Numerical Analysis and Statistics on Tensor Parameter Spaces

Chris Peterson

SIAM - AG11 - Tensors

Oct. 7, 2011

Overview

- Karcher mean / Normal mean - finding representatives for a set of points on a Grassmann manifold

- A clustering example with images.

- Representing multi-scale clustering on the Grassmann manifold

This talk includes joint work with Michael Kirby and Justin Marks

It has benefited from discussions with many friends including:

Ross Beveridge
Bruce Draper
Holger Kley
Jen-Mei Chang

Normal Mean / Karcher Mean

The goal of this section is to describe a method for obtaining a representative for a cluster of points on a Grassmann manifold.

The representative is obtained as the output of a particular iterative algorithm applied to the cluster of points.

The algorithm is based on an explicit description of the tangent spaces to the manifold arising from a standard realization as a quotient space.

The algorithm depends upon a pair of mutually inverse maps, one for each point of the manifold.

From the cluster of points, the algorithm exploits these maps in a predictor/corrector loop until converging, with prescribed tolerance, to a fixed point.

The fixed point acts as a representative of the cluster.

One view of the Grassmannian

An $n \times p$ matrix Y, with $Y^T Y = I$, is a representative for the p-dimensional subspace spanned by the columns of Y.

If Q is an orthonormal $p \times p$ matrix then Y and YQ represent the same point on $Gr(p, n)$ since they have the same column space.

Proceeding in this manner, we can obtain the identification $Gr(p, n) = O(n)/(O(n-p) \times O(p))$

Of course, there are many other ways to think of $Gr(p, n)$.

Principal Angles

Given two k-dimensional spaces A, B, the first principal angle, θ_1, between the two spaces is obtained by finding vectors $u_1 \in A$ and $v_1 \in B$ that have the smallest possible angle between them.

The second principal angle, θ_2, is obtained by finding vectors $u_2 \in A$ and $v_2 \in B$ that have the smallest possible angle subject to the constraint that $u_1 \cdot u_2 = 0$ and $v_1 \cdot v_2 = 0$.

Continuing in this way we obtain a vector of principal angles: $\Theta(A, B) = (\theta_1, \theta_2, \ldots, \theta_k)$.

It is an interesting and useful fact that the principal angles can be determined from the Singular Value Decomposition of a certain matrix built from x and y.

In particular, let A, B be orthonormal matrix representatives for V_x and V_y.

In the singular value decomposition $A^T B = U \Sigma V^T$, the diagonal entries of Σ are the cosines of the principal angles.

Any orthogonally invariant (resp. unitarily invariant) metric on $Gr(p, n)$ is an appropriate function of the principal angles.

For instance, the Fubini-Study metric (from the Plucker embedding) is $d(A, B) = \cos^{-1}(\prod_{i=1}^{k} \cos\theta_i)$

The metric corresponding to $Gr(p, n) = O(n)/(O(n-p) \times O(p))$ is $d(A, B) = (\sum_{i=1}^{k} \theta_i^2)^{1/2} = ||\Theta(A, B)||_2$

There is also an embedding of $Gr(k, n)$ into $\mathbb{R}^{(n^2+n-2)/2}$ that leads to the metric $||\sin\Theta(A, B)||_2$

Set-up for description of algorithm

Let $P \in Gr(p, n)$.

Let N_p denote a neighborhood of P on $Gr(p, n)$.

Let T_P denote the tangent space to Gr(p,n) at P.

Suppose that for each point $P \in Gr(p, n)$, you have maps
$F_P : N_P \to T_P$
$G_P : T_P \to Gr(p, n)$
such that F_P and G_P are inverses (on suitably restricted sets).

Consider a cluster of points $\mathcal{C} \subset Gr(p, n)$.
An outline of the iterative algorithm is:

1) Pick a point near $\mathcal{C}$, call it P.
2) Determine the Tangent space, T_P, to $Gr(p, n)$ at P
3) Use F_P to map each point in $\mathcal{C}$ to T_P.
4) Use the points in this linear space to produce a representative Q
5) Compute $G_P(Q)$ to obtain a new point on $Gr(p, n)$.
6) Update P to $G_P(Q)$ and go to step 2)

There are multiple possibilities for F_P, G_P and for step 4.

In particular, the principle angles can be used in many ways to produce Q from $F_P(\mathcal{C})$.

This leads to algorithms that can be optimized for different purposes such as suppressing the effect of outliers, etc.

The idea for the algorithm came from the Karcher mean algorithm (which uses Exp and Log maps in the context of lie groups to move back and forth between the lie algebra and the lie group)

Set-up for an example with digital images

Black and white images as rectangular arrays

When a picture is captured by a digital camera as a gray scale image, the pictured is stored as a rectangular array of numbers.

Each number in the array corresponds to a pixel in the image.

