# Principal Component Analysis of Image Gradient Orientations for Face Recognition

Georgios Tzimiropoulos[†], Stefanos Zafeiriou[†] and Maja Pantic[†,*]
† Department of Computing, Imperial College London
180 Queen's Gate, London SW7 2AZ, UK.
∗ EEMCS, University of Twente, 5 Drienerlolaan,
7522 NB Enschede, The Netherlands.
{gt204, s.zafeiriou,m.pantic}@imperial.ac.uk

*Abstract*— We introduce the notion of Principal Component Analysis (PCA) of image gradient orientations. As image data is typically noisy, but noise is substantially different from Gaussian, traditional PCA of pixel intensities very often fails to estimate reliably the low-dimensional subspace of a given data population. We show that replacing intensities with gradient orientations and the $\ell_2$ norm with a cosine-based distance measure offers, to some extend, a remedy to this problem. Our scheme requires the eigen-decomposition of a covariance matrix and is as computationally efficient as standard $\ell_2$ intensity-based PCA. We demonstrate some of its favorable properties for the application of face recognition.

## NOTATION

| | |
|---|---|
| $\mathcal{S}$, $\{.\}$ | set |
| $\Re$ | set of reals |
| $\mathcal{C}$ | set of complex numbers |
| $x$ | scalar or complex |
| $j$ | $j^2 = -1$ |
| $e^{j\theta}$ | Euler form: $\cos\theta + j\sin\theta$ |
| $\mathbf{x}$ | column vector |
| $\mathbf{X}$ | matrix |
| $\mathbf{I}_{m \times m}$ | $m \times m$ identity matrix |
| $\mathbf{x}(k)$ | $k$-th element of vector $\mathbf{x}$ |
| $N(\mathcal{X})$ | cardinality of set $\mathcal{X}$ |
| $\|.\|$ | $\ell_2$ norm |
| $\|.\|_F$ | Frobenius norm |
| $\mathbf{X}^H$ | conjugate transpose of $\mathbf{X}$ |
| $\text{Re}[x]$, $\text{Im}[x]$ | real and imaginary part of x |
| $U[a,b]$ | uniform distribution in $[a,b]$ |
| $\mathbb{E}[.]$ | mean value operator |
| $x \sim U[a,b]$ | $x$ follows $U[a,b]$ |

## I. INTRODUCTION

Provision for mechanisms capable of handling gross errors caused by possible arbitrarily large model deviations is a typical prerequisite in computer vision. Such deviations are not unusual in real-world applications where data contains artifacts due to occlusions, illumination changes, shadows, reflections or the appearance of new parts/objects. In most cases, such phenomena cannot be described by a mathematically well-defined generative model and are usually referred as outliers in learning and parameter estimation.

In this paper, we propose a new avenue for Principal Component Analysis (PCA), perhaps the most classical tool for dimensionality reduction and feature extraction in pattern recognition. Standard PCA estimates the $k-$rank linear subspace of the given data population, which is optimal in a least-squares sense. Unfortunately $\ell_2$ norm of pixel intensities enjoys optimality properties only when image noise is i.i.d. Gaussian; for data corrupted by outliers, the estimated subspace can be arbitrarily biased.

Robust formulations to PCA, such as robust covariance matrix estimators [1], [2], are computationally prohibitive for high dimensional data such as images. Robust approaches, well-suited for computer vision applications, include $\ell_1$ [3], [4], robust energy function [5] and weighted combination of nuclear norm and $\ell_1$ minimization [6], [7]. $\ell_1$-based approaches can be computationally efficient, however the gain in robustness is not always significant. The M-Estimation framework of [5] is robust but suitable only for relatively low dimensional data or off-line processing. Under weak assumptions [7], the convex optimization formulation of [6], [7] perfectly recovers the low dimensional subspace of a data population corrupted by sparse arbitrarily large errors; nevertheless efficient reformulations of standard PCA can be orders of magnitude faster.

In this paper we look at robust PCA from a completely different perspective. Our scheme *does not* operate on pixel intensities. In particular, we replace pixel intensities with gradient orientations. We define a notion of pixel-wise image dissimilarity by looking at the distribution of gradient orientation differences; intuitively this must be approximately uniform in $[0, 2\pi)$. We then assume that local orientation mismatches caused by outliers can be also well-described by a uniform distribution which, under some mild assumptions, is canceled out when we apply the cosine kernel. This last observation has been noticed in recently proposed schemes for image registration [8]. Following this line of research, we show that a cosine-based distance measure has a functional form which enables us to define an explicit mapping from the space of gradient orientations into a high-

dimensional sphere where essentially linear complex PCA is performed. The mapping is one-to-one and therefore PCA-based reconstruction in the original input space is direct and requires no further optimization. Similarly to standard PCA, the basic computational module of our scheme requires the eigen-decomposition of a covariance matrix, while high dimensional data can be efficiently analyzed following the strategy suggested in [9].

