Similarity-Preserving Binary Signature for Linear Subspaces

Jianqiu Ji*, **Jianmin Li***, **Shuicheng Yan**†, **Qi Tian**‡, **Bo Zhang***
*State Key Laboratory of Intelligent Technology and Systems,
Tsinghua National Laboratory for Information Science and Technology (TNList),
Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China
jijq10@mails.tsinghua.edu.cn, {lijianmin, dcszb}@mail.tsinghua.edu.cn
†Department of Electrical and Computer Engineering, National University of Singapore, Singapore, 117576
eleyans@nus.edu.sg
‡Department of Computer Science, University of Texas at San Antonio, qi.tian@utsa.edu

Abstract

Linear subspace is an important representation for many kinds of real-world data in computer vision and pattern recognition, e.g. faces, motion videos, speeches. In this paper, first we define pairwise angular similarity and angular distance for linear subspaces. The angular distance satisfies non-negativity, identity of indiscernibles, symmetry and triangle inequality, and thus it is a metric. Then we propose a method to compress linear subspaces into compact similarity-preserving binary signatures, between which the normalized Hamming distance is an unbiased estimator of the angular distance. We provide a lower bound on the length of the binary signatures which suffices to guarantee uniform distance-preservation within a set of subspaces. Experiments on face recognition demonstrate the effectiveness of the binary signature in terms of recognition accuracy, speed and storage requirement. The results show that, compared with the exact method, the approximation with the binary signatures achieves an order of magnitude speed-up, while requiring significantly smaller amount of storage space, yet it still accurately preserves the similarity, and achieves high recognition accuracy comparable to the exact method in face recognition.

Introduction

In computer vision and pattern recognition applications, linear subspace is an important representation for many kinds of real-world data, e.g. faces (Basri and Jacobs 2003)(He et al. 2005)(Cai et al. 2006)(Cai et al. 2007), motion videos (Liu and Yan 2011)(Basri, Hassner, and Zelnik-Manor 2011)(Liu et al. 2013), speeches (Basri, Hassner, and Zelnik-Manor 2011) and so on. In the following, we use the term subspace for short. Once subspace is used as a representation, the first key problem is to define a proper pairwise similarity or distance between subspaces. Furthermore, storing and processing large number of subspaces with high ambient dimensions require expensive storage and computational cost. A possible solution would be representing subspaces with short signatures that preserve the pairwise similarity.

In this paper, first we define the concept of angle, pairwise angular similarity and angular distance between subspaces, and show that the angular distance is a metric that satisfies non-negativity, identity of indiscernibles, symmetry and triangle inequality. Then we develop a method to produce similarity-preserving binary signatures for subspaces, the Hamming distance between which provides an unbiased estimate of the pairwise angular distance. We provide a lower bound on the length of the binary signatures that suffices to guarantee uniform distance-preservation within a set of subspaces, which is similar to the Johnson-Lindenstrauss Lemma on random projection. We conduct face recognition experiments on the union of two popular face datasets to show that the angular distance as a distance measurement is discriminative. The results also show that using the binary signature would achieve more than 60 times faster recognition with no accuracy loss, and the storage space for the binary signatures is at most 1/190 that of storing the subspaces.

To the best of our knowledge, this is the first work on developing similarity-preserving binary signature for subspaces. The advantages of similarity-preserving binary signature are two-fold: the storage space is significantly reduced due to the use of compact binary signatures; on the other hand, pairwise angular similarity or distance can be efficiently estimated by computing the Hamming distance between the corresponding binary signatures, and Hamming distance can be computed very fast in modern CPUs (He, Wen, and Sun 2013). When the ambient dimension is high and the number of subspaces is huge, the similarity-preserving binary signature would serve as a fundamental tool towards different kinds of large-scale applications, e.g. clustering, approximate nearest neighbor search and so on.

Related Work

In recent years, intensive attention and research efforts have been devoted to developing small codes for vectors (Charikar 2002)(Andoni and Indyk 2006)(Torralba, Fergus, and Weiss 2008)(Weiss, Torralba, and Fergus 2008)(Kulis and Darrell 2009)(Gong and Lazebnik 2011)(Wang, Kumar, and Chang 2012)(Ji et al. 2012)(He, Wen, and Sun 2013), in purpose of database compression and fast approximate nearest neighbor search. To the best of our knowledge, little research has been done on developing small codes for subspaces.

Various definitions of distance and similarity have been proposed for subspaces (Basri, Hassner, and Zelnik-Manor 2011)(Wang, Wang, and Feng 2006)(Edelman, Arias, and Smith 1999). The distance definition in (Wang, Wang, and Feng 2006) is not shown to be metric, and all of these previous similarities or distances are not shown to admit a possible way to generate similarity-preserving binary signatures.

