Classification on the Grassmannians: Theory and Applications

Jen-Mei Chang

Department of Mathematics and Statistics
California State University, Long Beach
jchang9@csulb.edu

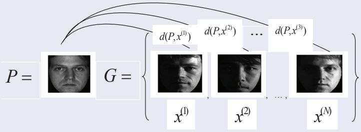
AMS Fall Western Section Meeting
Special Session on *Metric and Riemannian Methods in Shape Analysis*
October 9, 2010 at UCLA

Outline

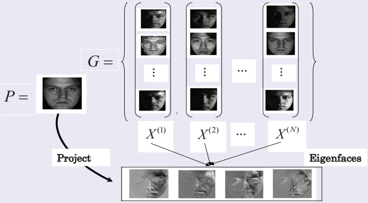
Architectures

Historically

- single-to-single

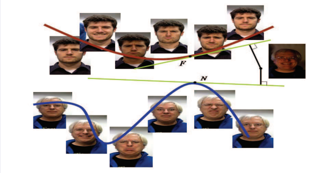


- single-to-many

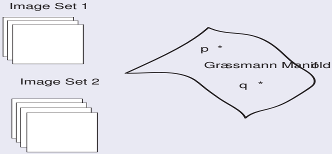


Currently

- subspace-to-subspace



- many-to-many



Some Approaches

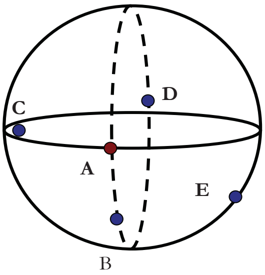
- Single-to-Single
  1. Euclidean distance of feature points.
  2. Correlation.
- Single-to-Many
  1. Subspace method [Oja, 1983].
  2. Eigenfaces (a.k.a. Principal Component Analysis, KL-transform) [Sirovich & Kirby, 1987], [Turk & Pentland, 1991].
  3. Linear/Fisher Discriminate Analysis, Fisherfaces [Belhumeur et al., 1997].
  4. Kernel PCA [Yang et al., 2000].

Some Approaches

- Many-to-Many
  1. Tangent Space and Tangent Distance - Tangent Distance [Simard et al., 2001], Joint Manifold Distance [Fitzgibbon & Zisserman, 2003], Subspace Distance [Chang, 2004].
  2. Manifold Density Divergence [Fisher et al., 2005].
  3. Canonical Correlation Analysis (CCA):
     - Mutual Subspace Method (MSM) [Yamaguchi et al., 1998],
     - Constrainted Mutual Subspace Method (CMSM) [Fukui & Yamaguchi, 2003],
     - Multiple Constrained Mutual Subspace Method (MCMSM) [Nishiyama et al., 2005],
     - Kernel CCA [Wolf & Shashua, 2003],
     - Discriminant Canonical Correlation (DCC) [Kim et al., 2006],
     - Grassmann method [Chang et al., 2006a].

A Quick Comparison

1. Training/Preprocessing.
   - others — yes.
   - proposed — nearly none.
2. Geometry.

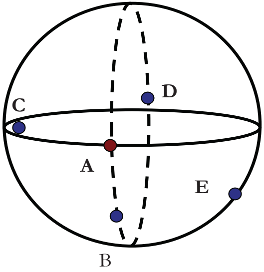


- others — **similarity** measures (e.g., maximum canonical correlation in MSM and sum of canonical correlations in DCC).
- proposed — classification is done on Grassmann manifold, hence Grassmannian **distances/metrics**.

By introducing the idea of Grassmannian, we are able to use many existing tools such as the Grassmannian metrics and Karcher mean to study the geometry of the data sets.

A Quick Comparison

1. Training/Preprocessing.
   - others — yes.
   - proposed — nearly none.
2. Geometry.

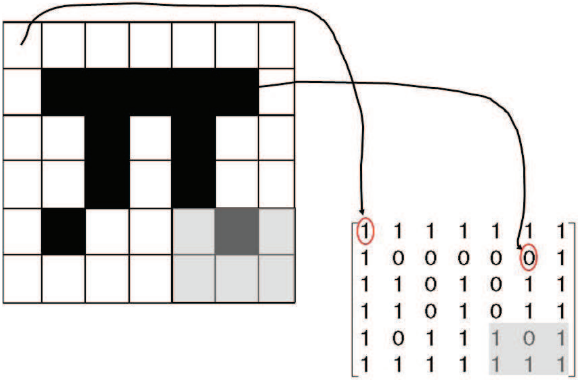


- others — **similarity** measures (e.g., maximum canonical correlation in MSM and sum of canonical correlations in DCC).
- proposed — classification is done on Grassmann manifold, hence Grassmannian **distances/metrics**.

By introducing the idea of Grassmannian, we are able to use many existing tools such as the Grassmannian metrics and Karcher mean to study the geometry of the data sets.