## II. $\ell_2$-BASED PCA OF PIXEL INTENSITIES

Let us denote by $\mathbf{x}_i \in \Re^p$ the $p-$dimensional vector obtained by writing image $\mathbf{I}_i \in \Re^{m_1 \times m_2}$ in lexicographic ordering. We assume that we are given a population of $n$ samples $\mathbf{X} = [\mathbf{x}_1 | \cdots | \mathbf{x}_n] \in \Re^{p \times n}$. Without loss of generality, we assume zero-mean data. PCA finds a set of $k < n$ orthonormal bases $\mathbf{B}_k = [\mathbf{b}_1 | \cdots | \mathbf{b}_k] \in \Re^{p \times k}$ by minimizing the error function

$$\epsilon(\mathbf{B}_k) = ||\mathbf{X} - \mathbf{B}_k \mathbf{B}_k^T \mathbf{X}||_F^2. \tag{1}$$

The solution is given by the eigenvectors corresponding to the $k$ largest eigenvalues obtained from the eigen-decomposition of the covariance matrix $\mathbf{X}\mathbf{X}^T$. Finally, the reconstruction of $\mathbf{X}$ from the subspace spanned by the columns of $\mathbf{B}_k$ is given by $\tilde{\mathbf{X}} = \mathbf{B}_k \mathbf{C}_k$, where $\mathbf{C}_k = \mathbf{B}_k^T \mathbf{X}$ is the matrix which gathers the set of projection coefficients.

For high dimensional data and Small Sample Size (SSS) problems (i.e. $n \ll p$), an efficient implementation of PCA in $O(n^3)$ (instead of $O(p^3)$) was proposed in [9]. Rather than computing the eigen-analysis of $\mathbf{X}\mathbf{X}^T$, we compute the eigen-analysis of $\mathbf{X}^T\mathbf{X}$ and make use of the following theorem

**Theorem I**
Define matrices $\mathbf{A}$ and $\mathbf{B}$ such that $\mathbf{A} = \mathbf{\Gamma}\mathbf{\Gamma}^H$ and $\mathbf{B} = \mathbf{\Gamma}^H\mathbf{\Gamma}$ with $\mathbf{\Gamma} \in \mathcal{C}^{m \times r}$. Let $\mathbf{U}_A$ and $\mathbf{U}_B$ be the eigenvectors corresponding to the non-zero eigenvalues $\mathbf{\Lambda}_A$ and $\mathbf{\Lambda}_B$ of $\mathbf{A}$ and $\mathbf{B}$, respectively. Then, $\mathbf{\Lambda}_A = \mathbf{\Lambda}_B$ and $\mathbf{U}_A = \mathbf{\Gamma}\mathbf{U}_B\mathbf{\Lambda}_A^{-\frac{1}{2}}$.

## III. RANDOM NUMBER GENERATION FROM GRADIENT ORIENTATION DIFFERENCES

We formalize an observation for the distribution of gradient orientation differences which does not appear to be well-known in the pattern recognition community [1]. Consider a set of images $\{\mathbf{J}_i\}$. At each pixel location, we estimate the image gradients and the corresponding gradient orientation [2]. We denote by $\{\mathbf{\Phi}_i\}$, $\mathbf{\Phi}_i \in [0, 2\pi)^{m_1 \times m_2}$ the set of orientation images and compute the orientation difference image

$$\Delta\mathbf{\Phi}_{il} = \mathbf{\Phi}_i - \mathbf{\Phi}_l. \tag{2}$$

We denote by $\phi_i$ and $\Delta\phi_{il} \triangleq \phi_i - \phi_l$ the $p-$dimensional vectors obtained by writing $\mathbf{\Phi}_i$ and $\Delta\mathbf{\Phi}_{il}$ in lexicographic

---

[1]This observation has been somewhat noticed in [10] without any further comments on its implications.

[2]More specifically, we compute $\Phi_i = \arctan G_{i,y}/G_{i,x}$, where $G_{i,x} = h_x \star I_i$, $G_{i,y} = h_y \star I_i$ and $h_x, h_y$ are filters used to approximate the ideal differentiation operator along the image horizontal and vertical direction respectively. Possible choices for $h_x, h_y$ include central difference estimators of various orders and discrete approximations to the first derivative of the Gaussian.

ordering and $\mathcal{P} = \{1, \ldots, p\}$ the set of indices corresponding to the image support. We introduce the following definition.
**Definition** Images $\mathbf{J}_i$ and $\mathbf{J}_l$ are pixel-wise dissimilar if $\forall k \in \mathcal{P}$, $\Delta\phi_{il}(k) \sim U[0, 2\pi)$.
Not surprisingly, nature is replete with images exemplifying Definition 1. This, in turn, makes it possible to set up a naive image-based random number generator. To confirm this, we used more than $70,000$ pairs of image patches of resolution $200 \times 200$ randomly extracted from natural images [11]. For each pair, we computed $\Delta\phi_{il}$ and formulated the following null hypothesis

- $H_0$: $\forall k \in \mathcal{P}$ $\Delta\phi_{il}(k) \sim U[0, 2\pi)$.