Frequently, the numbers in the array lie between 0 and 255.

The numbers record the energy arriving from the corresponding portion of the image.

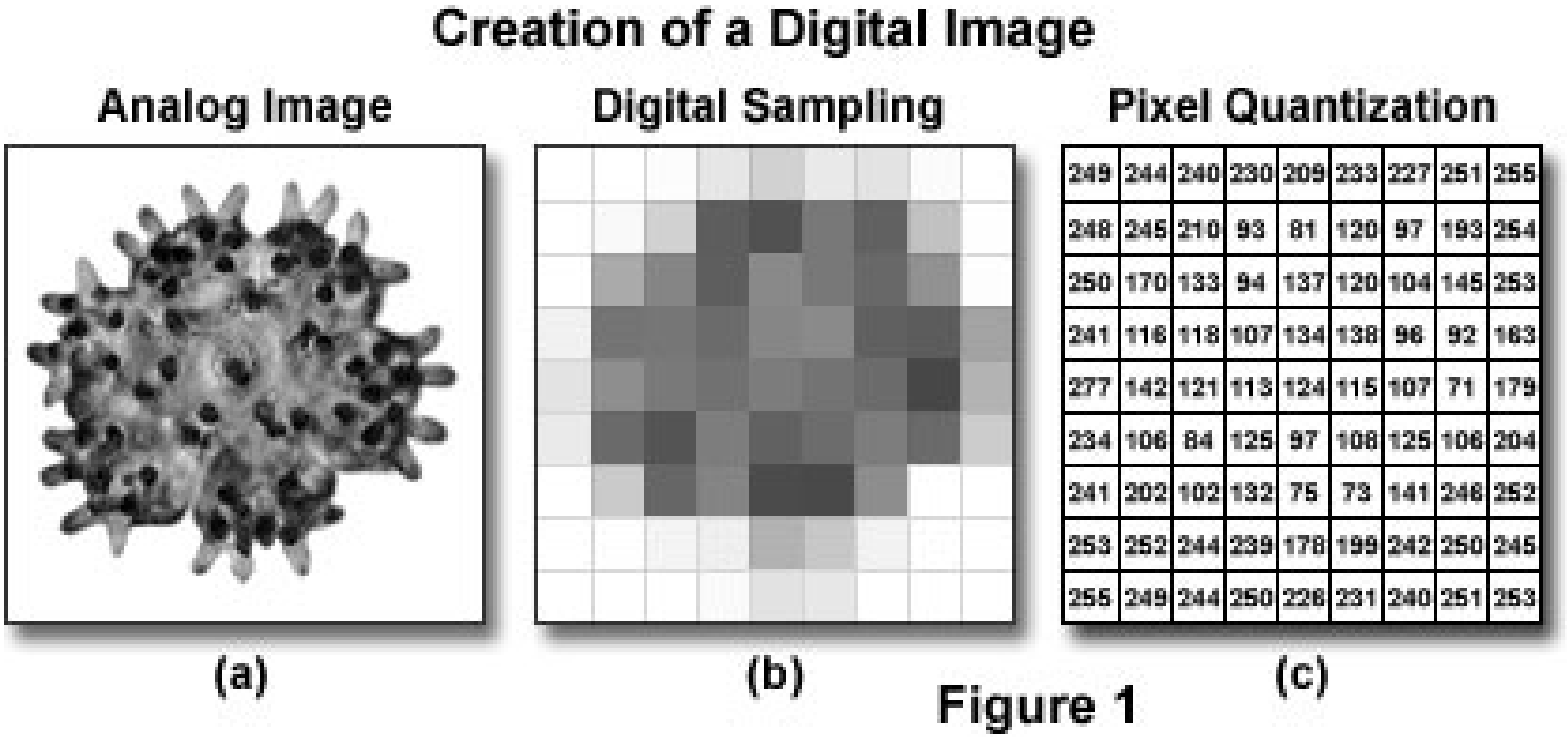


Figure 1

http://micro.magnet.fsu.edu/primer/digitalimaging/digitalimagebasics.html

Color images as rectangular hyper-arrays

When a picture is captured by a digital camera as a color image, the picture is typically stored as a three sheeted rectangular array of numbers.