The works most related to ours are the approximate nearest subspace search methods (Basri, Hassner, and Zelnik-Manor 2011)(Wang et al. 2013). These methods provide ways of conducting approximate nearest neighbor search in terms of specific distances in a database of subspaces. However, their methods are designed specifically for approximate nearest neighbor search, and they do not provide a compact representation, or more specificly, binary representation for subspaces.

The proposed binary signature for subspaces in this work can be used in not only approximate nearest neighbor search, but also applications where only the pairwise similarity or distance is exploited, e.g. clustering, kernel methods. In other words, generating binary signatures for subspaces may be regarded as a general preprocessing step.

Pairwise Similarity Measurement for Linear Subspaces

Principal Angles between two Linear Subspaces

Principal angles between subspaces serve as fundamental tools in mathematics, statistics and related applications, e.g. data mining. The concept was first introduced by Jordan (Jordan 1875) in 1875. Principal angles provide information about the relative position of two subspaces. In this work, they are the building blocks of our definition of the angle, angular similarity and angular distance between subspaces.

Formally, for two subspaces $\mathcal{P}$ and $\mathcal{Q}$ of $\mathcal{R}^d$, denote the dimension $dim(\mathcal{P})$ of $\mathcal{P}$ by $d_{\mathcal{P}}$ and the dimension $dim(\mathcal{Q})$ of $\mathcal{Q}$ by $d_{\mathcal{Q}}$. Assume that $d_{max} = d_{\mathcal{P}} \geq d_{\mathcal{Q}} = d_{min}$. Then the principal angles between them, $\theta_1, \theta_2, ..., \theta_{d_{min}} \in [0, \pi/2]$, are defined recursively as (Golub and Van Loan 1996; Knyazev, Merico, and Argentati 2002):

$$\cos(\theta_i) = \max_{u_i \in \mathcal{P}} \max_{v_i \in \mathcal{Q}} \frac{u_i^T v_i}{\|u_i\| \|v_i\|}$$

subject to

$$u_i^T u_j = 0 \text{ and } v_i^T v_j = 0, \text{ for } j = 1, 2, ..., i-1$$

Principal angles between two subspaces can be computed via singular value decomposition (Golub and Van Loan 1996; Knyazev, Merico, and Argentati 2002). Assume that P is a $d \times d_{\mathcal{P}}$ matrix with orthonormal columns which form an orthonormal basis of the subspace $\mathcal{P}$, and Q is a $d \times d_{\mathcal{Q}}$ matrix with orthonormal columns which form an orthonormal basis of the subspace $\mathcal{Q}$. Then the cosine of each of the principal angles equals a singular value of $P^T Q$. Formally, assume that the reduced singular value decomposition (SVD) of $P^T Q$ is

$$Y^T P^T Q Z = diag(\sigma_1, \sigma_2, ..., \sigma_{d_{min}}), \tag{1}$$

where $1 \geq \sigma_1 \geq \sigma_2 \geq ... \geq \sigma_{d_{min}} \geq 0$ are the singular values of $P^T Q$ and $Y \in \mathcal{R}^{d_{\mathcal{P}} \times d_{\mathcal{Q}}}$, $Z \in \mathcal{R}^{d_{\mathcal{Q}} \times d_{\mathcal{Q}}}$ have orthonormal columns.

Then

$$\theta_i = \arccos(\sigma_i), i = 1, 2, ..., d_{min} \tag{2}$$

Angle, Angular Similarity and Angular Distance between Subspaces

As mentioned before, principal angles depict the relative position of two subspaces. In this section we formally define the angle, pairwise similarity measurement and distance for linear subspaces, based on principal angles.

Definition 1. *For two subspaces $\mathcal{P}$ and $\mathcal{Q}$ of $\mathcal{R}^d$, denote the dimension $dim(\mathcal{P})$ of $\mathcal{P}$ by $d_{\mathcal{P}}$ and the dimension $dim(\mathcal{Q})$ of $\mathcal{Q}$ by $d_{\mathcal{Q}}$. Assume that $d_{max} = d_{\mathcal{P}} \geq d_{\mathcal{Q}} = d_{min}$, and $\theta_1, \theta_2, ..., \theta_{d_{min}}$ are the principal angles between $\mathcal{P}$ and $\mathcal{Q}$. Then the angle between $\mathcal{P}$ and $\mathcal{Q}$ is defined as*

$$\theta_{\mathcal{P},\mathcal{Q}} = \arccos \frac{\sum_{i=1}^{d_{min}} \cos^2 \theta_i}{\sqrt{d_{max}}\sqrt{d_{min}}}. \tag{3}$$

