Learning to Hash Faces Using Large Feature Vectors
Cassio dos Santos Jr., Ewa Kijak, Guillaume Gravier, William Robson Schwartz

To cite this version:
Cassio dos Santos Jr., Ewa Kijak, Guillaume Gravier, William Robson Schwartz. Learning to Hash Faces Using Large Feature Vectors. International Workshop on Content-based Multimedia Indexing, 2015, Prague, Czech Republic. hal-01186444

HAL Id: hal-01186444
https://hal.archives-ouvertes.fr/hal-01186444
Submitted on 28 Aug 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Learning to Hash Faces Using Large Feature Vectors

Cassio E. dos Santos Jr.*, Ewa Kijak†, Guillaume Gravier‡, William Robson Schwartz*
* Department of Computer Science, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
† IRISA & Inria Rennes (CNRS, Univ. Rennes 1), Campus de Beaulieu, Rennes, France
cass@dcc.ufmg.br, ewa.kijak@irisa.fr, guillaume.gravier@irisa.fr, william@dcc.ufmg.br

Abstract—Face recognition has been largely studied in past years. However, most of the related work focus on increasing accuracy and/or speed to test a single pair probe-subject. In this work, we present a novel method inspired by the success of locality sensing hashing (LSH) applied to large general purpose datasets and by the robustness provided by partial least squares (PLS) analysis when applied to large sets of feature vectors for face recognition. The result is a robust hashing method compatible with feature combination for fast computation of a short list of candidates in a large gallery of subjects. We provide theoretical support and practical principles for the proposed method that may be reused in further development of hash functions applied to face galleries. The proposed method is evaluated on the FERET and FRGCv1 datasets and compared to other methods in the literature. Experimental results show that the proposed approach is able to speedup 16 times compared to scanning all subjects in the face gallery.

I. INTRODUCTION

Face recognition consists of three tasks described as follow [1]. Verification considers 1-1 tests where the goal is to verify whether two samples belong to the same subject. Identification consists in a 1-N test where a sample is compared to a gallery containing \(N\) subjects. Open-set is the same as identification but considers that the probe subject may not be enrolled in the gallery. In this work, we consider face identification, which is important for numerous applications, including identification of clients in social media, human-computer interaction, search for interviews of a specific person in TV broadcast footages, search for suspects in videos from cameras for forensics or surveillance purposes, and in image databases. These applications present distinct aspects and challenges that have been investigated on other works and, as a result, it is possible to find a large variety of face identification approaches in the literature.

Although several works for face identification have been developed, only few of them target scalability regarding large galleries [2]–[4]. Our goal is to develop an approach to return the correct identity of a probe sample in an affordable time even for galleries containing thousands of subjects. One possible approach is to reduce the computational burden to test pairs of probe and subject in the gallery. However, the number of tests required to return the correct identity will be linear to the number of enrolled subjects. We are interested in, given a probe, discard with low computational cost most of the candidates in the gallery that are less likely to correspond to the identity of the probe sample. In this paper, we achieve both low cost to filter candidates and sub-linear complexity with respect to the number of subjects.

To address face identification on large galleries, we consider the well-known LSH method for large general purpose datasets and PLS applied to face identification using feature combination. The proposed approach consists in generating hash functions based on PLS and large feature vectors to create a short list of candidates for further use in face identification methods. Theoretical and practical discussion presented on the design of the algorithm imply that at least \(\lceil \log_2(N) \rceil\) hash function evaluations are required to compute the candidate list for a gallery with \(N\) subjects. Furthermore, the combination of different feature descriptors using PLS is shown to increase significantly the recognition rate with the candidate list compared to single feature descriptors. Provided that hash functions are based on PLS, a simple dot product is required to compute each hash function, thus, reducing the time necessary to retrieve the candidate list.

The remainder of the paper is organized as follow. In Section II, we review works in the literature related to face identification and large-scale image retrieval. In Section III, we present the hashing scheme based on PLS regression. In Section IV, we evaluate the proposed approach on the FRGC and FERET datasets. Finally, we conclude this paper with final remarks in Section V.