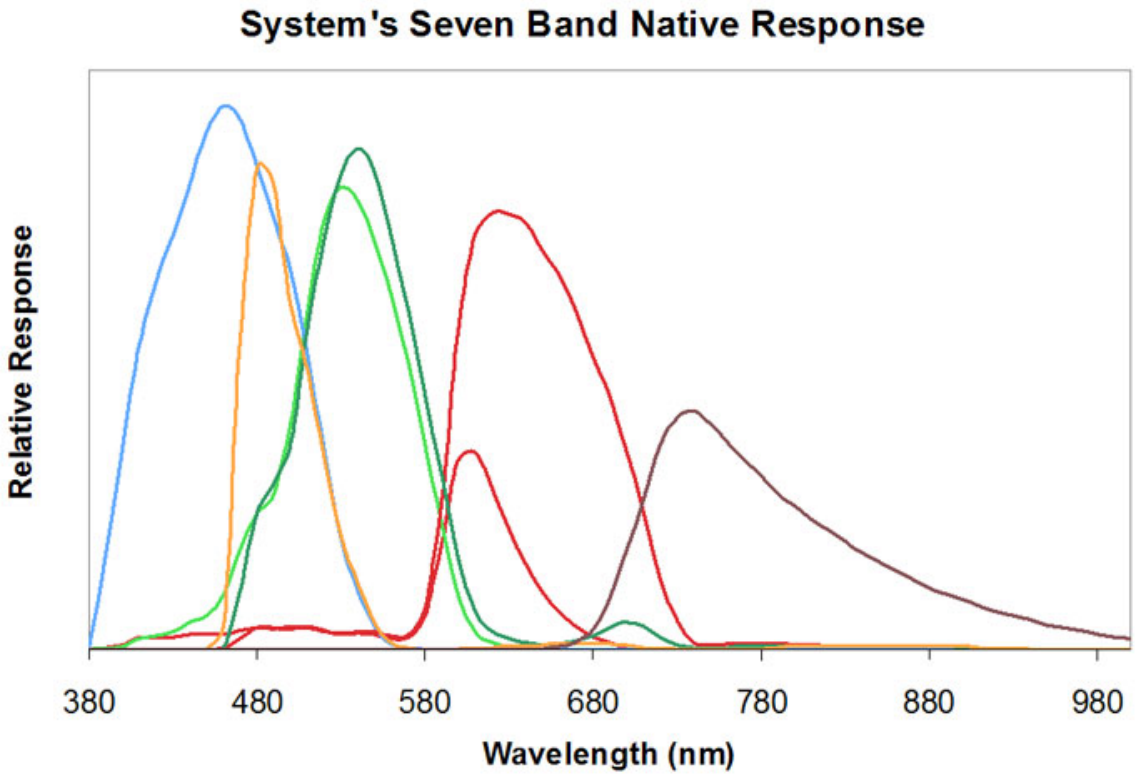
Each sheet in the array corresponds to energy arriving in a particular color band.

Some higher end cameras produce data with more sheets.

Black and white images are contractions of these sheets into a two dimensional array.

A pixel value (roughly) corresponds to an integral of a frequency dependent energy function against a frequency response curve.

Below are some curves from a higher end multi-spectral camera.



http://www.fluxdata.com/products/high-resolution-3-ccd-multispectral-camera

Most digital cameras have the ability to collect data from outside the visible spectrum.

It can be surprising when you "shift" data outside the visible spectrum to lie in the visible spectrum.

Below is a flower imaged in the visible spectrum and the same flower imaged in the UV-spectrum.

http://www.naturfotograf.com

Digital images as vectors

A gray scale digital picture is a point in $\mathbb{R}^{r \times c}$.

By flattening the matrix, we obtain a point in $\mathbb{R}^{rc \times 1}$.

A color image is a point in $\mathbb{R}^{r \times c \times s}$ and this can be flattened into a single point in $\mathbb{R}^{rcs \times 1}$, or into s points in $\mathbb{R}^{rc \times 1}$, etc.

Once flattened, a collection of images can then be stored as columns of a data matrix.

A specific example with digital images

Illumination space

We now consider the illumination space of an object.

More precisely, we will consider the special geometry of the span of a collection of images of a fixed object under varying illumination conditions.

As a specific example, we start with a collection of gray scale images of a person under different lighting.

Images of a person under different illumination conditions




Illumination images form a convex set

If $\mathbf{A}, \mathbf{B}$ are vectors representing two images of the same object under two different illumination conditions, then any convex combination of $\mathbf{A}$ and $\mathbf{B}$ represents the same object under a convex combination of these lighting conditions.

As a consequence, the set of vectors representing images of a fixed object collected under varying illumination conditions form a convex set.

In concrete terms, any weighted "average" of two such images is a valid image.

An average illumination image.

The picture in the middle is artificial. It was created by adding the matrices and dividing by 8.

Singular values of an illumination data matrix

Basri and Jacobs (2000) showed that the set of all reflectance functions produced by Lambertian objects under distant, isotropic lighting lies close to a 9D linear subspace.

An implication of this result is that the set of images of a convex Lambertian object obtained under a wide variety of lighting conditions can be well approximated by a low-dimensional linear subspace.

Thus, if the columns of a data matrix consist of images of a fixed object taken under a wide variety of illumination conditions, then the singular values of the matrix decay rapidly.