Their angular similarity is

$$\begin{aligned} sim(\mathcal{P}, \mathcal{Q}) &= 1 - \frac{\theta_{\mathcal{P},\mathcal{Q}}}{\pi} \\ &= 1 - \frac{1}{\pi} \arccos \frac{\sum_{i=1}^{d_{min}} \cos^2 \theta_i}{\sqrt{d_{max}}\sqrt{d_{min}}}. \end{aligned} \tag{4}$$

And their angular distance is

$$\begin{aligned} d(\mathcal{P}, \mathcal{Q}) &= 1 - sim(\mathcal{P}, \mathcal{Q}) \\ &= \frac{1}{\pi} \arccos \frac{\sum_{i=1}^{d_{min}} \cos^2 \theta_i}{\sqrt{d_{max}}\sqrt{d_{min}}}. \end{aligned} \tag{5}$$

Since principal angles are invariant to the choice of orthonormal basis, this definition of angle, angular similarity and angular distance is also **invariant to the choice of orthonormal basis**. We may regard this angle definition as a generalization of the angle between two vectors. Indeed, when $d_{\mathcal{P}} = d_{\mathcal{Q}} = 1$, this definition degenerates to the angle between two vectors. It is easy to see that the angular distance function $d(\cdot, \cdot)$ satisfies the following three properties: non-negativity, identity of indiscernibles and symmetry. Formally, for any subspaces $\mathcal{O}$, $\mathcal{P}$,$\mathcal{Q}$ of $\mathcal{R}^d$:

Non-negativity: $d(\mathcal{P}, \mathcal{Q}) \geq 0$

Identity of Indiscernibles: $d(\mathcal{P}, \mathcal{Q}) = 0$ if and only if $\mathcal{P} = \mathcal{Q}$

Symmetry: $d(\mathcal{P}, \mathcal{Q}) = d(\mathcal{Q}, \mathcal{P})$

Besides, later we will show that $d(\cdot, \cdot)$ also satisfies triangle inequality as described below, and thus it is a valid metric.

Triangle Inequality: $d(\mathcal{O}, \mathcal{Q}) \leq d(\mathcal{O}, \mathcal{P}) + d(\mathcal{P}, \mathcal{Q})$

Note that to compute the angle $\theta_{\mathcal{P},\mathcal{Q}}$ between two subspaces $\mathcal{P}$ and $\mathcal{Q}$, it is not necessary to perform the SVD of $P^T Q$. Since $\|P^T Q\|_F^2 = \sum_{i=1}^{d_{min}} \cos^2 \theta_i$, we may first compute $\|P^T Q\|_F^2$, then scale it by $\frac{1}{\sqrt{d_{max}}\sqrt{d_{min}}}$, and take the arccos of the value.

Besides these nice properties, in the subsequent section we will show that it is possible to generate binary signatures for subspaces, and the normalized Hamming distance between the binary signatures is an unbiased estimator of the angular distance. The experiment results on face recognition show that the angular distance as a distance measurement is very discriminative.

Similarity-Preserving Binary Signature for Linear Subspaces

In this section we develop a way of producing binary signatures for linear subspaces, of which the Hamming distance preserves the similarity. We will show that the normalized Hamming distance between the binary signatures is an unbiased estimator of the pairwise angular distance (Definition 1) between the corresponding subspaces. We also provide a lower bound for the length of the binary signatures to guarantee uniform distance-preservation within a set of subspaces. In addition, we also show that the angular distance as defined satisfies triangle inequality.

Assume that P is a $d \times d_{\mathcal{P}}$ matrix with orthonormal columns which form an orthonormal basis of the subspace $\mathcal{P}$, and Q is a $d \times d_{\mathcal{Q}}$ matrix with orthonormal columns which form an orthonormal basis of the subspace $\mathcal{Q}$. Then we may represent subspaces $\mathcal{P}$ and $\mathcal{Q}$ by their orthographic projection matrices PP^T and QQ^T respectively. To generate binary signatures for $\mathcal{P}$, we use an operator $g(\cdot)$ to transform PP^T into a vector, as introduced in (Basri, Hassner, and Zelnik-Manor 2011). For a $d \times d$ symmetric matrix Z, the operator rearranges the elements of the upper triangular part of Z together with the diagonal elements scaled by $\frac{1}{\sqrt{2}}$, into a vector. Formally, it is defined as:

Definition 2. *(Basri, Hassner, and Zelnik-Manor 2011)* $g(Z) = [\frac{z_{1,1}}{\sqrt{2}}, z_{1,2}, ..., z_{1,d}, \frac{z_{2,2}}{\sqrt{2}}, z_{2,3}, ..., \frac{z_{d,d}}{\sqrt{2}}]^T \in \mathcal{R}^{d'}$ *where* $d' = d(d+1)/2$*, for any* $d \times d$ *symmetric matrix* Z.

