Partial Least Squares for Face Hashing

Cassio E. dos Santos Jr., William Robson Schwartz
Department of Computer Science, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
cass@dcc.ufmg.br, william@dcc.ufmg.br

***Abstract*—Face identification is an important research topic for applications such as surveillance, forensics, and human-computer interaction. In the past few years, a myriad of methods for face identification has been proposed in the literature, with just a few among them focusing on scalability. In this work, we propose a simple but efficient approach for scalable face identification based on partial least squares (PLS) and random independent hash functions inspired by locality-sensitive hashing (LSH), resulting in the *PLS for hashing* (PLSH) approach. The original PLSH approach is further extended using feature selection to reduce the computational cost to evaluate the PLS-based hash functions, resulting in the state-of-the-art *extended PLSH* approach (ePLSH). The proposed approach is evaluated in the dataset FERET and in the dataset FRGCv1. The results show significant reduction in the number of subjects evaluated in the face identification (reduced to 0.3% of the gallery), providing averaged speedups up to 233 times compared to evaluating all subjects in the face gallery and 58 times compared to previous works in the literature.**

I. Introduction

In this work, we focus on the face identification task. Specifically, the main goal is to provide a face identification approach scalable to galleries consisting of numerous subjects and on which common face identification approaches would probably fail on responding in low computational time. There are several applications for a scalable face identification method: surveillance scenarios, human-computer interaction and social media.

The few aforementioned applications show the importance of performing face identification fastly and, in fact, several works in the literature have been developed in the past years motivated by these same types of applications (surveillance, forensics, human-computer interaction, and social media). However, most of the works focus on developing fast methods to evaluate one test face and a single subject enrolled in the gallery. These methods usually develop low computational cost feature descriptors for face images that are discriminative and with low memory footprint enough to process several images per second. Note that these methods still depend on evaluating all subjects in the face gallery. Therefore, if the number of subjects in the gallery increases significantly, these methods will not be able to respond fastly and new methods shall be developed to scale the face identification to this larger gallery.

Face identification methods usually consists of a face representation or description in the feature vector where mathematical models can be applied to determine the face identity. In this case, it is used one model to determine each identity in the face gallery, therefore, being necessary a number of models equal to the gallery size. Note that the parameters in each model are learned using samples for each subject in the face gallery and every model must be evaluated to correctly identify a test sample. In this work, we propose a method to reduce the number of models evaluated in the face identification by eliminating identities that are somewhat clearly not the identity in the test sample. Figure 1 illustrates the common face identification pipeline employed in practice and the main component tackled in this work.

There is an extensive literature of works regarding large-scale image retrieval that could be employed in face identification. However, most of these works focus on returning a list containing images from the dataset that are similar to the test image. Although reasonable to recover images in large datasets, such approaches are not suitable to apply directly to the face identification task. The models from subjects in the face gallery should optimally be described regarding the discriminative features related to each subject identity, which might consume less memory, specially if several samples per subject are available, and less computational time since only discriminative features are evaluated to determine the face identity.

The proposed approach is inspired by the family of methods regarded as *locality-sensitive hashing* (LSH), which are the most popular large-scale image retrieval method in the literature, and the *partial least squares* (PLS), which has been explored intensively in numerous past works regarding face recognition. We call the proposed approach PLS *for hashing*, abbreviated to PLSH and ePLSH in its extension. The main goal in LSH is to approximate the representation of samples in the high dimensional space using a small binary representation where the search can be implemented efficiently employing a hash structure to approximate near-identical binary representations. The idea in LSH is to generate random hash functions to map the feature descriptor in the high dimensional representation to bits in the binary representation.

In the PLSH approach, the random projection in the aforementioned example is replaced by a PLS regression, which provides discriminability among subjects in the face gallery and allow us to employ a combination of different feature descriptors to generate a robust description of the face image. PLSH is able to provide significant improvement over the brute-force approach (evaluating all subjects in the gallery) and compared to other approaches in the literature. Furthermore, since the evaluation of hash functions in PLSH requires a dot product between the feature and regression vectors, additional speedup can be achieved by employing feature selection