Mathematical Setup

An r-by-c gray scale digital image corresponds to an r-by-c matrix, X, where each entry enumerates one of the 256 possible gray levels of the corresponding pixel.



Mathematical Setup

Realize the data matrix, X, by its columns and concatenate columns into a single column vector, **x**.

$$
\begin{bmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 0 & 0 & 0 & 0 & 0 & 1 \\
1 & 1 & 0 & 1 & 0 & 1 & 1 \\
1 & 1 & 0 & 1 & 0 & 1 & 1 \\
1 & 0 & 1 & 1 & 1 & 0 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1
\end{bmatrix}
\longrightarrow
\begin{bmatrix}
1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ \hline
1 \\ 0 \\ 1 \\ 1 \\ 0 \\ 1 \\ \hline
\vdots \\ \hline
1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1
\end{bmatrix}
$$

Mathematical Setup

That is,

$$X = \begin{bmatrix} \mathbf{x}_1 & | & \mathbf{x}_2 & | & \cdots & | & \mathbf{x}_c \end{bmatrix} \in \mathbb{R}^{r \times c} \quad \longrightarrow \quad \mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_c \end{bmatrix} \in \mathbb{R}^{rc \times 1}$$

Thus, an image J whose matrix representation, X, can be realized as a column vector of length equaling J's resolutions.

$$\boxed{\text{Image} \rightarrow \text{Matrix} \rightarrow \text{Vector}}$$

Mathematical Setup

- Now, for a subject i, we collect k distinct images, which corresponds to k column vectors, $\mathbf{x}_j^{(i)}$ for $j = 1, 2, \ldots, k$.
- Store them into a single data matrix $X^{(i)}$ so that

$$X^{(i)} = \left[\mathbf{x}_1^{(i)} \;\middle|\; \mathbf{x}_2^{(i)} \;\middle|\; \cdots \;\middle|\; \mathbf{x}_k^{(i)} \right].$$

  Note that $\operatorname{rank}(X^{(i)}) = k$ with each $x_j^{(i)} \in \mathbb{R}^n$ being an image of resolution n.
- Associate an orthonormal basis matrix to the column space of $X^{(i)}$ (obtained via, e.g., QR or SVD), $\mathcal{R}(X^{(i)})$. Then $\mathcal{R}(X^{(i)})$ is a k-dimensional vector subspace of $\mathbb{R}^n$.

Mathematical Setup

- Now, for a subject i, we collect k distinct images, which corresponds to k column vectors, $\mathbf{x}_j^{(i)}$ for $j = 1, 2, \ldots, k$.
- Store them into a single data matrix $X^{(i)}$ so that

$$X^{(i)} = \left[\mathbf{x}_1^{(i)} \mid \mathbf{x}_2^{(i)} \mid \cdots \mid \mathbf{x}_k^{(i)} \right].$$

  Note that $\mathrm{rank}(X^{(i)}) = k$ with each $x_j^{(i)} \in \mathbb{R}^n$ being an image of resolution n.
- Associate an orthonormal basis matrix to the column space of $X^{(i)}$ (obtained via, e.g., QR or SVD), $\mathcal{R}(X^{(i)})$. Then $\mathcal{R}(X^{(i)})$ is a k-dimensional vector subspace of $\mathbb{R}^n$.

Mathematical Setup

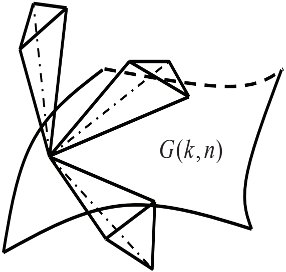
- Now, for a subject i, we collect k distinct images, which corresponds to k column vectors, $\mathbf{x}_j^{(i)}$ for $j = 1, 2, \ldots, k$.
- Store them into a single data matrix $X^{(i)}$ so that

$$X^{(i)} = \left[\mathbf{x}_1^{(i)} \quad | \quad \mathbf{x}_2^{(i)} \quad | \quad \cdots \quad | \quad \mathbf{x}_k^{(i)} \right].$$

  Note that $\text{rank}(X^{(i)}) = k$ with each $x_j^{(i)} \in \mathbb{R}^n$ being an image of resolution n.
- Associate an orthonormal basis matrix to the column space of $X^{(i)}$ (obtained via, e.g., QR or SVD), $\mathcal{R}(X^{(i)})$. Then $\mathcal{R}(X^{(i)})$ is a k-dimensional vector subspace of $\mathbb{R}^n$.

Grassmann Framework

These k-dimensional linear subspaces of $\mathbb{R}^n$ are all elements of a parameter space called the **Grassmannian (Grassmann manifold)**, $G(k,n)$, where n is the ambient resolution dimension.



