PAMI*, 2000.

[20] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge," in *CVPR*. IEEE, 2005.

[21] T.-A. Pham, S. Barrat, M. Delalandre, and J.-Y. Ramel, "An efficient tree structure for indexing feature vectors," *ICPRL*, 2015.

[22] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in *ECCV*. Springer, 2006.

[23] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *IJCV*, 2004.

[24] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *IJCV*, 2001.

[25] S. Liao, Z. Lei, D. Yi, and S. Z. Li, "A benchmark study of large-scale unconstrained face recognition," in *IJCB*. IEEE, 2014.

[26] W. R. Schwartz, H. Guo, J. Choi, and L. S. Davis, "Face identification using large feature sets," *TIP*, 2012.

[27] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *TPAMI*, 2006.

[28] N.-S. Vu, H. M. Dee, and A. Caplier, "Face recognition using the poem descriptor," *ICPR*, 2012.

[29] M. Muja and D. G. Lowe, "Fast matching of binary features," in *CRV*. IEEE, 2012.

[30] M. Datar, N. Immorlica, P. Indyk, and V. S. Mirrokni, "Locality-sensitive hashing scheme based on p-stable distributions," in *Proceedings of the twentieth annual symposium on Computational geometry*. ACM, 2004.