We will show that the angle between the two vectors $g(PP^T)$ and $g(QQ^T)$ equals the angle $\theta_{\mathcal{P},\mathcal{Q}}$ between $\mathcal{P}$ and $\mathcal{Q}$. Formally, we have

Lemma 1. *Assume that* P *is a* $d \times d_{\mathcal{P}}$ *matrix with orthonormal columns which form an orthonormal basis of the subspace* $\mathcal{P} \subseteq \mathcal{R}^d$*, and* Q *is a* $d \times d_{\mathcal{Q}}$ *matrix with orthonormal columns which form an orthonormal basis of the subspace* $\mathcal{Q} \subseteq \mathcal{R}^d$*. Then* $\theta_{g(PP^T),g(QQ^T)} = \theta_{\mathcal{P},\mathcal{Q}}$.

Proof. Assume that $d_{max} = d_{\mathcal{P}} \geq d_{\mathcal{Q}} = d_{min}$ and $\theta_1, \theta_2, ..., \theta_{d_{min}}$ are the principal angles between $\mathcal{P}$, $\mathcal{Q}$. It is shown (Basri, Hassner, and Zelnik-Manor 2011) that $\|PP^T - QQ^T\|_F^2 = d_{\mathcal{P}} + d_{\mathcal{Q}} - 2\sum_{i=1}^{d_{min}} \cos^2\theta_i = 2\|g(PP^T) - g(QQ^T)\|_2^2$.

Since $2\|g(PP^T)\|_2^2 = d_{\mathcal{P}}$, $2\|g(QQ^T)\|_2^2 = d_{\mathcal{Q}}$,

$$\begin{aligned} & 2\|g(PP^T) - g(QQ^T)\|_2^2 \\ = \ & d_{\mathcal{P}} + d_{\mathcal{Q}} - 2\sqrt{d_{\mathcal{P}}}\sqrt{d_{\mathcal{Q}}}\cos\theta_{g(PP^T),g(QQ^T)} \end{aligned} \quad (6)$$

Then $\theta_{g(PP^T),g(QQ^T)} = \arccos \frac{\sum_{i=1}^{d_{min}} \cos^2\theta_i}{\sqrt{d_{max}}\sqrt{d_{min}}} = \theta_{\mathcal{P},\mathcal{Q}}$ □

Therefore the problem of generating similarity-preserving binary signatures for subspaces reduces to that of generating angle-preserving binary signatures for vectors.

Sign-random-projection (Charikar 2002) is a probabilistic method for generating binary signatures for vectors. The function is a combination of sign function $sgn(\cdot)$ together with a random projection. Formally, a sign-random-projection function $h_v(\cdot) : \mathcal{R}^d \rightarrow \{0, 1\}$ is defined as

Definition 3. *(Charikar 2002)*

$$h_v(a) = sgn(v^T a)$$

for any vector $a \in \mathcal{R}^d$, where v is a $d \times 1$ random vector, each entry of which is an independent standard normal random variable, and $sgn(\cdot)$ is defined as:

$$sgn(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

It is proven that (Goemans and Williamson 1995), for any two given vectors a and b:

$$Pr[h_v(a) = h_v(b)] = 1 - \frac{\theta_{a,b}}{\pi} \quad (7)$$

where $\theta_{a,b} = \arccos(\frac{a^T b}{\|a\|_2\|b\|_2})$, and $\|\cdot\|_2$ denotes the ℓ_2-norm of a vector.

Then the function for generating binary signatures of only 1 bit for subspaces is a combination of the sign-random-projection function $h_v(\cdot)$ together with the operator $g(\cdot)$ as defined in Definition 2, i.e. $h_v(g(\cdot))$ (note that here v is a $d' \times 1$ vector). To be concise, we define

Definition 4. $h_v(\mathcal{P}) = h_v(g(PP^T))$.

Formally, by Lemma 1 and Equation (7), we have

$$\begin{aligned} & Pr[h_v(\mathcal{P}) = h_v(\mathcal{Q})] \\ = \ & Pr[h_v(g(PP^T)) = h_v(g(QQ^T))] \\ = \ & 1 - \frac{\theta_{g(PP^T),g(QQ^T)}}{\pi} \\ = \ & 1 - \frac{\theta_{\mathcal{P},\mathcal{Q}}}{\pi} \\ = \ & sim(\mathcal{P}, \mathcal{Q}) \end{aligned} \quad (8)$$

Thus the function family $h_v(g(\cdot))$ is a locality-sensitive hash family for subspaces, and by the following lemma proven in (Charikar 2002), we have that the angular distance $d(\cdot, \cdot)$ defined in Definition 1 satisfies **triangle inequality**.