which was tested using the Kolmogorov-Smirnov test [12]. For a significance level equal to 0.01, the null hypothesis was accepted for $94.05\%$ of the image pairs with mean $p$-value equal to $0.2848$. In a similar setting, we tested Matlab's random generator. The null hypothesis was accepted for $99.48\%$ of the cases with mean $p$-value equal to $0.501$. Fig. 1 (a)-(b) show a typical pair of image patches considered in our experiment. Fig. 1 (c) and (d) plot the histograms of the gradient orientation differences and 40,000 samples drawn from Matlab's random number generator respectively.
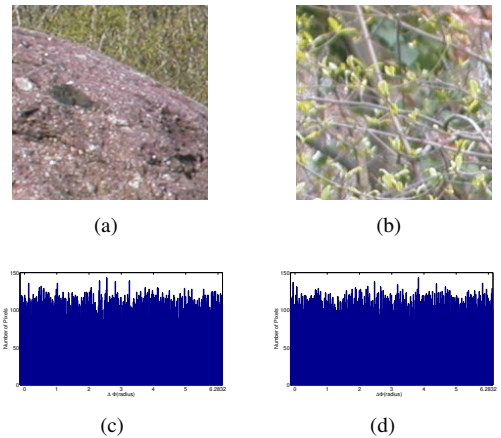


Fig. 1.    (a)-(b) An image pair used in our experiment, (c) Image-based random number generator: histogram of 40,000 gradient orientation differences and (d) Histogram of 40,000 samples drawn from Matlab's random number generator.

## IV. PCA OF GRADIENT ORIENTATIONS

### A. Cosine-based correlation of gradient orientations

Given the set of our images $\{\mathbf{I}_i\}$, we compute the corresponding set of orientation images $\{\mathbf{\Phi}_i\}$ and measure image correlation using the cosine kernel

$$s(\phi_i, \phi_l) \triangleq \sum_{k \in \mathcal{P}} \cos[\Delta\phi_{il}(k)] = cN(\mathcal{P}) \tag{3}$$

where $c \in [-1, 1]$. Notice that for highly spatially correlated images $\Delta\phi_{il}(k) \approx 0$ and $c \to 1$.

Assume that there exists a subset $\mathcal{P}_2 \subset \mathcal{P}$ corresponding to the set of pixels corrupted by outliers. For $\mathcal{P}_1 = \mathcal{P} - \mathcal{P}_2$,

we have

$$s_1(\phi_i, \phi_l) = \sum_{k \in \mathcal{P}_1} \cos[\Delta\phi_{il}(k)] = c_1 N(\mathcal{P}_1) \quad (4)$$

where $c_1 \in [-1, 1]$.

Not unreasonably, we assume that in $\mathcal{P}_2$, the images are pixel-wise dissimilar according to Definition 1. For example, Fig. 2 (a)-(b) show an image pair where $\mathcal{P}_2$ is the part of the face occluded by the scarf and Fig. 2 (c) plots the distribution of $\Delta\phi$ in $\mathcal{P}_2$.
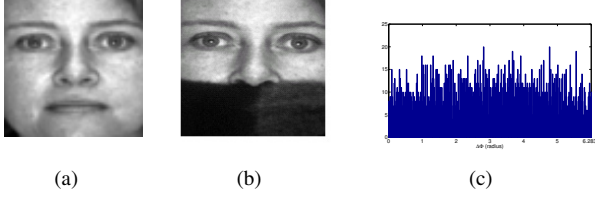


(a)  (b)  (c)

Fig. 2. (a)-(b) An image pair used in our experiments. (c) The distribution of $\Delta\phi$ for the part of face occluded by the scarf.

Before proceeding for $\mathcal{P}_2$, we need the following theorem.

**Theorem II**

Let $u(.)$ be a random process and $u(t) \sim U[0, 2\pi]$ then:

- $\mathbb{E}[\int_{\mathcal{X}} \cos u(t) dt] = 0$ for any non-empty interval $\mathcal{X}$ of $\Re$.
- If $u(.)$ is mean ergodic, then $\int_{\mathcal{X}} \cos u(t) dt = 0$.

**Proof**: Let us define the random process $z(t) = \cos u(t)$. Let also $f_U(u) = U[0, 2\pi]$ and we assumed that $u \sim f_U(u)$. The integral $s = \int_a^b z(t) dt$ of the stochastic process $z(t)$ is a random variable $s$ [12]. By interpreting the above as a Riemannian integral and using the linearity of the expectation operator, we conclude that

$$\begin{aligned} E\{s\} &= \int_a^b E\{z(t)\} dt = \int_a^b \int_{-\infty}^{+\infty} \cos(u) f_U(u) du \\ &= \int_a^b \int_0^{2\pi} \cos(u) du = 0 \end{aligned} \quad (5)$$

which shows that the integral $\int_{\mathcal{X}} \cos u(t) dt$ is equal to zero in mean value. By further assuming mean-ergodicity, the time average is equal to the mean, thus we get

$$E(Z) \propto \int z(t) dt \equiv \int_{\mathcal{X}} \cos u(t) dt = 0 \quad (6)$$

which proves the Theorem.