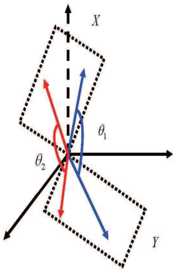
Definition

The *Grassmannian G(k,n)* or the *Grassmann manifold* is the set of k-dimensional subspaces in an n-dimensional vector space K^n for some field K, i.e.,

$$G(k,n) = \{W \subset K^n \mid \dim(W) = k\}.$$

Principal Angles [Björck & Golub, 1973]

It turns out that any attempt to construct an unitarily invariant metric on $G(k,n)$ yields something that can be expressed in terms of the **principal angles** [Stewart & Sun, 1990].



Definition

(**Principal Angles**) If X and Y are two subspaces of $\mathbb{R}^m$, then the principal angles $\theta_k \in \left[0, \frac{\pi}{2}\right]$, $1 \leq k \leq q$ between X and Y are defined recursively by

$$\cos(\theta_k) = \max_{u \in X} \max_{v \in Y} u^T v = u_k^T v_k$$

s.t. $\|u\| = \|v\| = 1$, $u^T u_i = 0$, $v^T v_i = 0$ for $i = 1, 2, \ldots, k-1$ and $q = \min\{\dim(X), \dim(Y)\} \geq 1$.

SVD-based Algorithm for Principal Angles

[Knyazev et al., 2002] For $A \in \mathbb{R}^{n\times p}$ and $B \in \mathbb{R}^{n\times q}$.

1. Find orthonormal bases Q_a and Q_b for A and B such that
$$Q_a^T Q_a = Q_b^T Q_b = I \quad \text{and} \quad \mathcal{R}(Q_a) = \mathcal{R}(A), \mathcal{R}(Q_b) = \mathcal{R}(B).$$
2. Compute SVD for cosine: $Q_a^T Q_b = Y \Sigma Z^T$, $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_q)$.
3. Compute matrix
$$B = \begin{cases} Q_b - Q_a(Q_a^T Q_b) & \text{if } \mathrm{rank}(Q_a) \geq \mathrm{rank}(Q_b); \\ Q_a - Q_b(Q_b^T Q_a) & \text{otherwise.} \end{cases}$$
4. Compute SVD for sine: $[Y, \mathrm{diag}(\mu_1, \ldots, \mu_q), Z] = svd(B)$.
5. Compute the principal angles, for $k = 1, \ldots, q$:
$$\theta_k = \begin{cases} \arccos(\sigma_k) & \text{if } \sigma_k^2 < \frac{1}{2}; \\ \arcsin(\mu_k) & \text{if } \mu_k^2 \leq \frac{1}{2}. \end{cases}$$

Grassmannian Distances [Edelman et al., 1999]

Metric	**Mathematical Expression**
Fubini-Study	$d_{FS}(\mathcal{X}, \mathcal{Y}) = \cos^{-1}\left(\prod_{i=1}^{k} \cos\theta_i\right)$
Geodesic (Arc Length)	$d_g(\mathcal{X}, \mathcal{Y}) = \|\theta\|_2$
Chordal (Projection F-norm)	$d_c(\mathcal{X}, \mathcal{Y}) = \|\sin\theta\|_2$
Projection 2-norm	$d_{p2}(\mathcal{X}, \mathcal{Y}) = \|\sin\theta\|_\infty$
Chordal 2-norm	$d_{c2}(\mathcal{X}, \mathcal{Y}) = \left\|2\sin\frac{1}{2}\theta\right\|_F$
Chordal F-norm	$d_{cF}(\mathcal{X}, \mathcal{Y}) = \left\|2\sin\frac{1}{2}\theta\right\|_2$

Various Realizations of the Grassmannian

1. First, as a quotient (homogeneous space) of the orthogonal group,

$$G(k,n) = O(n)/O(k) \times O(n-k). \tag{1}$$

2. Next, as a submanifold of projective space,

$$G(k,n) \subset \mathbb{P}(\Lambda^q \mathbb{R}^n) = \mathbb{P}^{\binom{n}{k}-1}(\mathbb{R}) \tag{2}$$

via the Plücker embedding.

3. Finally, as a submanifold of Euclidean space,

$$G(k,n) \subset \mathbb{R}^{(n^2+n-2)/2} \tag{3}$$

via a projection embedding described in [Conway et al., 1996].

The Corresponding Grassmannian Distances

1. The standard invariant Riemannian metric on orthogonal matrices $O(n)$ descends via (1) to a Riemannian metric on the homogeneous space $G(k, n)$. We call the resulting geodesic distance function on the Grassmannian the *arc length* or *geodesic* distance and denote it d_g.
2. If one prefers the realization (2), then the Grassmannian inherits a Riemannian metric from the *Fubini-Study* metric on projective space (see, e.g., [Griffiths & Harris, 1978]).
3. One can restrict the usual Euclidean distance function on $\mathbb{R}^{(n^2+n-2)/2}$ to the Grassmannian via (3) to obtain the *projection F* or *chordal* distance d_c.