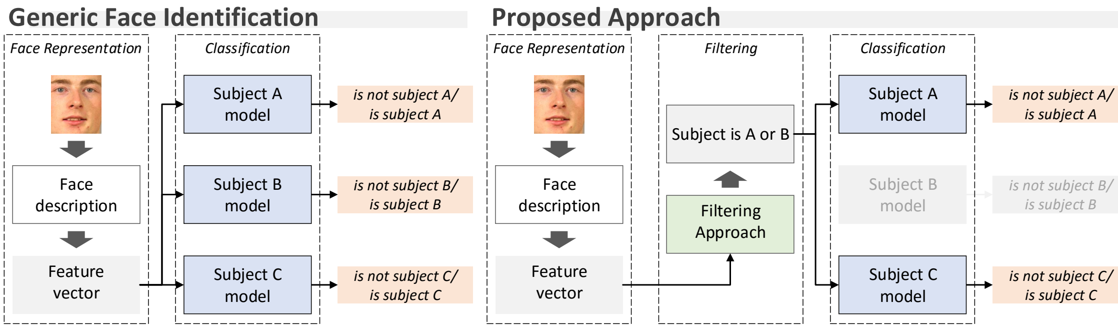

Fig. 1: Common face identification pipeline and the proposed pipeline with the filtering approach which is used to reduce the number of evaluations in the classification step with low computational cost. The filtering approach is the main contribution in this work and it is tailored considering recent advances in large-scale image retrieval and face identification based on PLS.

methods, resulting on the extended version of PLSH (ePLSH).

The following contributions are presented in this work. (i) A fast approach for face identification that support a combination of several feature descriptors and high dimensional feature vectors. (ii) The proposed approach presents at least comparable performance with other methods in the literature and up to 58 times faster when enough samples per subject are available for train. (iii) Extensive discussion and experimentation regarding alternative implementations that may guide future development in scalable face identification methods. (iv) The proposed approach is easy to implement and to deploy in practice since only two trade-off parameters need to be estimated.

This work resulted in the following publications in which the former presents PLSH and the latter presents ePLSH.

- C. E. Santos Jr, E. Kijak, G. Gravier, and W. R. Schwartz, "Learning to hash faces using large feature vectors," in *Content-Based Multimedia Indexing (CBMI), 13th IEEE International Workshop on*, 2015, pp. 1–6
- ——, "Partial least squares for face hashing," *Elsevier Neurocomputing – Special Issue on Binary Representation Learning in Computer Vision*, pp. 1–44, 2016, (to appear)

II. Methodology

This section describes the methods considered in the proposed approach, namely PLS for regression (Section II-A) and PLS for face identification (Section II-B). The proposed PLSH is described in Section II-C and in Section II-D, we describe a PLSH extension (ePLSH), which consists in employing PLS-based feature selection to improve the performance of PLSH.

A. Partial least squares regression

PLS is a regression method that combines ordinary least squares applied to a latent subspace of the feature vectors. Several works have employed PLS for face identification [3], face verification [4], and open-set face recognition [5]. These works consider PLS mainly due to the robustness to combine several feature descriptors, capability to deal with thousands of dimensions, and robustness to unbalanced classes. In this work, we consider PLS due to the high accuracy presented when used to retrieve candidates in PLSH and the low computational cost to test samples since only a single dot product between the regression coefficients and the feature vector is necessary to estimate the PLS response.

PLS is calculated as follows. The p-dimensional latent subspace is estimated by decomposing the zero mean matrices $X_{n\times d}$, with n feature vectors and d dimensions, and Y_n, with response values, in

$$\begin{aligned} X_{n\times d} &= T_{n\times p}P_{d\times p}^T + E_{n\times d}, \\ Y_{n\times 1} &= U_{n\times p}Q_{p\times 1} + F_{n\times 1}, \end{aligned} \tag{1}$$

where $T_{n\times p}$ and $U_{n\times p}$ denote latent variables from feature vectors and response values, respectively. The matrix $P_{d\times p}$ and the vector Q_p represent loadings and the matrix E and the vector F are residuals from the transformation. PLS algorithms compute P and Q such that the covariance between U and T is maximum [6]. We consider the nonlinear iterative PLS (NIPALS) algorithm [7] which calculates the maximum covariance between the latent variables $T = \{t_1, ..., t_p\}$ and $U = \{u_1, ..., u_p\}$ using the matrix $W_{d\times p} = \{w_1, ..., w_p\}$, such that

$$\arg\max[cov(t_i, u_i)]^2 = \arg\max_{|w_i|=1}[cov(Xw_i, Y)]^2.$$

The regression vector β between T and U is calculated using matrix W according to

$$\beta = W(P^TW)^{-1}(T^TT)^{-1}T^TY. \tag{2}$$

The PLS regression response $\hat{y}$ for a probe feature vector $x_{1\times d}$ is calculated according to $\hat{y} = \bar{y} + \beta^T(x - \bar{x})$, where $\bar{y}$ and $\bar{x}$ denote average values of Y and elements of X, respectively. The PLS model is defined as the variables necessary to estimate $\hat{y}$, which are β, $\bar{x}$ and $\bar{y}$.