We also make use of the following approximation

$$\int_{\mathcal{X}} \cos[\Delta\phi_{il}(t)] dt \propto \sum_{k \in \mathcal{P}} \cos[\Delta\phi_{il}(k)] \quad (7)$$

where with some abuse of notation, $\Delta\phi_{il}$ is defined in the continuous domain on the left hand side of (7). Completely analogously, the above theorem and approximation hold for the case of the sine kernel.

Using the above results, for $\mathcal{P}_2$, we have

$$s_2(\phi_i, \phi_l) = \sum_{k \in \mathcal{P}_2} \cos[\Delta\phi_{il}(k)] \simeq 0 \quad (8)$$

It is not difficult to verify that $\ell_2$-based correlation i.e. the inner product between two images will be zero if and only

if the images have interchangeably black and white pixels. Our analysis and (8) show that cosine-based correlation of gradient orientations allows for a much broader class of uncorrelated images. Overall, unlike $\ell_2$-based correlation where the contribution of outliers can be arbitrarily large, $s(.)$ measures correlation as $s(\phi_i, \phi_l) = s_1(\phi_i, \phi_l) + s_2(\phi_i, \phi_l) \simeq c_1 N(\mathcal{P}_1)$, i.e. the effect of outliers is approximately canceled out.

### B. The principal components of image gradient orientations

To show how (3) can be used as a basis for PCA, we first define the distance

$$d^2(\phi_i, \phi_l) = \sum_{k=1}^p \{1 - \cos[\Delta\phi_{il}(k)]\} \quad (9)$$

We can write (9) as follows

$$\begin{aligned} d^2(\phi_i, \phi_l) &= \frac{1}{2} \sum_{k=1}^p \{2 - 2\cos[\phi_i(k) - \phi_l(k)]\} \\ &= \frac{1}{2} ||e^{j\phi_i} - e^{j\phi_l}||^2 \end{aligned} \quad (10)$$

where $e^{j\phi_i} = [e^{j\phi_i(1)}, \ldots, e^{j\phi_i(p)}]^T$. The last equality makes the basic computational module of our scheme apparent. We define the mapping from $[0, 2\pi)^p$ onto a subset of complex sphere with radius $\sqrt{N(\mathcal{P})}$

$$\mathbf{z}_i(\phi_i) = e^{j\phi_i} \quad (11)$$

and apply linear complex PCA to the transformed data $\mathbf{z}_i$.

Using the results of the previous subsection, we can remark the following

**Remark I** If $\mathcal{P} = \mathcal{P}_1 \cup \mathcal{P}_2$ with $\Delta\phi_{il}(k) \sim U[0, 2\pi]$, $\forall k \in \mathcal{P}_2$, then $\text{Re}[\mathbf{z}_i^H \mathbf{z}_l] \simeq c_1 N(\mathcal{P}_1)$

**Remark II** If $\mathcal{P}_2 = \mathcal{P}$, then $\text{Re}[\mathbf{z}_i^H \mathbf{z}_l] \simeq 0$ and $\text{Im}[\mathbf{z}_i^H \mathbf{z}_l] \simeq 0$.

Further geometric intuition about the mapping $\mathbf{z}_i$ is provided by the chord between vectors $\mathbf{z}_i$ and $\mathbf{z}_l$

$$\text{crd}(\mathbf{z}_i, \mathbf{z}_l) = \sqrt{(\mathbf{z}_i - \mathbf{z}_l)^H (\mathbf{z}_i - \mathbf{z}_l)} = \sqrt{2d^2(\phi_i, \phi_l)} \quad (12)$$

Using crd$(.)$, the results of Remark 1 and 2 can be reformulated as $\text{crd}(\mathbf{z}_i, \mathbf{z}_l) \simeq \sqrt{2((1 - c_1)N(\mathcal{P}_1) + N(\mathcal{P}_2))}$ and $\text{crd}(\mathbf{z}_i, \mathbf{z}_l) \simeq \sqrt{2N(\mathcal{P})}$ respectively.

Overall, **Algorithm 1** summarizes the steps of our PCA of gradient orientations.

**Algorithm 1.** *Estimating the principal subspace*

**Inputs:** A set of $n$ orientation images $\mathbf{\Phi}_i$, $i = 1, \ldots, n$ of $p$ pixels and the number $k$ of principal components.

**Step 1.** Obtain $\phi_i$ by writing $\mathbf{\Phi}_i$ in lexicographic ordering.

**Step 2.** Compute $\mathbf{z}_i = e^{j\phi_i}$, form the matrix of the transformed data $\mathbf{Z} = [\mathbf{z}_1 | \cdots | \mathbf{z}_n] \in \mathcal{C}^{p \times n}$ and compute the matrix $\mathbf{T} = \mathbf{Z}^H \mathbf{Z} \in \mathcal{R}^{n \times n}$.