Grassmannian Semi-Distances

- Often time, the data set is compact and fixed.
- First few principal angles contain discriminatory information and are less sensitive to noise.
- Thus, it is natural to consider the nested subspaces. Define the ℓ-truncated principal angle vector $\theta^\ell := (\theta_1, \theta_2, \ldots, \theta_\ell)$. Then we have example ℓ-truncated Grassmannian semi-distances:

$$d_g^\ell := \|\theta^\ell\|_2, \qquad d_{FS}^\ell := \cos^{-1} \prod_{i=1}^{\ell} \cos\theta_i,$$

$$d_c^\ell := \|\sin\theta^\ell\|_2 \qquad d_{cF}^\ell := \|2\sin\tfrac{1}{2}\theta^\ell\|_2.$$

Separation Gap & Grassmann Separable

Given a set of image sets $\mathcal{P} = \{X_1, X_2, \ldots, X_m\}$, where $X_i \in \mathbb{R}^{n \times k_i}$ and each X_i belongs to one of the subject class C_j.

- Let **cardinality** of a set of images be the number of distinct images used.
- The distances between different realizations of subspaces for the same class are called **match distances** while for different classes they are called **non-match distances**.
- $W_i = \{j \mid X_j \in C_i\}$, the within-class set of subject i, and $B_i = \{j \mid X_j \notin C_i\}$, the between-class set of subject i.

Separation Gap & Grassmann Separable

- Let M **be the maximum of the match distances**

$$M = \max_{1\le i\le m} \max_{j\in W_i} d(X_i, X_j)$$

  and m **be the minimum of the non-match distances**

$$m = \min_{1\le i\le m} \min_{k\in B_i} d(X_i, X_k),$$

  then define the **separation gap** to be $g_s = m - M$.

- Then we say the set $\mathcal{P}$ is **Grassmann separable** if the separation gap is positive. i.e.,

$$\boxed{g_s > 0 \Leftrightarrow \mathcal{P} \text{ is Grassmann separable}}$$

Separation Gap & Grassmann Separable

- Let M **be the maximum of the match distances**

$$M = \max_{1 \le i \le m} \max_{j \in W_i} d(X_i, X_j)$$

  and m **be the minimum of the non-match distances**

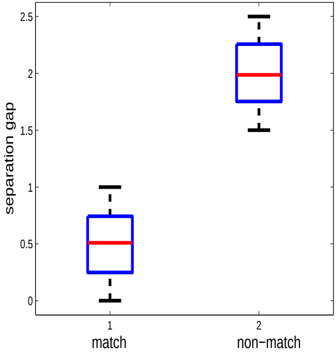
$$m = \min_{1 \le i \le m} \min_{k \in B_i} d(X_i, X_k),$$

  then define the **separation gap** to be $g_s = m - M$.

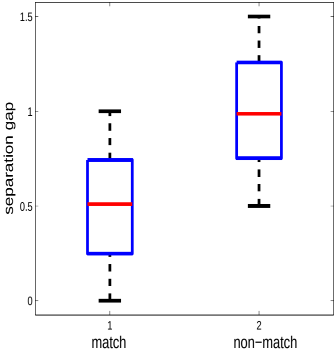
- Then we say the set $\mathcal{P}$ is **Grassmann separable** if the separation gap is positive. i.e.,

$$\boxed{g_s > 0 \Leftrightarrow \mathcal{P} \text{ is Grassmann separable}}$$

A Graphical Example



Grassmann separable



Non-Grassmann separable

Measure of Classification Rates

- **False accept rate (FAR)** is the ratio of the number of false acceptances divided by the number of identification attempts.
- **False reject rate (FRR)** is the ratio of the number of false rejections divided by the number of identification attempts.
- Given match and non-match distances for a set of classes, the **false accept rate (FAR) at a zero false reject rate (FRR)** (defined, e.g., in [Mansfield & Wayman, 2002]) is **the ratio of the number of non-match distances that are smaller than the maximum of the match distances divided by the number of non-match distances.**

$$\boxed{\text{zero percent FAR at a zero FRR} \Longleftrightarrow g_s > 0}$$

Measure of Classification Rates

- **False accept rate (FAR)** is the ratio of the number of false acceptances divided by the number of identification attempts.
- **False reject rate (FRR)** is the ratio of the number of false rejections divided by the number of identification attempts.
- Given match and non-match distances for a set of classes, the **false accept rate (FAR) at a zero false reject rate (FRR)** (defined, e.g., in [Mansfield & Wayman, 2002]) is **the ratio of the number of non-match distances that are smaller than the maximum of the match distances divided by the number of non-match distances.**

$$\boxed{\text{zero percent FAR at a zero FRR} \Longleftrightarrow g_s > 0}$$

Measure of Classification Rates

- **False accept rate (FAR)** is the ratio of the number of false acceptances divided by the number of identification attempts.
- **False reject rate (FRR)** is the ratio of the number of false rejections divided by the number of identification attempts.
- Given match and non-match distances for a set of classes, the **false accept rate (FAR) at a zero false reject rate (FRR)** (defined, e.g., in [Mansfield & Wayman, 2002]) is **the ratio of the number of non-match distances that are smaller than the maximum of the match distances divided by the number of non-match distances.**

$$\boxed{\text{zero percent FAR at a zero FRR} \iff g_s > 0}$$

Empirical fact

Images of a single person seen under variations of illumination appear to be more difficult to recognize than images of different people [Zhao et al., 2003].