B. Face identification based on partial least squares

The proposed approach consists in filtering subjects in the gallery using methods for large-scale image retrieval. For a given face identification approach, the evaluation of all subjects in the gallery (without filtering) is regarded as the *brute-force* approach, which is undesirable since the asymptotic time complexity is linear with the number of subjects enrolled in the gallery. The filtering approach consists in providing a shortlist to the face identification so that it evaluates only subjects presented in that shortlist.

The filtering and face identification pipeline consists of the following steps. Different feature descriptors are extracted from a probe sample and concatenated in the first step (feature extraction). Then, the combined feature vector is presented to the filtering step, which employs large-scale image retrieval methods to generate a list of candidates sorted in decreasing order of probability that the candidate is the subject in the probe. Then, a small number of high probability candidates in the list is provided to the face identification method, which evaluates subjects following the order in the candidate list until the face identification returns a subject in the face gallery. In this case, speedup is achieved because it is not necessary to evaluate the remaining subjects in the candidate list once a gallery match is found, reducing therefore, the computational cost compared to the brute-force approach.

To evaluate the filtering and face identification pipeline, we consider the face identification method described by Schwartz et al. [3], which consists in employing a large feature set concatenated to generate a high dimensional feature descriptor. Then, a PLS model is learned for each subject in the gallery following a *one-against-all* classification scheme: samples from the subject are learned with response equal to $+1$ and samples from other subjects with response equal to -1. Test samples are presented to each PLS model and associated to the identity related to the model that returns the maximum score. We consider the evaluation of all PLS models as the brute-force approach and, in the proposed pipeline, only PLS models that correspond to subjects in the candidate list are evaluated.

C. Partial least squares for face hashing (PLSH)

The PLSH method is based on two principles: (i) data dependent hash functions and (ii) hash functions generated independently among each other. Data dependent hash functions provide better performance in general. Hash functions generated independently are necessary to induce uniform distribution of binary codes among subjects in the gallery [8].

PLSH consists of the learn and the test steps. In the learn, for each hash model, subjects in the face gallery are randomly divided into two balanced subgroups, *positive* and *negative*. Then, a PLS regression model, regarded as hash function in this work, is learned to discriminate the subjects in the positive subset (response $+1$) from the subjects in the negative subset (response -1). The association of one subject to one of the two subsets consists in sampling from a Bernoulli distribution with parameter p equal to 0.5 and associating that subject to the positive subset in case of "success". Note that, the association to each subset can be viewed as a bit in the Hamming embedding and the Bernoulli distribution with p equal to 0.5 is important to distribute the Hamming strings uniformly among the subjects in the face gallery. A PLSH *hash model* is defined as a PLS model and the subjects in the positive subset necessary to evaluate the test samples.

In the test, the test sample (probe sample) is presented to each PLSH hash model to obtain a regression value r. We define a vote-list of size equal to the number of subjects in the gallery initially with zeros, then, each position of the vote-list is increased by r according to the indexes of subjects in the positive subset of the same PLSH hash model. Note that this scheme allows us to store half of the subject indexes to increment the vote-list since it will be equivalent to increment subjects in the negative set by $|r|$ when r is negative (the differences among pairs of votes will be the same). Finally, the list of subjects is sorted in decreasing order of values and presented as candidates for the face identification.

1) Relationship with Hamming embedding: We do not estimate the Hamming embedding directly since there is no binary string associated to any face sample. However, PLSH is equivalent to estimating the Hamming embedding for a test sample and comparing it with the binary strings generated for each subject in the gallery. In addition, each bit of the test binary string is weighted by the absolute value of the PLS regression response.