Lemma 2. *(Charikar 2002) For any similarity function* $sim(\cdot, \cdot)$ *that admits a locality-sensitive hash family* $h \in \mathcal{F}$ *operating on domain* $\mathcal{D}$ *that satisfies:*

$$Pr[h(a) = h(b)] = sim(a, b), \quad (9)$$

for any a*,* $b \in \mathcal{D}$*. Then* $1 - sim(\cdot, \cdot)$ *satisfies triangle inequality.*

To produce binary signatures of length K for subspaces, we need to generate K standard normal random vectors $v_1, v_2, ... v_K$ and form a $K \times d'$ random projection matrix A with each v_i^T as one row, and thus each entry of A is an independently sampled standard normal variable. Therefore the final function is

Definition 5.

$$\begin{aligned} h_A(\mathcal{P}) &= sgn(A\, g(PP^T)) \\ &= [h_{v_1}(\mathcal{P}), h_{v_2}(\mathcal{P}), ..., h_{v_K}(\mathcal{P})]^T \end{aligned} \quad (10)$$

where $A = [v_1, v_2, ...v_K]^T$*, and the sign function* $sgn(\cdot)$ *is applied element-wisely to its input.*

It can be shown that the normalized Hamming distance between two binary signatures $h_A(\mathcal{P})$ and $h_A(\mathcal{Q})$ is an unbiased estimator of the angular distance between the two subspaces $\mathcal{P}$ and $\mathcal{Q}$. Formally, by Equation (8) and the linearity of expectation, it can be shown that

$$\mathbb{E}[d_H(h_A(\mathcal{P}), h_A(\mathcal{Q}))/K] = \frac{\theta_{\mathcal{P},\mathcal{Q}}}{\pi} = d(\mathcal{P}, \mathcal{Q}) \quad (11)$$

where $d_H(\cdot,\cdot)$ outputs the Hamming distance between two binary vectors. Therefore the binary signature for subspaces is similarity-preserving.

The variance of this estimator is

$$Var[d_H(h_A(\mathcal{P}), h_A(\mathcal{Q}))/K] = \frac{\theta_{\mathcal{P},\mathcal{Q}}}{K\pi}(1 - \frac{\theta_{\mathcal{P},\mathcal{Q}}}{\pi}) \quad (12)$$

This shows that the longer the length of the binary signatures, the smaller the variance of the estimator.

The following lemma gives a lower bound of K to ensure that with constant probability, the binary signatures preserve the angular distance between every pair of the subspaces in a fixed set of n subspaces within a constant small error ϵ.

Lemma 3. *For any* $0 < \epsilon < 1, 0 < \delta < \frac{1}{2}$ *and any set* S *of* n *subspaces of* $\mathcal{R}^d$*, let* $K \geq K_0 = \frac{1}{2\epsilon^2} \ln \frac{n(n-1)}{\delta}$ *and* $h_A(\cdot)$ *as defined in Definition 5, and each entry of* A *is an independent standard normal variable. Then with probability at least* $1 - \delta$*, for all* $\mathcal{P}, \mathcal{Q} \in S$*:*

$$|d_H(h_A(\mathcal{P}), h_A(\mathcal{Q}))/K - d(\mathcal{P}, \mathcal{Q})| \leq \epsilon \quad (13)$$

Proof. Denote $A = [v_1, v_2, ...v_K]^T$. For a fixed pair of subspaces $\mathcal{P}_0, \mathcal{Q}_0 \in S$. Define random variables

$$X_i = \begin{cases} 1, & h_{v_i}(\mathcal{P}_0) \neq h_{v_i}(\mathcal{Q}_0) \\ 0, & h_{v_i}(\mathcal{P}_0) = h_{v_i}(\mathcal{Q}_0) \end{cases}$$

for $i = 1, 2, ..., K$. Then $d_H(h_A(\mathcal{P}_0), h_A(\mathcal{Q}_0)) = \sum_{i=1}^{K} X_i$. By Hoeffding's inequality,

$$\begin{aligned} & Pr[|d_H(h_A(\mathcal{P}_0), h_A(\mathcal{Q}_0))/K - d(\mathcal{P}_0, \mathcal{Q}_0)| \geq \epsilon] \\ = & Pr[|\frac{1}{K}\sum_{i=1}^{K} X_i - \mathbb{E}(\frac{1}{K}\sum_{i=1}^{K} X_i)| \geq \epsilon] \\ \leq & 2\exp(-2\epsilon^2 K) \end{aligned} \quad (14)$$