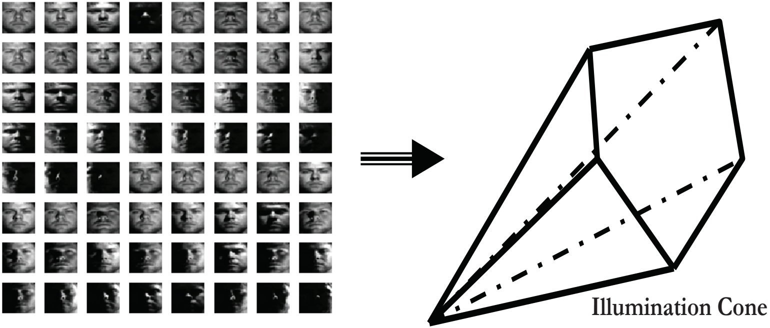
Subject 1

Subject 2

Can you tell who this is?

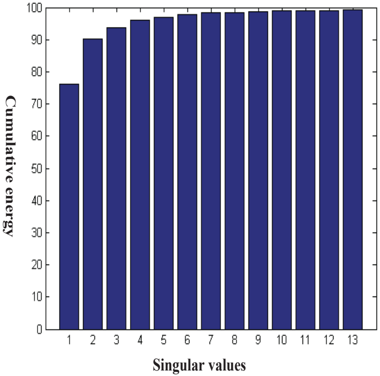
Geometric facts - 1

The set of m-pixel monochrome images of an object seen under general lighting conditions forms a convex polyhedral cone (illumination cone) in $\mathbb{R}^m$ [Belhumeur & Kriegman, 1998].

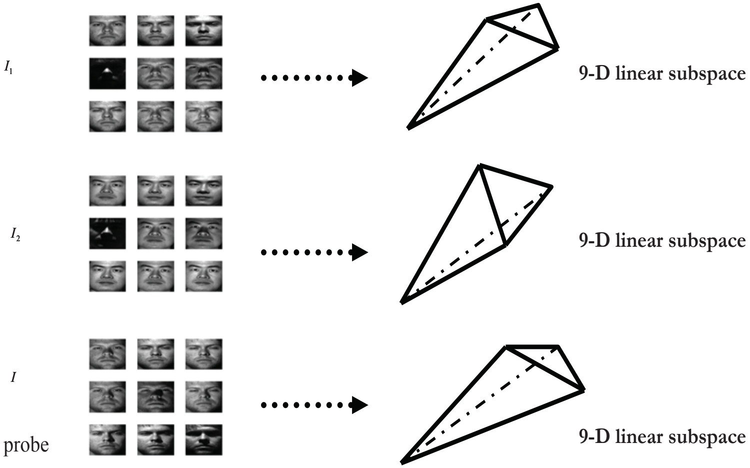


Geometric facts - 2

The illumination cone can be approximated by a 9-dimensional linear subspace [Basri & Jacobs, 2003], i.e., the illumination cone is low-dimensional and linear.

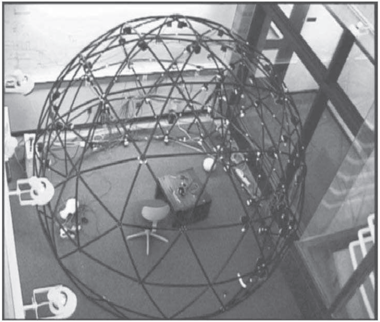


Grassmann Set-up

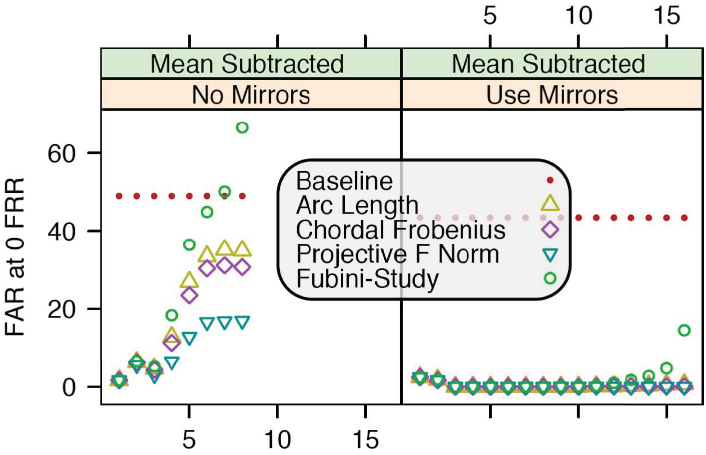


Yale Face Database B (YDB)

10 subjects, 64 illumination conditions, 9 poses



Classification Result [Chang et al., 2006a]

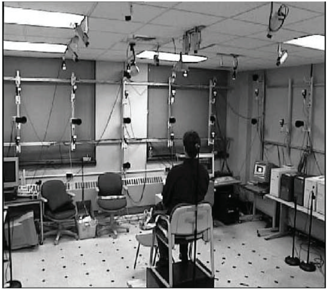


CMU-PIE

We fix the frontal pose, neutral expression and select the "illum" and "lights" subsets of CMU-PIE (68 subjects, 13 poses, 43 lightings, 4 expressions) [Sim et al., 2003] for experiments.