II. RELATED WORKS

Due to lack of space, we briefly review face identification methods based on regular 2D images We refer the reader to [5] for a more detailed presentation of different methods for face identification.

A. Face Identification

Face identification methods are divided in two categories [6]: appearance-based (holistic), feature-based (local). In holistic methods [3], [7], the whole face image is represented using feature descriptors extracted from a regular grid or from overlapping rectangular cells. The advantage of holistic methods is the easy encoding of overall geometric disposition of patches in the final descriptor. However, a preprocessing step is usually required to account for misalignment, lighting conditions, and pose variations. Feature descriptors commonly employed in holistic methods include local binary patterns (LBP) and Gabor filters [7].

Methods based on local description [8], [9] represent face images in a sparse manner by extracting feature descriptors from regions around fiducial points, such as nose tip or corners, eye and mouth. Regarding the detection of interest points, the authors in [9] consider a fiducial point detector while the authors in [8] estimate salient points in the face image for further match with other salient points from other face images. The advantage of local description is the robustness regarding pose changes and partial occlusions. That is, even if some points are
occluded or shaded, other points can be used for identification, differently from the holistic representation. Although robust, local description methods neglect overall appearance of the face image and they often produce ambiguous description since only a few pixels of the image are considered to compute the final descriptor.

In recent years, sparse representation-based classification (SRC) [10] has been exploited, yielding good performances in face identification datasets. The method consider probe images as a linear combination of a dictionary composed by samples in the face gallery. Although the original approach requires a large number of samples per subject, SRC has been further extended to support single samples [11]. In [12], the authors propose to represent the dictionary as centroids of subject samples and their relative differences in order to account for uncontrolled acquired images in the face gallery.

B. Large-scale Image Retrieval

In the last section, we discussed works related to face identification and, since we are interested on methods to speedup face identification in large face galleries, in this section, we discuss some works related to large-scale image retrieval and scalable face identification. The goal in image retrieval is to retrieve the closest image in the gallery to a test sample considering a similarity metric between two or more images. In face identification, the metric is related to the likelihood that an image, or a set of images, represents the same subject in the face gallery. Although efficient algorithms to the exact closest match exist, they result in poor performance on high dimensional feature spaces, which is often the case when dealing with image retrieval. To solve this problem, several algorithms for approximate search have been proposed in the literature, including the popular locality sensing hashing (LSH) [13] which we further describe.

The idea of LSH is to use hash functions to map similar inputs to the same position in a hash table. Regarding the type of the hash function considered, LSH methods can be divided in two categories [13]: data independent and data dependent. On one hand, data independent hash functions are defined regardless of the data distribution and their main advantage is the fast enrollment of new samples in the dataset. An example of data independent hash function is presented in [2] and consists in generating random regression vectors using a \( p \)-stable distribution. On the other hand, data dependent hash functions analyze the data to take more advantage of the data distribution. A hash function based on maximum margin separation among images in the gallery was presented in [14].

Regarding fast face identification, methods based on distance metrics usually employ compact representation of face images, such as binary patterns [15] to speedup image-image comparisons. In SRC methods, fast optimization algorithms have been proposed to compute the linear transformation among probe and gallery samples [16]. Although the aforementioned methods provide considerable speedup over probe-gallery pair comparison, their complexity depends on the number of subjects enrolled in the face gallery. To tackle this problem, a rejection tree was proposed in [3] to narrow down quickly the number of subjects in the gallery to a small list of candidates, and a rejection cascade was proposed in [4] to discard the majority of subjects in the initial steps of the test using low computational cost weak classifiers.

Different from the aforementioned methods, the approach proposed in this work considers the benefits of LSH. In this context, our approach is more similar to [2], which uses random regression vectors to hash faces. However, instead of building random regressions, we consider PLS models inspired by the LSH principles presented in [14]: (i) data dependent hash functions and (ii) hash functions generated independently. The advantage of using PLS is that we can use combination of features in a high dimensional descriptor vector to achieve higher recognition rates [3].