**Step 3.** Compute the eigen-decomposition of $\mathbf{T} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$ and denote by $\mathbf{U}_k \in \mathcal{C}^{p \times k}$ and $\mathbf{\Lambda}_k \in \mathcal{R}^{k \times k}$ the $k$-reduced set. Compute the principal subspace from $\mathbf{B}_k = \mathbf{Z}\mathbf{U}_k \mathbf{\Lambda}_k^{-\frac{1}{2}} \in \mathcal{C}^{p \times k}$.

**Step 4.** Reconstruct using $\tilde{\mathbf{Z}} = \mathbf{B}_k \mathbf{B}_k^H \mathbf{Z}$.

**Step 5.** Go back to the orientation domain using $\tilde{\mathbf{\Phi}} = \angle \tilde{\mathbf{Z}}$.

Let us denote by $\mathcal{Q} = \{1, \ldots, n\}$ the set of image indices and $\mathcal{Q}_i$ any subset of $\mathcal{Q}$. We can conclude the following

**Remark III** If $\mathcal{Q} = \mathcal{Q}_1 \cup \mathcal{Q}_2$ with $\mathbf{z}_i^H \mathbf{z}_l \simeq 0 \ \forall i \in \mathcal{Q}_2$, $\forall l \in \mathcal{Q}$ and $i \neq l$, then, $\exists$ eigenvector $\mathbf{b}_l$ of $\mathbf{B}_n$ such that $\mathbf{b}_l \simeq \frac{1}{N(\mathcal{P})} \mathbf{z}_i$.

A special case of Remark III is the following

**Remark IV** If $\mathcal{Q} = \mathcal{Q}_2$, then $\frac{1}{N(\mathcal{P})} \mathbf{\Lambda} \simeq \mathbf{I}_{n \times n}$ and $\mathbf{B}_n \simeq \frac{1}{N(\mathcal{P})} \mathbf{Z}$.

To exemplify Remark IV, we computed the eigen-spectrum of 100 natural image patches. In a similar setting, we computed the eigen-spectrum of samples drawn from Matlab's random number generator. Fig. 3 plots the two eigen-spectrums.
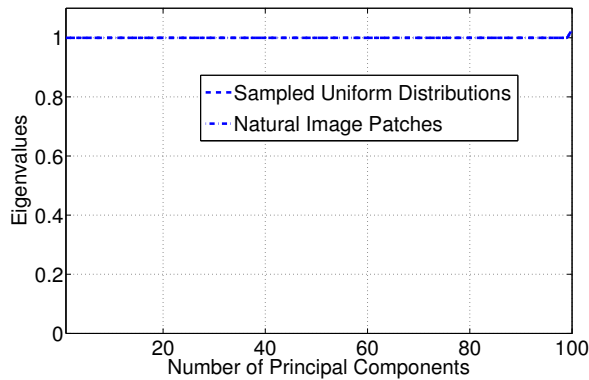


Fig. 3. The eigen-spectrum of natural images and the eigen-spectrum of samples drawn from Matlab's random number generator.

Finally, notice that our framework also enables the direct embedding of new samples. **Algorithm** 2 summarizes the procedure.

**Algorithm 2.** *Embedding of new samples*

**Inputs:** An orientation image $\mathbf{\Theta}$ of $p$ pixels and the principal subspace $\mathbf{B}_k$ of **Algorithm 1**.

**Step 1.** Obtain $\boldsymbol{\theta}$ by writing $\mathbf{\Theta}$ in lexicographic ordering.

**Step 2.** Compute $\mathbf{z} = e^{j\boldsymbol{\theta}}$ and reconstruct using $\tilde{\mathbf{z}} = \mathbf{B}_k \mathbf{B}_k^H \mathbf{z}$.

**Step 3.** Go back to the orientation domain using $\tilde{\boldsymbol{\theta}} = \angle \tilde{\mathbf{z}}$.

## V. RESULTS

### A. Face reconstruction

The estimation of a low-dimensional subspace from a set of a highly-correlated images is a typical application of PCA [13]. As an example, we considered a set of 50 aligned face images of image resolution $192 \times 168$ taken from the Yale B face database [14]. The images capture the face of the same subject under different lighting conditions. This setting usually induces cast shadows as well as other specularities. Face reconstruction from the principal subspace is a natural candidate for removing these artifacts.

We initially considered two versions of this experiment. The first version used the set of original images. In the second version, 20% of the images was artificially occluded by a $70 \times 70$ "Baboon" patch placed at random spatial locations. For both experiments, we reconstructed pixel intensities and gradient orientations with $\ell_2$ intensiy-based PCA and PCA of gradient orientations respectively using the first 5 principal components.

Fig. 4 and Fig. 5 illustrate the quality of reconstruction for 2 examples of face images considered in our experiments. While PCA-based reconstruction of pixel intensities is visually appealing in the first experiment, Fig. 4 (g)-(h) clearly illustrate that, in the second experiment, the reconstruction suffers from artifacts. In contrary, Fig. 5 (e)-(f) and (g)-(h) show that PCA-based reconstruction of gradient orientations not only reduces the effect of specularities but also reconstructs the gradient orientations corresponding to the "face" component only.