(a) lights: 21 illumination conditions with background lights **on**.

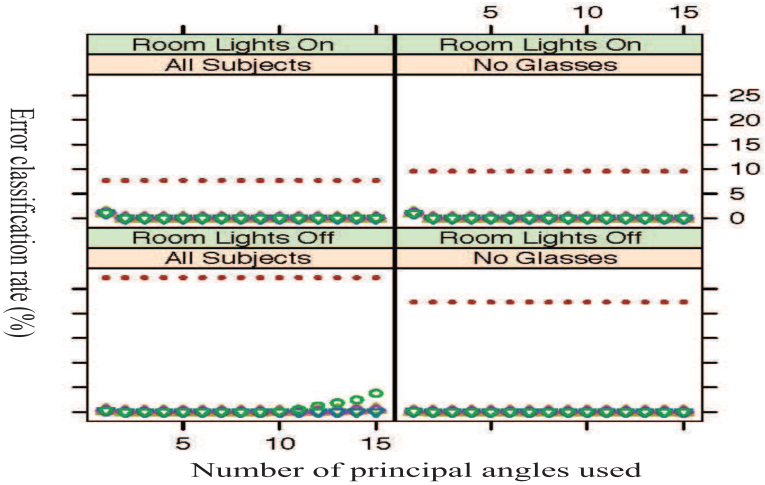
(b) illum: 21 illumination conditions with background lights **off**.



(a) "lights" subset

(b) "illum" subset

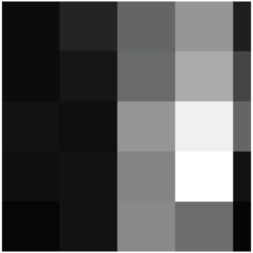
Classification Result [Chang et al., 2006a]



Patch Collapsing [Chang et al., 2007b]

If the data set is Grassmann separable using subject illumination subspaces of this kind of image [Chang et al., 2006a]:

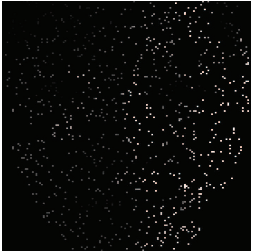
The data set is still Grassmann separable using subject illumination subspaces of this kind of image [Chang et al., 2007b]:



Patch Projection [Chang et al., 2007c]

If the data set is Grassmann separable using subject illumination subspaces of this kind of image [Chang et al., 2006a]:

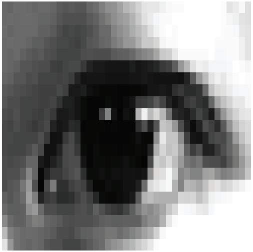
The data set is still Grassmann separable using subject illumination subspaces of this kind of image [Chang et al., 2007c]:



Patch Projection [Chang et al., 2007c]

If the data set is Grassmann separable using subject illumination subspaces of this kind of image [Chang et al., 2006a]:

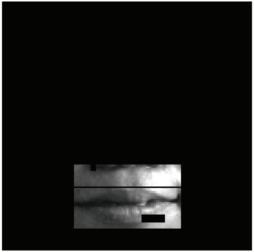
The data set is still Grassmann separable using subject illumination subspaces of this kind of image [Chang et al., 2007c]:



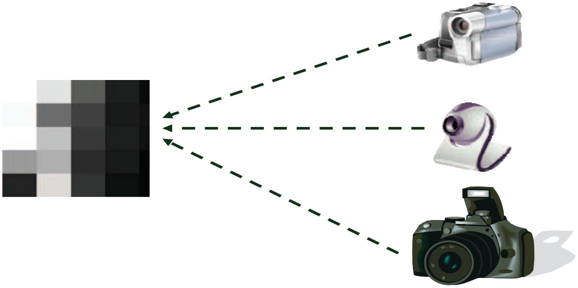
Patch Projection [Chang et al., 2007c]

If the data set is Grassmann separable using subject illumination subspaces of this kind of image [Chang et al., 2006a]:

The data set is still Grassmann separable using subject illumination subspaces of this kind of image [Chang et al., 2007c]:



Potential Use: Low-Res. Illumination Camera



Large private databases of facial imagery can be stored at a resolution that is sufficiently low to prevent recognition by a human operator yet sufficiently high to enable machine recognition.