III. PROPOSED APPROACH

The proposed method is inspired by two works in the literature: The first considers face identification based on large feature sets and PLS for simultaneous dimensionality reduction and classification [3], and the second based on independent hash functions [14]. An overview of the approach and a brief summary of its principles are presented in Figure 1. In the next sections, we explain Partial Least Squares for dimension reduction and regression (Section III-A) and the proposed hashing function based on PLS (Section III-B).

A. PLS regression and face identification

PLS is a statistical method composed of a dimensionality reduction step followed by a regression step in the low dimensional space. Dimensionality reduction consists in determining latent variables as linear transformation of the original feature vectors and target values, then, ordinary least squares is used to predict target values using latent variables from feature vectors. The advantages of PLS for face identification are robustness to unbalanced classes and support for high dimensional feature vectors. These advantages are presented in [3] and [17] where one sample per subject in the gallery is available for training and where several feature descriptors are concatenated in order to account for weaknesses of single feature descriptors.
The relationship among feature descriptors and target values is given as $X = TP^T + E$ and $Y = UQ^T + F$, where $X_{n \times d}$ denotes a zero-mean feature descriptor matrix with $n$ samples and $d$ dimensions, $Y_d$ denotes a zero-mean target vector where each row $y_i$ corresponds to the $i$-th feature vector $x_i$ of $X$. The matrices $T_{n \times p}$ and $U_{n \times p}$ denotes latent variables from feature vectors and target values, respectively. The matrix $P_{p \times d}$ and the vector $Q$ are loading matrices similar to PCA transformations. Finally, $E$ and $F$ represent residuals. PLS algorithms compute $P$ and $Q$ such that the covariance between $U$ and $T$ is maximum. In order to compute PLS, we consider the NIPALS algorithm [18] which output a weight matrix $W_{d \times p} = \{w_1, ..., w_p\}$ such that $\text{cov}(\{t_1, u_i\})^2 = \arg \max_{y_i=1} |\text{cov}(xw_i, y)|^2$. The regression vector $\beta$ between $T$ and $U$ is calculated using least squares according to $\beta = W(P^TW)^{-1}TY$. A PLS regression response $\hat{y}$ for a probe feature vector $x$ is calculated according to $\hat{y} = \hat{y} + \beta^T(x - \bar{x})$, where $\hat{y}$ and $\bar{x}$ denotes average values of $Y$ and elements of $X$, respectively. A PLS model is then defined as the variables required to estimate $\hat{y}$ ($\beta$, $\bar{x}$ and $\hat{y}$).

To evaluate the gain obtained by the proposed approach with a real face identification method, we consider the work based on PLS described in [3]. The face identification method consists in learning one PLS model for each subject in the gallery following a one-against-all classification scheme. In this context, target values are set to +1 if the sample refers to the subject being considered or −1 otherwise. During test, samples are presented to each PLS model and their identities are assigned to the subject in the gallery related to the PLS model with maximum regression response.

B. PLS hashing

The proposed approach is based on two principles: (i) hash functions that consider the distribution of the data (data dependent) and (ii) hash functions generated independently among each other. As discussed in [14], independently generated hash functions are desirable to achieve uniform distribution of data in the hash table. Both principles are achieved following the steps presented in Figure 1. In the next paragraphs, we provide theoretical support, details, and practical design of the approach.

The approach consists in two steps: learning and test. On learning, we randomly split subjects in the gallery in two subsets: positive and negative. The split is performed as follows. For each subject, we sample from a Bernoulli distribution with parameter $p = 0.5$ and associate the subject to the positive subset in case of “success”. Then, a PLS model is learned 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. This process is repeated several times1. A hash models is defined by a single PLS model and the subjects in the positive subset.

In the test step, we extract the same feature descriptors employed on the training for the probe image and present them to each PLS model to obtain a regression value $r$. Then, we increase by $r$ each position of a weight-vector (initially zero) according to the indexes of subjects in the positive subset.