To demonstrate the aforementioned claims, consider that PLS responses can be only $+1$ or -1, such that any test sample can be represented by the sequence $X = \{+1, -1\}^H$, where H denotes the number of PLSH hash models. Consider also that each subject s in the face gallery is represented by the binary string $Y_s = \{1, 0\}^H$, where $y_i \in Y_s$ is set to 1 if the subject s was associated to the positive subset of the i-th PLSH hash model in the train step, or 0, otherwise. In this context, the weight w_s given by PLSH to each subject in the gallery is calculated as

$$w_s = \sum_{i=1}^{H} x_i y_i.$$

Note that the maximum w_s is equal to the sum of $+1$ elements in X, which occurs when $y_i = 1$, if $x_i = +1$, and $y_i = 0$, otherwise. Similarly, the minimum weight is equal to the sum of -1 elements in X, which occurs when $y_i = 1$, if $x_i = -1$, and $y_i = 0$, otherwise. If we transform X onto a binary string $\hat{X}$ such that $\hat{x}_i = 1$, if the corresponding x_i is $+1$, and $\hat{x}_i = 0$, otherwise; we can calculate the Hamming distance between $\hat{X}$ and Y_s. In fact, the exactly same Hamming distance can be calculate using w_s as

$$d(X, Y)_{\mathbb{H}} = w_{\max} - w_s, \tag{3}$$

where $w_{\max}$ denotes maximum possible w_s. The same analogy can be applied to the weighted Hamming distance if we consider x_i assuming any real number. In this case, the weight of each bit α_i is the absolute value of r and the weighted Hamming distance is equivalent to Equation 3.

D. Feature selection for face hashing (ePLSH)

The algorithms for PLSH described in Section II-C require a dot product between the PLS regression vector and the feature descriptor to calculate each hash function. This section describes methods to reduce the computational cost to evaluate hash functions. To discriminate PLSH with the feature selection version and to maintain consistence with the nomenclature given in our publications, PLSH with feature selection is called *extended PLSH* (ePLSH) in the rest of this work.

In practice, ePLSH is equivalent to PLSH when all features are considered to evaluate hash functions. The main advantage of ePLSH is the possibility of employing thousands of additional hash functions, resulting in considerable increase of the recognition rate while keeping low computational cost to calculate the hash functions. The common feature setup considered in the PLSH and in the ePLSH approaches consists in combining four feature descriptors, which leads to a feature vector with 120,059 dimensions. However, we show in our experiments that, for the feature set considered in this work, about 500 dimensions with an increased number of hash functions provides better candidate lists than PLSH with about the same computational cost.

The ePLSH consists of two steps: *train* and *test*. In the train, it calculates the β regression vector following the same procedure of PLSH. Then, the indexes of the k more discriminative features are stored. Considering that the range of values in the feature vector is known (zero mean and unit variance in our experiments), it is possible to calculate an approximated score using only the more discriminative features. However, if only such features are used to calculate the regression value without rebuilding the PLS model, the result would not be accurate because of the large number of remaining features, even though they present a very low contribution individually. To tackle this issue, we learn a new PLS model to replace the full feature version in PLSH, which is performed by eliminating the dimensions from the matrix X that do not correspond to the k select features and recalculate β using Equation 2.

We define the ePLSH *hash model* as the PLS model, the subjects in the positive subset and the k selected features. Finally, the test step is carried in the same manner as in PLSH, but with the difference that only features selected in the ePLSH hash model are considered to calculate the regression score.

E. Early-stop search heuristic

To stop the search for the correct subject in the candidate list, we employ the heuristic described by Schwartz et al. [3]. For a short number of initial samples (15), all subjects in the candidate list are evaluated and the median value of the scores is taken as threshold for the remaining test samples. Then, subjects in the candidate list are evaluated until a score equal or higher than the threshold is obtained or the end of the list is reached.

Note that, in practice, the candidate list size is a percentage of the subjects enrolled in the gallery and most of the candidates with low weights can be discarded because they rarely corresponds to the probe sample. In this case, the worst case scenario consists in evaluating all subjects in the candidate list for every probe sample. However, the early-stop search heuristic alone is shown to reduce the number of tests in the face identification up to 63% without degrading the recognition rate so the speedup achieved is usually higher than the ratio of the gallery size divided by the number of subjects in the candidate list.