Karcher Mean

- How should we choose subject subspace representations given a set of images?
- Patch collapsing (e.g., low res. images) and projections (e.g., lip and nose feature patches) provide one way of compression. In particular, compression in n for points in $G(k,n)$.
- What about compression in the other parameter, k?
- To this end, we will use another geometric concept, Karcher mean, on the Grassmann manifold to accomplish this.

Karcher Mean

- How should we choose subject subspace representations given a set of images?
- Patch collapsing (e.g., low res. images) and projections (e.g., lip and nose feature patches) provide one way of compression. In particular, compression in n for points in $G(k,n)$.
- What about compression in the other parameter, k?
- To this end, we will use another geometric concept, Karcher mean, on the Grassmann manifold to accomplish this.

Karcher Mean

- How should we choose subject subspace representations given a set of images?
- Patch collapsing (e.g., low res. images) and projections (e.g., lip and nose feature patches) provide one way of compression. In particular, compression in n for points in $G(k, n)$.
- What about compression in the other parameter, k?
- To this end, we will use another geometric concept, Karcher mean, on the Grassmann manifold to accomplish this.

Karcher Mean

- How should we choose subject subspace representations given a set of images?
- Patch collapsing (e.g., low res. images) and projections (e.g., lip and nose feature patches) provide one way of compression. In particular, compression in n for points in $G(k,n)$.
- What about compression in the other parameter, k?
- To this end, we will use another geometric concept, Karcher mean, on the Grassmann manifold to accomplish this.

Notions of Mean

- For a set of points $\{x^{(1)}, x^{(2)}, \ldots, x^{(P)}\} \in \mathbb{R}^n$, its Euclidean mean is the x that minimizes the sum squared distance

$$\sum_{i=1}^{P} d^2\left(x - x^{(i)}\right),$$

where d is the straight-line distance defined by the vector 2-norm.

- Given the points $p_1, \ldots, p_m \in G(k,n)$, the Karcher mean is the point q^* that minimizes the sum of the squares of the geodesic distance between q^* and p_i's, i.e.,

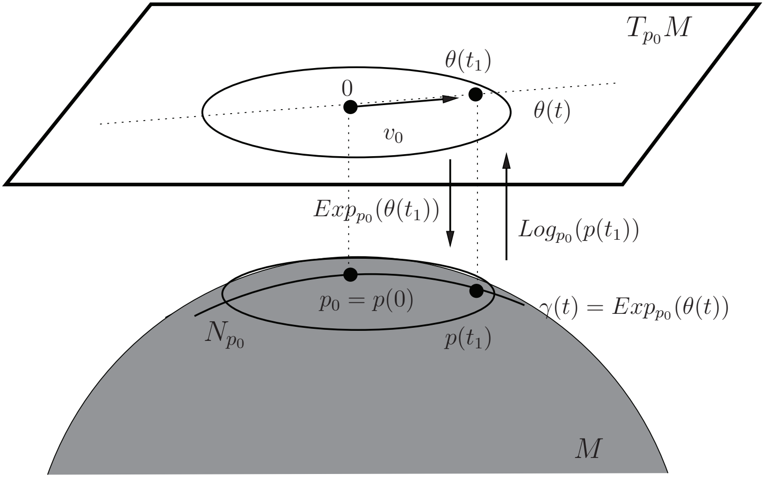
$$q^* = \arg\min_{q \in G(k,n)} \frac{1}{2m} \sum_{j=1}^{m} d^2(q, p_j),$$

where $d(q,p)$ is the geodesic distance between p and q on $G(k,n)$.

Notions of Mean

- For a set of points $\{x^{(1)}, x^{(2)}, \ldots, x^{(P)}\} \in \mathbb{R}^n$, its Euclidean mean is the x that minimizes the sum squared distance
$$\sum_{i=1}^{P} d^2\left(x - x^{(i)}\right),$$
where d is the straight-line distance defined by the vector 2-norm.
- Given the points $p_1, \ldots, p_m \in G(k,n)$, the Karcher mean is the point q^* that minimizes the sum of the squares of the geodesic distance between q^* and p_i's, i.e.,
$$q^* = \operatorname*{arg\,min}_{q \in G(k,n)} \frac{1}{2m} \sum_{j=1}^{m} d^2(q, p_j),$$
where $d(q,p)$ is the geodesic distance between p and q on $G(k,n)$.