1We repeat 150 times on FERET and 10-35 on FRGC datasets. The number of repetitions depends on $\lceil \log_2(\#\text{subjects}) \rceil$ and the dataset difficulty.
Note that independence when drawing bits implies in independent hash functions. In the other hand, if we instead assign codes systematically among subjects, e.g., following a binary counting sequence \((001, 010, \ldots, 111)\), the probability to assign a specific code to a subject will depend on codes already assigned to other subjects, breaking independence among the hash functions. If we systematically assign codes to subjects, we also limit the combinatorial number of binary subsets resulting in hash functions biased toward the order of codes assigned to subjects. Figure 2 illustrates the advantage of independent hash functions.

In practice, we learn more than \(B\) hash functions to reduce the number of collisions in the hash table when we independently draw bits from a probability distribution. We also expect that some hash functions will miss some bits (change one bit for another), However, considering unbiased also expect that some hash functions will miss some bits independently draw bits from a probability distribution. We reduce the number of collisions in the hash table when we independent hash functions. In the other hand, if we instead assign codes systematically among subjects, e.g., following a binary counting sequence \((001, 010, \ldots, 111)\), the probability to assign a specific code to a subject will depend on codes already assigned to other subjects, breaking independence among the hash functions. Figure 2 illustrates the advantage of independent hash functions.

Considering the Bernoulli distribution, Equation 1 is rewritten as

\[
P(\text{code}) = p^k(1-p)^{B-k} = \frac{1}{N}, \forall k \in \{0, \ldots, B\}, \quad (2)
\]

where \(k\) is the number of bits in the code that are equal to 1. Expanding Equation 2, we have

\[
P(\text{code}) = p^B = p^{B-1}(1-p) = \ldots = (1-p)^B = \frac{1}{N}, \quad (3)
\]

implying in \(p = 1 - p = 0.5\). It can also be solved as

\[
p^B = \frac{1}{N} \implies log_p(\frac{1}{N}) = log_2(N) \implies p = 0.5. \quad (4)
\]

It is possible to demonstrate that \(p = 0.5\) minimizes the expected number of collisions in the hash table. We experimented changing \(p\) to 0.3 and 0.7 and both values resulted in poor performance. Based on the aforementioned discussion, we can conclude that (i) the robustness of the hash functions depends on how well a classifier can distinguish between two random subsets of subjects, and (ii) each subset must hold half of the subjects.

IV. EXPERIMENTS

Herein we evaluate the proposed approach. In Section IV-A we describe general setup, such as the number of factors on PLS models and parameters regarding the feature descriptors considered. In Section IV-B, we evaluate feature combination, number of hash functions, stability, and results with face identification on FERET dataset, since it is the dataset with the highest number of subjects considered in our experiments. In Section IV-C, we evaluate the proposed approach on the FRGC dataset and compare with other methods in the literature.

A. Experimental Setup

Feature descriptors considered are HOG, Gabor filters, and LBP. To compute Gabor features, we convolve the face image with squared filters of size 16 in 8 scales, equally distributed between \([0, \pi/2]\), and 5 orientations, equally distributed between \([0, \pi]\), resulting in 40 convolved images. The convolved images are downsampled by a factor of 4 and concatenated to form the Gabor feature descriptors. Two feature descriptor setups are used for HOG. The first consists in block size of \(16 \times 16\) pixels with stride of 4 pixels and cell size equal to 4 pixels. The second consists in blocks of \(32 \times 32\) pixels with stride of 8 pixels and cell size 8 pixels. For LBP, we consider the feature descriptor as the image resulted from the LBP transformation. The final feature vector is computed by concatenating features from the two HOG setups, Gabor and regular LBP applied to the image. The size of the final feature vector is 93,196.

The only parameter to build the PLS models is the number of factors (number of dimensions of the generated latent subspace). We tested varying the number of factors between 10 and 20 on FERET but the result was similar for any number of factors. Therefore, we consider 20 factors in all experiments.

All experiments were conducted on an Intel Xeon W3550 processor, 3.07 GHz with 64 GB of RAM running Fedora 16 operating system. All tests were performed using a single CPU core and no more than 8 GB of RAM were required during learn or test.