III. Experimental setup

We evaluate PLSH and ePLSH in two standard face identification datasets, FERET and FRGCv1. The facial recognition technology (FERET) dataset [9] consists of $1,196$ images, one per subject for training, and four test sets designed to evaluate the effects of lightning conditions, facial expression and aging on face identification methods. The test sets are: *fb*, consisting of $1,195$ images taken with different facial expressions; *fc*, consisting of 194 images taken in different lightning conditions; *dup1*, consisting of 722 images taken between 1 minute and $1,031$ days after the gallery image; *dup2*, is a subset of *dup1* and consists of 234 images taken 18 months after the gallery image. In our experiments, all images were cropped in the face region using annotated coordinates of the face, scaled to 128×128 pixels and normalized using the self-quotient image (SQI) method to remove lightning effects [10].

The face recognition grand challenge dataset (FRGC) [11] consists of 275 subjects and samples that include 3D models of the face and 2D images taken with different illumination conditions and facial expressions. We follow the same protocol described by Yuan et al. [12], which considers only 2D images and consists in randomly selecting different percentages of samples from each subject to compose the face gallery and using the remaining samples to test. The process is repeated five times and the mean and standard deviation of the rank-1 recognition rate and speedup (considering the brute-force approach) are reported. The samples were cropped in the facial region, resulting in size 138×160 pixels, and scaled to 128×128 pixels.

All experiments regarding parameter validation were performed on the FERET dataset, since it is the dataset with the largest number of subjects ($1,196$ in total). FERET consists of four test sets and we use *dup2* to validate parameters since it is considered the hardest of the dataset.

We consider four feature descriptors in this work, CLBP [13], Gabor filters [14], HOG [15] and SIFT [16], which mainly captures information about texture and shape of the face image. This set of features was chosen because they present slightly better results in the face identification and indexing compared to the previous works [1], [3].

The error rate of the pipeline as described in Figure 1 results from errors induced by the filter approach (fail to return identity of test sample in the candidate list) and by the face identification approach (fail to identify correctly the subject in the candidate list). Therefore, to assess the performance of the filter approach alone, we provide results considering the *maximum achievable recognition rate* (MARR) [1], which is calculated considering that a perfect face identification method is employed for different percentages of candidates visited in the list. Note that the MARR value is the upper bound for the recognition rate achieved by the filter and face identification pipeline.

A. Experimental results

Results regarding MARR and rank-1 recognition rate for PLSH in all test sets from the FERET dataset are presented in Figures 2(a) and 2(b). For the test sets *fb* and *fc*, about 1% of subjects in the candidates list is enough to achieve more than 95% of the rank-1 recognition rate of the brute-force approach (presented in the legend of Figure 2(b) for each test set). However, for the test sets *dup1* and *dup2*, about 5% of subjects in the candidate list ensured at least 95% of the brute-force rank-1 recognition rate. The theoretical speedup in the worst case can be calculated considering the 150 PLSH hash function evaluations and the 5% of the gallery size, which consists of 60 PLS projections. In this case, the number of PLS projections would be 210 compared to the $1,196$ projections necessary in the brute-force approach, which would still results in a 5.6 times speedup.

Results from ePLSH are presented in Figures 2(c) and 2(d). Using only 1% of subjects in the candidate list, it is possible to recover all subjects in the rank-1 recognition rate from brute-force approach for all four test sets. In this case, the rank-1 recognition rate from the ePLSH pipeline is the same as the brute-force approach, but with reduction to 1% of the subjects evaluated in the identification. Considering that the cost to evaluate all hash models in ePLSH is about the same as in PLSH, the theoretical speedup is 7.38 times compared to the brute-force approach in the worst case.

Results from the FRGC dataset for PLSH and ePLSH are presented in Table I along with results from three other methods as presented in the literature. The three methods are the cascade of rejection classifiers (CRC) from [12], the PLS-based search tree [3], and our previous published work [1], which consists of PLSH with the combination of HOG, Gabor filter and LBP feature descriptors. For PLSH and ePLSH, we vary the number of hash models and the maximum percentage of subjects visited in the candidate list and we present the results with rank-1 recognition rate close to 0.95 and higher speedups. In this way, it is possible to compare directly the maximum speedup achievable when using PLSH and ePLSH compared to the other approaches, which also provide rank-1 recognition rate close to 0.95.