This performance improvement becomes more evident by plotting the principal components for each method and experiment. Fig. 6 shows the 5 dominant Eigenfaces of $\ell_2$ intensity-based PCA. Observe that, in the second experiment, the last two Eigenfaces (Fig. 6 (i) and (j)) contain "Baboon" ghosts which largely affect the quality of reconstruction. In contrary, a simple visual inspection of Fig. 7 reveals that, in the second experiment, the principal subspace of gradient orientations (Fig. 7 (f)-(j)) is artifact-free which in turn makes dis-occlusion in the orientation domain feasible.

Finally, to exemplify Remark III, we considered a third version of our experiment where 20% of the images were replaced by *the same* $192 \times 168$ "Baboon" image. Fig. 8 (a)-(e) and (f)-(j) illustrate the principal subspace of pixel intensities and gradient orientations respectively. Clearly, we may observe that $\ell_2$ PCA was unable to handle the extra-class outlier. In contrary, PCA of gradient orientations successfully separated the "face" from the "Baboon" subspace, i.e. no eigenvector was corrupted by the Baboon image (the "Baboon" orientation image appeared as a separate eigenvector). Note that the "face" principal subspace is not the same as the one obtained in versions 1 and 2. This is because only 80% of the images in our dataset was used in this experiment.
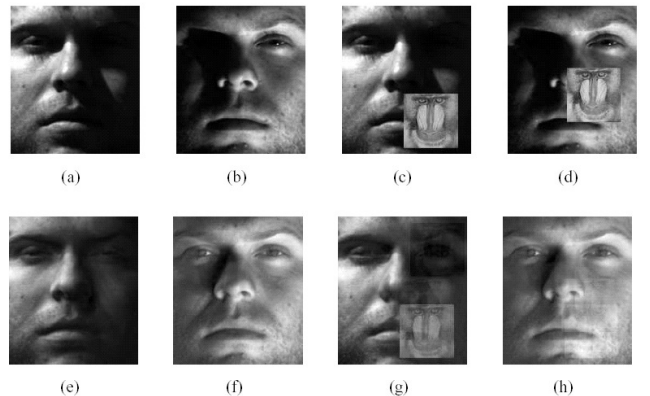


Fig. 4. PCA-based reconstruction of pixel intensities. (a)-(b) Original images used in version 1 of our experiment. (c)-(d) Corrupted images used in version 2 of our experiment. (e)-(f) Reconstruction of (a)-(b) with 5 principal components. (g)-(h) Reconstruction of (c)-(d) with 5 principal components.
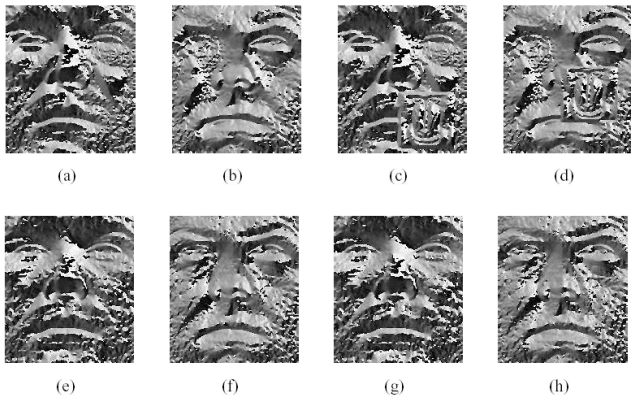
Fig. 5. PCA-based reconstruction of gradient orientations. (a)-(b) Original orientations used in version 1 of our experiment. (c)-(d) Corrupted orientations used in version 2 of our experiment. (e)-(f) Reconstruction of (a)-(b) with 5 principal components. (g)-(h) Reconstruction of (c)-(d) with 5 principal components.
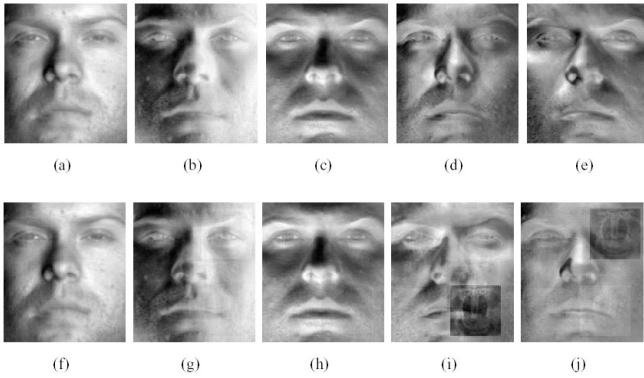