B. Results on the FERET dataset

The FERET [19] contains 1196 subjects, each having one image for training, and four test sets designed to evaluate robustness to illumination change, facial expression and aging. The test sets are: \(fb\), 1195 images taken with different expressions; \(fc\), 194 images taken in different lightning conditions; \(dup1\), 722 images taken between 1 minute and 1031 days after the training image; and \(dup2\), which is a subset of \(dup1\) with 234 images taken 18 months after the training image. All images are cropped in the face region and scaled to \(128 \times 128\) pixels. We use images with corrected illumination using the self-quotient images (SQI) method [20] kindly provided by the authors of [3].

Since \(dup2\) is considered the hardest test set of FERET, we use \(dup2\) to evaluate the combination of features, the stability, and the number of hash functions of the proposed approach. To isolate errors of the proposed approach from errors of the face identification, the plots are presented as the maximum achievable recognition rate, which is calculated considering that a perfect face identification method is employed for different percentages of candidates visited in the list. Preferable results present high maximum achievable recognition rate and low percentage of candidates visited, which, in general, represent curves close to the upper left corner of the plots. Finally, we evaluate the proposed approach on all four test sets from FERET dataset considering the face identification approach in [3].

Feature combination. To assess that different feature descriptors contribute to better accuracy of the hash functions, we present, in Figure 3, the maximum achievable recognition rate when providing different percentage of the candidate list to the face identification. According to the results, the feature combination enhances the recognition rate for about 9% compared to the best single features (Gabor feature descriptors).

Number of hash functions. Figure 4 shows the maximum recognition rate achievable for 50,100,150, and 200 hash functions. Significant improvement is achieved from 50 to 100 and from 100 to 150 hash functions. However, small improvement
were cropped in the facial region in the size variation in illumination and facial expression. The images [4]. FRGC is consists in 275 subjects and images presenting to compare the proposed approach with other methods [3], scanning all subjects. 95%

We can see that only one with rank-1 recognition rates achieved when presenting the candidate list to the face identification method (Figure 6b). Finally, we evaluate the proposed approach demonstrated weak performance when a few samples are available for training because we have to increase either the number of hash functions or the maximum number of candidates searched in the list. The conclusion is that the performance of the proposed approach depends on a high number of samples for training. Since the proposed approach does not consider face identities, we can try to include additional samples for learning on future works. We believe that the speedup of the proposed approach compared to the tree-based approach [3] is related to the independence of the hash functions since both approaches build PLS models in a similar manner.

V. CONCLUSIONS

We proposed a novel approach for hashing faces based on LSH and PLS regression. The hash functions are learned considering balanced random partitions of subjects in the face gallery, which we demonstrated to be the best option to avoid collision between two subjects. The performance of the proposed approach is simplified to how well a classifier distinguishes between two random subsets of subjects in the face gallery. To learn robust classifiers, we consider a combination of feature descriptors and PLS regression models, which are appropriated for high dimensional feature vectors. Finally, the proposed approach demonstrated weak performance when a reduced amount of samples is available for training.
Fig. 6: (a) Maximum achievable recognition rate of the proposed approach. (b) Rank-1 recognition rate achieved when the candidate list is presented to the face identification method [3]. Rank-1 recognition rate for scanning all candidates is shown in parenthesis for each experiment.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Test set</td>
<td>fb (97.65)</td>
<td>dup1 (86.98)</td>
<td>dup2 (84.61)</td>
</tr>
<tr>
<td>dup2</td>
<td>(98.96)</td>
<td>(88.98)</td>
<td>(84.61)</td>
</tr>
<tr>
<td>dup1</td>
<td>(86.98)</td>
<td>(84.61)</td>
<td>(84.61)</td>
</tr>
<tr>
<td>fb</td>
<td>(97.65)</td>
<td>(98.96)</td>
<td>(84.61)</td>
</tr>
</tbody>
</table>

TABLE I: Comparison between the proposed approach and other approaches in the literature. Higher speedups are shown in bold.

ACKNOWLEDGMENTS

The authors would like to thank the Brazilian National Research Council – CNPq (Grants #487529/2013-8 and #477457/2013-4), CAPES (Grant STIC-AMSUD 001/2013), and the Minas Gerais Research Foundation - FAPEMIG (Grants APQ-00567-14 and CEX – APQ-03195-13). This work was partially supported by the STIC-AmSud program.

REFERENCES