According to Table I, the speedup for PLSH and ePLSH decreases considerable as the number of samples per subject available for train reduce. The reason for that is the increase in the number of hash models and the maximum number of subjects visited in the candidate list to guarantee at least 0.95 rank-1 recognition rate. Even with reduced speedups considering 35% of samples available for train, ePLSH provides significant improvement over the speedup achieved by the tree-based approach (3.6 times faster), while PLSH provides competitive speedup.

The speedup provided by PLSH and ePLSH compared to the tree-based approach is noticed with 90% of the samples available for train, where PLSH is about 5 times faster than the tree-based approach while ePLSH is about 13 times faster than PLSH. Finally, in the worse case, ePLSH provides at least 14 times speedup considering the brute-force approach in the setup with 200 hash models and 10% of subjects in the candidate list.

IV. Conclusions

In this work, we proposed and evaluated PLSH and its extension ePLSH for face indexing. PLSH is inspired by the well-known locality-sensitive hashing for large-scale image retrieval and PLS for face identification, which provides fast and robust results for face indexing. Additional gain in speedup was achieved with the ePLSH, a method that employs PLS-based feature selection to reduce the computational cost to evaluate hash functions, enabling a large amount of additional hash functions to be employed and raising the indexing precision. We evaluated several parameters and alternative implementations of PLSH in the hope that they will be useful for future face indexing development. The experiments were conducted on two face identification standard datasets, FERET and FRGCv1, with $1,196$ and 275 subjects, respectively. Although these datasets do not provide enough number of subjects for a proper evaluation regarding scalability to large galleries, PLSH and ePLSH still provide significant improvement in speedup compared to other scalable face identification approaches in the literature.

The conclusions and considerations regarding PLSH and ePLSH are the following: (i) they support for high dimensional feature vectors, allowing different complementary feature descriptors to be employed to increase the robustness of the face indexing; (ii) they are easy to implement and deploy in practice since the only parameters needed to be set are the number of hash models and subjects in the candidate list. (iii) they do not provide good performances when the number of samples per subject is reduced and (iv) incremental enrollment of subjects in the framework requires re-training of the hash models, which may be prohibitive to perform in practice, specially for ePLSH which demands considerable more hash models.

Acknowledgments

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

References

[1] C. E. Santos Jr, E. Kijak, G. Gravier, and W. R. Schwartz, "Learning to hash faces using large feature vectors," in *Content-Based Multimedia Indexing (CBMI), 13th IEEE International Workshop on*, 2015, pp. 1–6.

[2] ——, "Partial least squares for face hashing," *Elsevier Neurocomputing – Special Issue on Binary Representation Learning in Computer Vision*, pp. 1–44, 2016, (to appear).

[3] W. R. Schwartz, H. Guo, J. Choi, and L. S. Davis, "Face identification using large feature sets," *Image Processing, IEEE Transactions on*, vol. 21, no. 4, pp. 2245–2255, 2012.

[4] H. Guo, W. R. Schwartz, and L. S. Davis, "Face verification using large feature sets and one shot similarity," in *Biometrics (IJCB), 2011 International Joint Conference on*. IEEE, 2011, pp. 1–8.