Descent Algorithm [Rahman et al., 2005]



An SVD-based Algorithm [Begelfor & Werman, 2003]

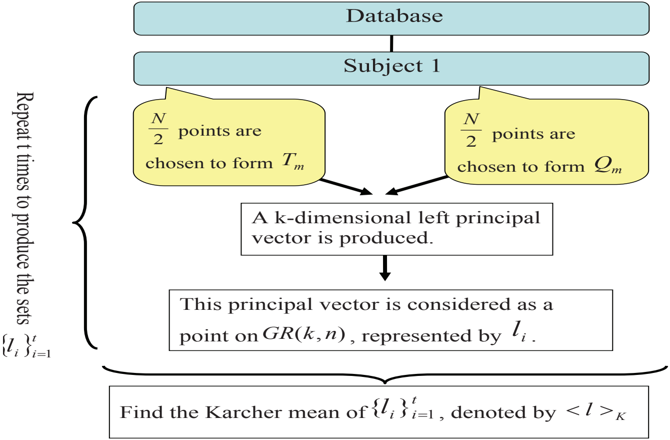
For points $p_1, p_2, \ldots, p_m \in G(k,n)$ and ϵ (machine zero), find the Karcher mean, q.

1. Set $q = p_1$.
2. Find
$$A = \frac{1}{m}\sum_{i=1}^{m} \mathrm{Log}_q(p_i).$$
3. If $||A|| < \epsilon$, return q, else, go to step 4.
4. Find the SVD $U\Sigma V^T = A$ and update
$$q \to qV\cos(\Sigma) + U\sin(\Sigma).$$
Go to step 2.

Note: the map in step 4 is the *Exponential map* that takes points from the tangent space back to the manifold.

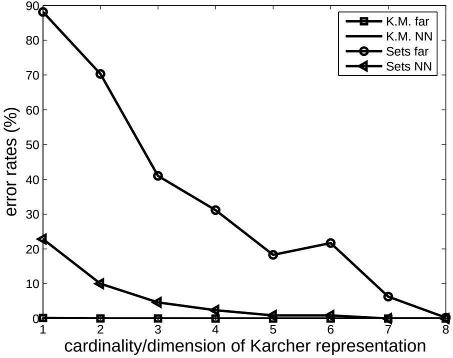
An Example Result

Compress data with k-d Karcher mean and compare the recognition result to results obtained using k raw images.



An Example Result

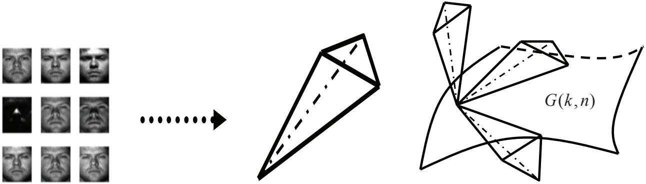
- 16 images used for gallery pts; 3 images used for probes.
- Lip patch on the CMU-PIE “lights” data set.



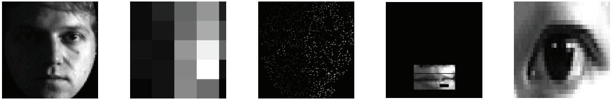
The fact that using a 1-d Karcher representation achieves a perfect recognition result while using 1 raw image in the gallery does not indicates that Karcher representations are able to pack useful information more efficiently.

Conclusions

- A novel geometric framework for a many-to-many architecture — Grassmann framework.

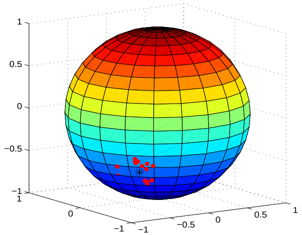


- Empirical results and new insights.



Conclusions

- A novel algorithm for Karcher compression.



Acknowledgements

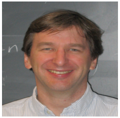
- National Science Foundation award# DMS-0434351 and the DOD-USAF-Office of Scientific Research under contract FA9550-04-1-0094.
- Research Group from Colorado State University:

Michael

Ross

Bruce

Holger



Chris

Selected References

[Chang et al., 2006a] J.-M. Chang, M. Kirby, H. Kley, J. R. Beveridge, C. Peterson, B. Draper, "Illumination face spaces are idiosyncratic", *Int'l Conf. on Image Proc. & Comp. Vision, 2: 390–396, 2006*.

[Chang et al., 2006b] J.-M. Chang, M. Kirby, H. Kley, J. R. Beveridge, C. Peterson, B. Draper, "Examples of set-to-set image classification", *7th Int'l Conf. on Math. in Signal Proc. Conf. Digest, 102–105, 2006*.

[Chang et al., 2007a] J.-M. Chang, M. Kirby, C. Peterson, "Set-to-set face recognition under variations in pose and illumination", *Proceedings of 2007 Biometric Symposium at the Biometrics Consortium Conference, 2007*.

[Chang et al., 2007b] J.-M. Chang, M. Kirby, H. Kley, J. R. Beveridge, C. Peterson, B. Draper, "Recognition of digital images of the human face at ultra low resolution via illumination spaces", *ACCV'07, LNCS, Springer, 4844: 733–743, 2007*.

[Chang et al., 2007c] J.-M. Chang, M. Kirby, C. Peterson, "Feature Patch Illumination spaces and Karcher compression for face recognition via Grassmannian", *submitted, 2008*.