Denote event $E = \{|d_H(h_A(\mathcal{P}), h_A(\mathcal{Q}))/K - d(\mathcal{P}, \mathcal{Q})| \leq \epsilon, for\ all\ \mathcal{P}, \mathcal{Q} \in S\}$. Then by union bound,

$$Pr[E] \geq 1 - \binom{n}{2} 2\exp(-2\epsilon^2 K) \geq 1 - \delta \quad (15)$$

□

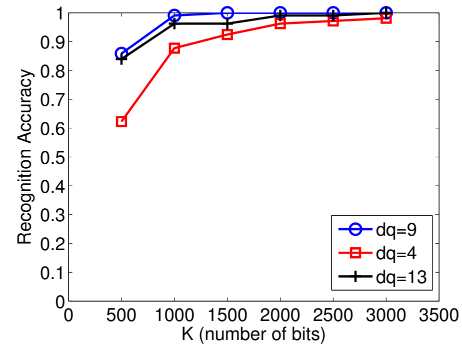

Figure 1: Face recognition accuracy achieved by using the binary signatures of various lengths K (number of bits) on Extended Yale database B + PIE database, with query subspaces of dimensions dq = 4, 9, 13. The exact nearest subspace search in terms of angular distance achieves 100% accuracy for dq = 4, 9, 13.

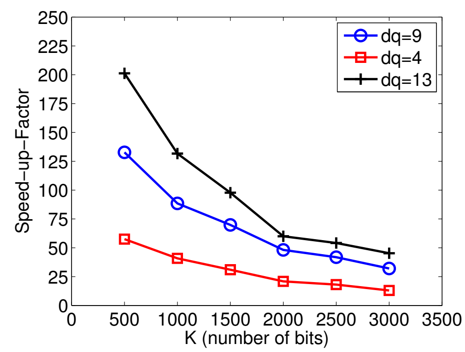

Figure 2: Speed-up-factor achieved by using the binary signatures of various lengths K (number of bits) on Extended Yale database B + PIE database, with query subspaces of dimensions dq = 4, 9, 13.

Remarks: Lemma 3 can generalize to any locality-sensitive hashing method that satisfies Equation (9). It shows that to guarantee the uniform distance-preservation within a set, the signature length K has nothing to do with the original ambient dimension of the input space, and it increases very slowly as the size of the set grows. This is similar to the Johnson-Lindenstrauss Lemma (Johnson and Lindenstrauss 1984; Dasgupta and Gupta 2003), a well-known theoretical result on random projection.

Experiment

In this section we demonstrate the effectiveness of the binary signature by applying it to face recognition.

Datasets: we use the union of two face datasets, the Extended Yale Face Database B (Georghiades, Belhumeur, and Kriegman 2001)(Lee, Ho, and Kriegman 2005) and the PIE database (Sim, Baker, and Bsat 2002). Note that both of these datasets are the processed versions[1] (He et al.

[1] http://www.cad.zju.edu.cn/home/dengcai/Data/FaceData.html

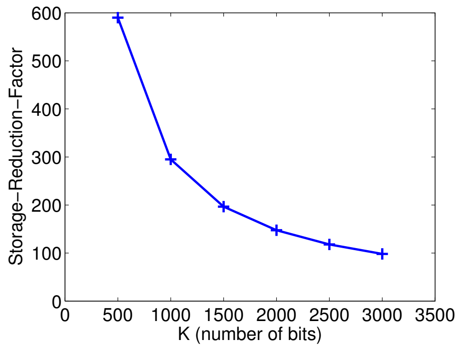

Figure 3: Storage-reduction-factor achieved by using the binary signatures of various lengths K (number of bits) on Extended Yale database B + PIE database.

2005)(Cai et al. 2006)(Cai et al. 2007)(Cai, He, and Han 2007). Each image in both datasets has been cropped and resized to 32×32. Extended Yale Face Database B contains 38 individuals each with about 64 near-frontal images under different illuminations. PIE database contains 68 individuals each with 170 images under 5 near frontal poses (C05, C07, C09, C27, C29), 43 different illumination conditions and 4 different expressions. We only use the frontal face images (C27) from PIE database. Therefore the union of these two datasets contains 106 individuals in total. The two datasets have been split into training sets and testing sets respectively. The training set of the Extended Yale Face Database B contains around 50 images per individual, and the rest (around 14) images are in the testing set. The training set of the PIE database has about 20 images per individual, and the rest (around 14) images are in the testing set.

Experiment Setup: each face image is vectorized to a 1024×1 vector. Several vectors of the same individual constitute a subspace. For each of these subspaces, we fit a subspace of dimension 9 by taking the first 9 principal components, and use this subspace to represent each individual in the training set. The same goes for the testing set, except that we test query subspaces with dimensions $dq = 4, 9, 13$ respectively. Thus there are 106 subspaces in the training set and testing set respectively.