Fig. 6. The 5 principal components of pixel intensities for (a)-(e) version 1 and (f)-(j) version 2 of our experiment.
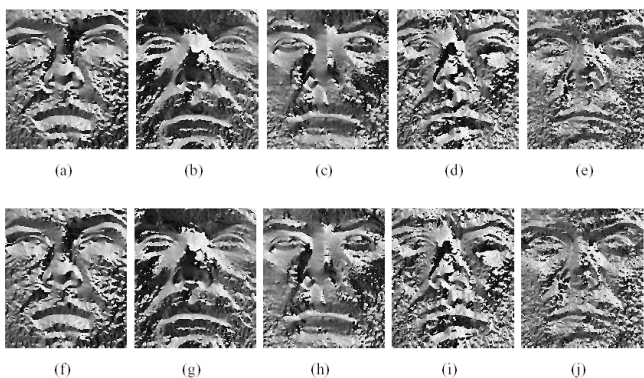


Fig. 7. The 5 principal components of gradient orientations for (a)-(e) version 1 and (f)-(j) version 2 of our experiment.

## B. Face recognition

PCA-based feature extraction for face recognition goes back to the classical work by Turk and Pentland [9] and still remains a standard benchmark for performance evaluation of new algorithms. We considered a single-sample-per-class experiment using aligned frontal-view face images taken
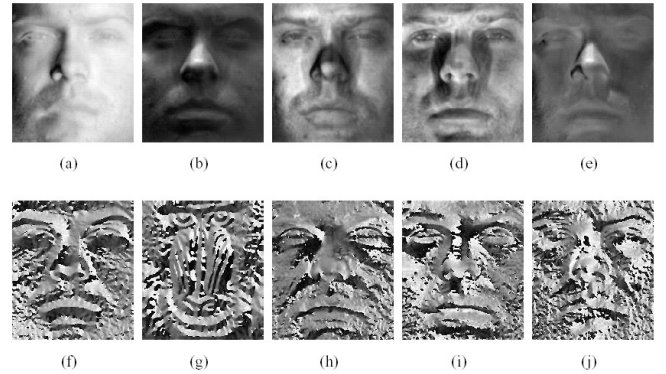


Fig. 8. (a)-(e) The 5 principal components of pixel intensities for version 3 of our experiment and (f)-(j) The 5 principal components of gradient orientations for the same experiment.

from the AR database [15]. The database consists of more than 4,000 frontal view facial images of 126 subjects. Each subject has up to 26 images taken in two sessions. The first session contains 13 images, numbered from 1 to 13, including different facial expressions (1-4), different lighting (5-7) and different occlusions under different lighting (8-13). The second session duplicated the first session two weeks later. We randomly selected a subset with 100 subjects. For training, we used 100 face images of 100 different subjects from session 1. We investigated the robustness of our scheme for the case of illumination variations and occlusions. In particular, we carried out the following experiments:

1) In experiment 1, we used the images 5-7 of session 2 for testing (different illumination).
2) In experiment 2, we used images 8-13 of session 2 for testing (occlusion by scarf or glasses under different illumination).

Note that the second experiment is very challenging since our single-sample-per-class training set does not allows us to find discriminant projection bases by exploiting class-label information, while the presence of the scarf or glasses in the testing set occludes approximately 25-40 % of the total face area.

Table I and Fig. 9 summarize the results for our single sample per class experiments. The robustness of the proposed scheme is evident. As our results show, PCA of gradient orientations achieves almost 100% recognition rate for the case of illumination changes (experiment 1) and approximately 94% recognition rate for the case of occlusions (experiment 2). The latter result is approximately 20% better than the best reported recognition rate [18], which is obtained using as testing set a subset of the occluded images with **no illumination** variations.

Moreover, the presented results should not be compared with those achieved by recent approaches such as the ones in [16] or [17] which use for training **8 images** per subject taken from **both sessions**, while the test images are also taken from **both sessions** (On the contrary, we used only **one sample** from the first session and tested on the second session). For our single-sample-per-class experiment, we

applied the method in [16] using

- features extracted using image resizing, intensity-based PCA, LaplacianFaces as described in [16] for experiment 1.
- the robust extended $\ell_1$ minimization formulation for experiment 2.

The third column of Table I shows the best results achieved by the method in [16]. As we may observe, our PCA of gradient orientations is not only significantly faster but also much more robust.

| Recognition rate | IGO-PCA | I-PCA | Best of [16] |
|---|---|---|---|
| Experiment 1 (illumination)% | 99.67 | 74 | 74 |
| Experiment 2 (occlusions)% | 94 | 28 | 32.33 |

TABLE I

RECOGNITION RESULTS ON AR DATABASE FOR EXPERIMENTS 1 AND 2 (IGO-PCA STANDS FOR PCA OF IMAGE GRADIENT ORIENTATIONS AND I-PCA STANDS FOR INTENSITY-BASED PCA.
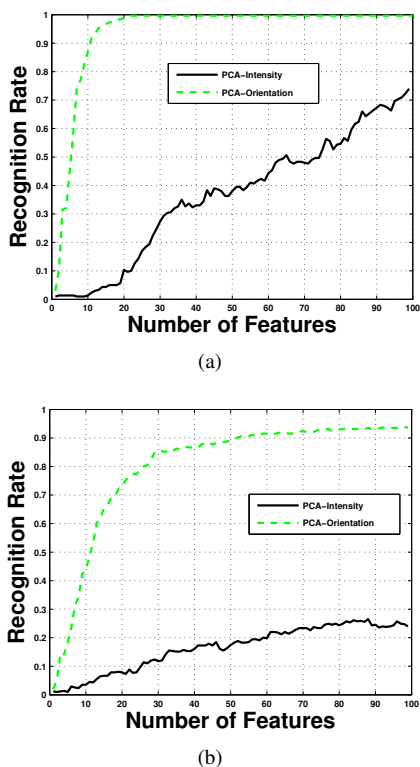