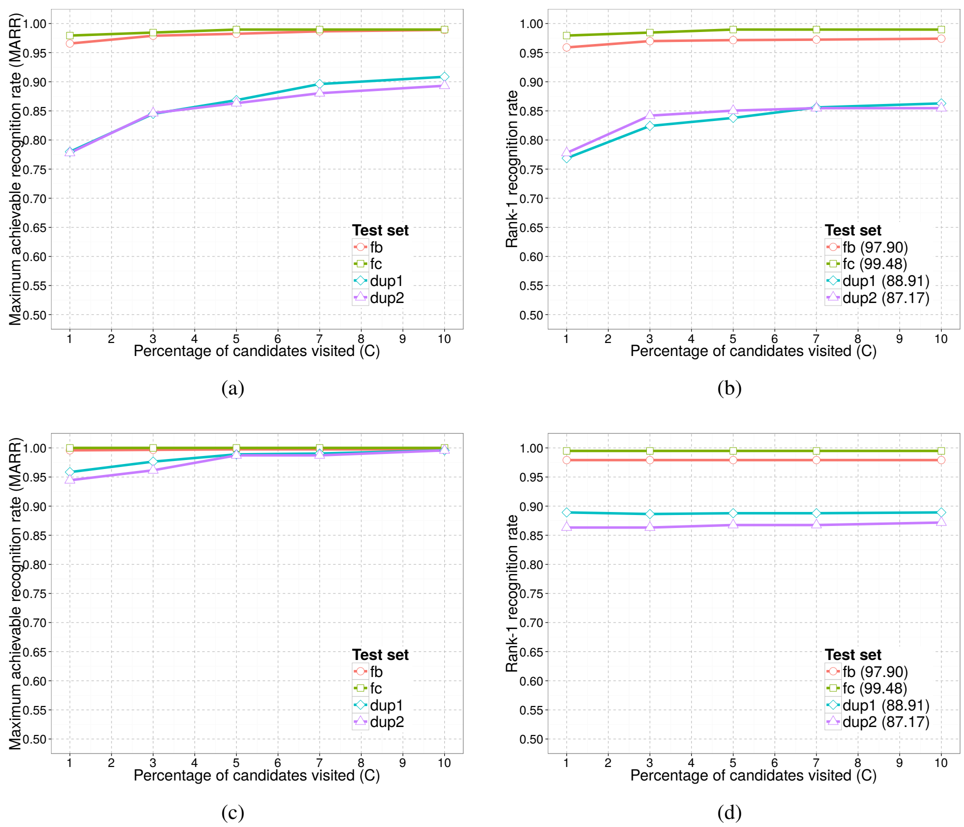

Fig. 2: Results on the FERET dataset. (a) PLSH MARR curves, (b) PLSH rank-1 recognition rate, (c) ePLSH MARR curves and (d) ePLSH rank-1 recognition rate. Number in parenthesis indicate rank-1 recognition rate for the brute force approach.

	% of samples for train	90%	79%	68%	57%	35%
CRC [12]	Speedup	1.58×	1.58×	1.60×	2.38×	3.35×
	Rank-1 rec. rate	80.5%	77.7%	75.7%	71.3%	58.0%
Tree-based [3]	Speedup	3.68×	3.64×	3.73×	3.72×	3.80×
	Rank-1 rec. rate	94.3%	94.9%	94.3%	94.46%	94.46%
PLSH	Speedup	18.24×	8.61×	6.95×	3.96×	3.49×
	Rank-1 rec. rate	95.31%	95.31%	93.60%	94.67%	94.60%
ePLSH	Speedup	**233.61**×	**98.93**×	**45.42**×	**22.29**×	**14.21**×
	Rank-1 rec. rate	96.03%	95.02%	95.98%	94.67%	94.44%

TABLE I: Comparison between the proposed approach and other approaches in the literature. The highest speedups are shown in bold. The full table with more information can be found in page 58 of the dissertation.

[5] C. E. Santos Jr and W. R. Schwartz, "Extending face identification to open-set face recognition," in *Graphics, Patterns and Images (SIBGRAPI), 27th IEEE Conference on*, 2014, pp. 188–195.

[6] R. Rosipal and N. Krämer, "Overview and recent advances in partial least squares," in *Subspace, latent structure and feature selection*. Springer, 2006, pp. 34–51.

[7] H. Wold, "Partial least squares," in *Encyclopedia of Statistical Science*, S. Kotz and N. Johnson, Eds. New York: Wiley, 1985, pp. 581–591.

[8] A. Joly and O. Buisson, "Random maximum margin hashing," in *Computer Vision and Pattern Recognition (CVPR). IEEE Conference on*, 2011, pp. 873–880.

[9] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1090–1104, 2000.

[10] H. Wang, S. Z. Li, and Y. Wang, "Face recognition under varying lighting conditions using self quotient image," in *Automatic Face and Gesture Recognition*, 2004, pp. 819–824.

[11] 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 *Computer vision and pattern recognition (CVPR). IEEE Conference on*, vol. 1, 2005, pp. 947–954.

[12] Q. Yuan, A. Thangali, and S. Sclaroff, "Face identification by a cascade of rejection classifiers," in *Computer Vision and Pattern Recognition (CVPR) - Workshops. IEEE Conference on*, 2005, pp. 152–152.

[13] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 12, pp. 2037–2041, 2006.

[14] T. Randen and J. H. Husoy, "Filtering for texture classification: A comparative study," *Pattern Analysis and Machine Intelligence*, vol. 21, no. 4, pp. 291–310, 1999.

[15] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society Conference on*, vol. 1, 2005, pp. 886–893.

[16] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.