The baseline is the exact nearest subspace search method in terms of angular distance. For each individual in the testing set (querying set), we search for the nearest subspace in the training set (database) in terms of angular distance as defined in Definition 1, and use its label as the prediction to that of the query. This is done by linear scanning the training set and returning the candidate subspace with the nearest angular distance to the query.

To test the effectiveness of the proposed binary signature, we generate a binary signature for each subspace in the training and the testing set, and thus each subspace is now represented by a binary signature. For each query we search for the nearest binary signature in the training set in terms of Hamming distance. This is also done by linear scanning the database. Since the length of the binary signatures is relatively small and Hamming distance can be computed very efficiently, linear scan in Hamming space is very fast in practice (He, Wen, and Sun 2013). We test different lengths K of the binary signatures ranging from 500 to 3000.

Results: The accuracy of face recognition using the binary signatures with different lengths (number of bits) are shown in Figure 1. Note that the exact nearest subspace search in terms of angular distance achieves 100% accuracy for dq = 4, 9, 13. Under different query subspace dimensions, as the length K of the binary signatures grows, the recognition accuracy gradually increases and converges to 1. This is due to the asymptotic convergence guarantee of the estimators produced by the binary signatures. The accuracy of $dq = 9$ is the highest for various lengths K, which is slight higher than that of $dq = 13$ for some values of K. Figure 2 shows the factor of speed-up over the exact method achieved by using the binary signatures, against different lengths K, for dq = 4, 9, 13. As the query subspace dimension grows, the speed-up-factor increases accordingly. This is because when fixing the dimension of the subspaces in the database, the computational cost of computing the angular distance grows approximately linearly with the query subspace dimension: the computational cost of computing the angular distance between two subspaces $\mathcal{P}$ and $\mathcal{Q}$ is $O(d \times d_{\mathcal{P}} \times d_{\mathcal{Q}})$.

When $dq = 9$ and $K = 1500$, the approximate method using the binary signatures achieves more than 60 times faster recognition speed than exact method, with 100% recognition accuracy. If each entry of a vector is stored with 32 bits, then a subspace $\mathcal{P}$ of $\mathcal{R}^d$ with dimension $d_{\mathcal{P}}$ requires $32 \times d \times d_{\mathcal{P}}$ to store (the $d \times d_{\mathcal{P}}$ matrix P of which the columns form an orthonormal basis of $\mathcal{P}$). In the experiment, $d = 1024$, $d_{\mathcal{P}} = 9$, then the storage requirement for each subspace in the database is 294912 bits. Therefore using the binary signatures of 1500 bits for each subspace reduces the storage by a factor of 196. Figure 3 shows the storage-reduction-factor achieved by using the binary signatures of various lengths K.

Conclusion

In this paper we formally define the angular similarity and angular distance between subspaces, and we show that the angular distance is a metric which satisfies non-negativity, identity of indiscernibles, symmetry and triangle inequality. Then we propose a method to produce compact binary signatures for subspaces. The normalized Hamming distance between the binary signatures is an unbiased estimator of the pairwise angular distance. We provide a lower bound on the length of the signatures which guarantees uniform angular-distance-preservation within a set of subspaces. The experiments on face recognition show that using the binary signature as representation achieves more than 60 times speed-up, and a reduction of storage space by a factor of more than 190, with no loss in recognition accuracy.

Acknowledgments

This work was supported by the National Basic Research Program (973 Program) of China (Grant Nos. 2013CB329403 and 2012CB316301), the National Natural Science Foundation of China (Grant No. 61332007, 61273023 and 91120011), the Beijing Natural Science Foundation (Grant No. 4132046), and the Tsinghua University Initiative Scientific Research Program (Grant No. 20121088071). This research is partially supported by the Singapore National Research Foundation under its International Research Centre @Singapore Funding Initiative and administered by the IDM Programme Office. This work was supported in part to Dr. Qi Tian by ARO grant W911NF-12-1-0057, Faculty Research Award by NEC Laboratories of America, and 2012 UTSA START-R Research Award respectively. This work was supported in part by National Science Foundation of China (NSFC) 61128007.

References

Andoni, A., and Indyk, P. 2006. Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions. In *Annual IEEE Symposium on Foundations of Computer Science*, 459–468.

Basri, R., and Jacobs, D. W. 2003. Lambertian reflectance and linear subspaces. *IEEE Trans. Pattern Anal. Mach. Intell.* 25(2):218–233.

Basri, R.; Hassner, T.; and Zelnik-Manor, L. 2011. Approximate nearest subspace search. *IEEE Trans. Pattern Anal. Mach. Intell.* 33(2):266–278.