(a)



(b)

Fig. 9. Single sample per class face recognition experiment on the AR database. (a) Experiment 1 (illumination changes) (b) Experiment 2 (occlusions-illumination changes).

## VI. CONCLUSIONS

We introduced a new concept: PCA of gradient orientations. Our framework is as simple as standard $\ell_2$ intensity-based PCA, yet much more powerful for efficient subspace-based data representation. Central to our analysis is the distribution of gradient orientation differences and the cosine kernel which provide us a consistent way to measure image dissimilarity. We showed how this dissimilarity measure can

be naturally used to formulate a robust version of PCA. We demonstrated some of the favorable properties of our framework for the application of face recognition. Extensions of our scheme span a wide range of theoretical topics and applications; from statistical machine learning and clustering to object recognition and tracking.

REFERENCES

[1] N.A. Campbell, *Robust procedures in multivariate analysis I: Robust Covariance estimation*, Applied Statistics, 29 (1980), pp. 231–237.
[2] C. Croux and G. Haesbroeck, *Principal component analysis based on robust estimators of the covariance or correlation matrix: influence functions and efficiencies*, Biometrika, 87 (2000), pp. 603.
[3] Q. Ke and T. Kanade, *Robust L1 norm factorization in the presence of outliers and missing data by alternative convex programming*, in IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR (2005).
[4] N. Kwak, *Principal component analysis based on L1-norm maximization*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 30 (2008), pp. 1672–1680.
[5] F.D.L. Torre and M.J. Black, *A framework for robust subspace learning*, International Journal of Computer Vision, 54 (2003), pp. 117–142.
[6] V. Chandrasekaran, S. Sanghavi, P.A. Parrilo and A.S. Willsky, *Rank-sparsity incoherence for matrix decomposition*, preprint, (2009).
[7] E.J. Candes, X. Li, Y. Ma, and J. Wright, *Robust principal component analysis?*, Arxiv preprint arXiv:0912.3599, (2009).
[8] G. Tzimiropoulos, V. Argyriou, S. Zafeiriou, and T. Stathaki, *Robust FFT-Based Scale-Invariant Image Registration with Image Gradients*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 32 (2010), pp. 1899–1906.
[9] M. Turk and A.P. Pentland, *Eigenfaces for recognition*, Journal of Cognitive Neuroscience, 3 (1991), pp. 71–86.
[10] A.J Fitch, A. Kadyrov, W.J. Christmas, and J. Kittler, *Orientation correlation*, in British Machine Vision Conference, 1 (2002), pp. 133–142.
[11] H.P. Frey, P. Konig, and W. Einhauser, *The Role of First and Second-Order Stimulus Features for Human Overt Attention*, Perception and Psychophysics, 69 (2007), pp. 153–161.
[12] A. Papoulis and S.U. Pillai, *Probability, random variables, and stochastic processes*, McGraw-Hill New York (2004).
[13] M. Kirby and L. Sirovich, *Application of the karhunen-loeve procedure for the characterization of human faces*, IEEE Transactions Pattern Analysis and Machine Intelligence, 12 (1990), pp. 103–108.
[14] A.S. Georghiades, P.N. Belhumeur and D.J. Kriegman, *From few to many: Illumination cone models for face recognition under variable lighting and pose*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 23 (2001), pp. 643–660.
[15] A.M. Martinez and R. Benavente, *The AR face database*, Tech. Rep., CVC Technical report (1998).
[16] J. Wright, A.Y Yang, A Ganesh, S.S. Sastry and Y. Ma, *Robust face recognition via sparse representation*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 31 (2009), pp. 210–227.
[17] Z. Zhou, A. Wagner, H. Mobahi, J. Wright and Y. Ma, *Face recognition with contiguous occlusion using Markov random fields*, in Proceedings of International Conference on Computer Vision, ICCV (2009).
[18] H. Jia, and A.M. Martinez, *Face recognition with occlusions in the training and testing sets*, in Proc. Conf. Automatic Face and Gesture Recognition, FG(2008).
[19] X. Tan, S. Chen, Z.H. Zhou and F. Zhang, *Recognizing partially occluded, expression variant faces from single training image per person with SOM and soft k-NN ensemble*, IEEE Transactions on Neural Networks, 16 (2005), pp. 875–887.
[20] A.M. Martinez, *Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 24 (2002), pp. 748–763.