Cai, D.; He, X.; Han, J.; and Zhang, H.-J. 2006. Orthogonal laplacianfaces for face recognition. *IEEE Transactions on Image Processing* 15(11):3608–3614.

Cai, D.; He, X.; Hu, Y.; Han, J.; and Huang, T. 2007. Learning a spatially smooth subspace for face recognition. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition Machine Learning (CVPR'07)*.

Cai, D.; He, X.; and Han, J. 2007. Spectral regression for efficient regularized subspace learning. In *Proc. Int. Conf. Computer Vision (ICCV'07)*.

Charikar, M. 2002. Similarity estimation techniques from rounding algorithms. In *ACM Symposium on Theory of Computing*, 380–388.

Dasgupta, S., and Gupta, A. 2003. An elementary proof of a theorem of johnson and lindenstrauss. *Random Struct. Algorithms* 22(1):60–65.

Edelman, A.; Arias, T. A.; and Smith, S. T. 1999. The geometry of algorithms with orthogonality constraints. *SIAM J. Matrix Anal. Appl.* 20(2):303–353.

Georghiades, A.; Belhumeur, P.; and Kriegman, D. 2001. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. Pattern Anal. Mach. Intelligence* 23(6):643–660.

Goemans, M. X., and Williamson, D. P. 1995. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM* 42(6):1115–1145.

Golub, G. H., and Van Loan, C. F. 1996. *Matrix Computations (3rd Ed.)*. Johns Hopkins University Press.

Gong, Y., and Lazebnik, S. 2011. Iterative quantization: A procrustean approach to learning binary codes. In *IEEE Conference on Computer Vision and Pattern Recognition*, 817–824.

He, X.; Yan, S.; Hu, Y.; Niyogi, P.; and Zhang, H.-J. 2005. Face recognition using laplacianfaces. *IEEE Trans. Pattern Anal. Mach. Intelligence* 27(3):328–340.

He, K.; Wen, F.; and Sun, J. 2013. K-means hashing: An affinity-preserving quantization method for learning binary compact codes. In *CVPR*, 2938–2945.

Ji, J.; Li, J.; Yan, S.; Zhang, B.; and Tian, Q. 2012. Super-bit locality-sensitive hashing. In *Advances in Neural Information Processing Systems 25*. 108–116.

Johnson, W. B., and Lindenstrauss, J. 1984. Extensions of Lipschitz mapping into Hilbert space. In *Conf. in modern analysis and probability*, volume 26 of *Contemporary Mathematics*, 189–206. American Mathematical Society.

Jordan, C. 1875. Essai sur la géométrie à n dimensions. *Bulletin de la Société Mathématique de France* 3:103–174.

Knyazev, A. V.; Merico; and Argentati, E. 2002. Principal angles between subspaces in an a-based scalar product: Algorithms and perturbation estimates. *SIAM J. Sci. Comput* 23:2009–2041.

Kulis, B., and Darrell, T. 2009. Learning to hash with binary reconstructive embeddings. In *Advances in Neural Information Processing Systems*, 1042–1050.

Lee, K.; Ho, J.; and Kriegman, D. 2005. Acquiring linear subspaces for face recognition under variable lighting. *IEEE Trans. Pattern Anal. Mach. Intelligence* 27(5):684–698.

Liu, G., and Yan, S. 2011. Latent low-rank representation for subspace segmentation and feature extraction. In *ICCV*, 1615–1622.

Liu, G.; Lin, Z.; Yan, S.; Sun, J.; Yu, Y.; and Ma, Y. 2013. Robust recovery of subspace structures by low-rank representation. *IEEE Trans. Pattern Anal. Mach. Intell.* 35(1):171–184.

Sim, T.; Baker, S.; and Bsat, M. 2002. The cmu pose, illumination, and expression (pie) database. In *FGR*, 53–58.

Torralba, A.; Fergus, R.; and Weiss, Y. 2008. Small codes and large image databases for recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, 1–8.

Wang, X.; Atev, S.; Wright, J.; and Lerman, G. 2013. Fast subspace search via grassmannian based hashing. In *The IEEE International Conference on Computer Vision (ICCV)*.

Wang, J.; Kumar, S.; and Chang, S.-F. 2012. Semi-supervised hashing for large-scale search. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34(12):2393–2406.

Wang, L.; Wang, X.; and Feng, J. 2006. Subspace distance analysis with application to adaptive bayesian algorithm for face recognition. *Pattern Recognition* 39(3):456–464.

Weiss, Y.; Torralba, A.; and Fergus, R. 2008. Spectral hashing. In *Advances in Neural Information Processing Systems*, 1753–1760.