Objects which do not reflect light very well are Lambertian, they have the effect of smoothing (rugs and fur are Lambertian).

Objects which reflect light very well are not Lambertian (mirror disco balls are not Lambertian).

Illumination data matrices of non-Lambertian objects have a slower decay of singular values.

Since the singular values of an illumination data matrix decay rapidly, there is a low dimensional linear space that captures the vast majority of the energy of an illumination data set.

For a given k, the singular value decomposition tells us which k-linear space to pick. This linear space is the illumination space. It is a subspace of a (much) larger space.

As a subspace, the illumination space can be identified with a single point in an appropriate Grassmannian.

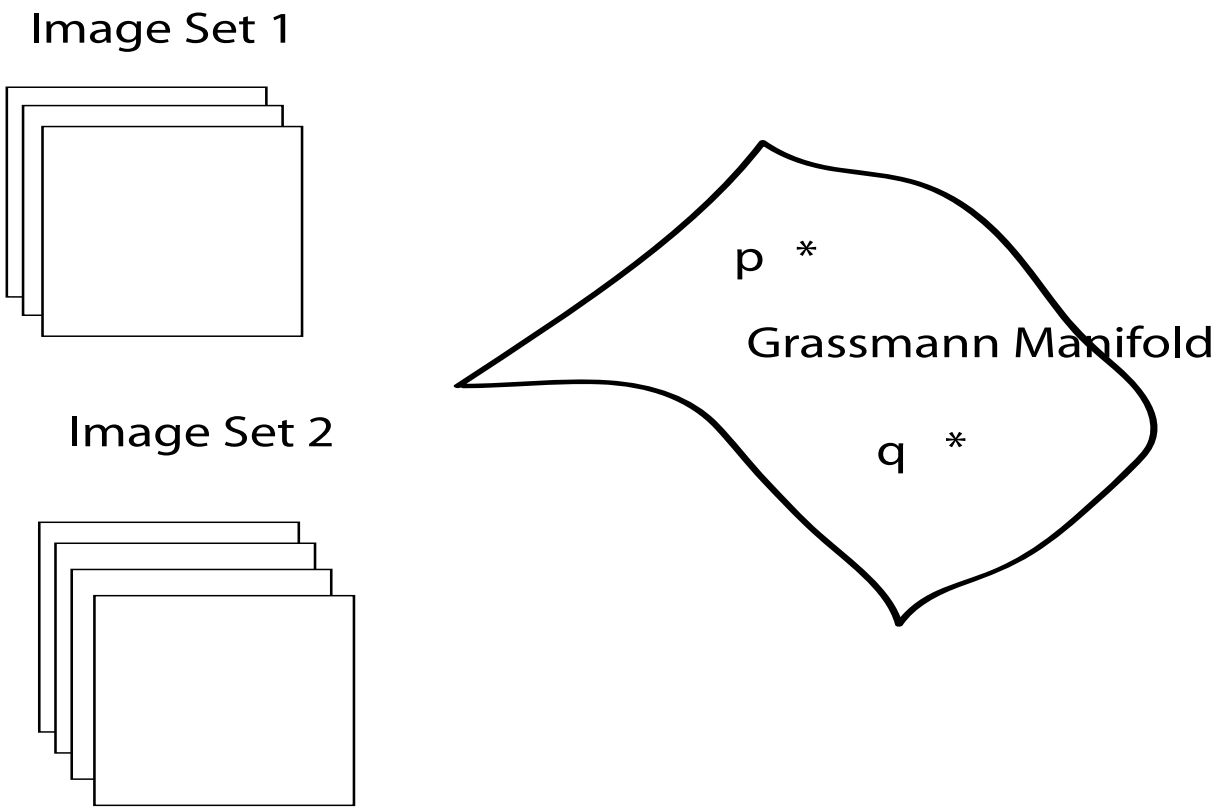


Image Set 1
Image Set 2
p
Grassmann Manifold
q

The CMU-PIE data base is a database (developed at Carnegie-Mellon University) containing images of people with variations in **I**llumination, **P**ose and **E**xpression.

For the following experiment, we restrict our attention to images drawn from the CMU-PIE database with a variation only in Illumination.

Description of experiment

For each of the 67 subjects in the CMU-PIE data base, we randomly selected two disjoint sets of 10 images to produce two 10-dimensional estimates of the illumination space for the subject.

The process of random selection was repeated 10 times to generate a total of 670 matching subspaces and 44,220 non-matching subspaces.

Let D be a distance measure on a Grassmann variety. When the largest distance between any two matching subspaces is less than the smallest distance between any two non-matching subspaces, the data is called D-Grassmann separable.

Using the first principal angle, we observed a significant gap between matching and non-matching subspaces (approximately 16°) when subspaces are realized as points in $G(10, 22080)$.

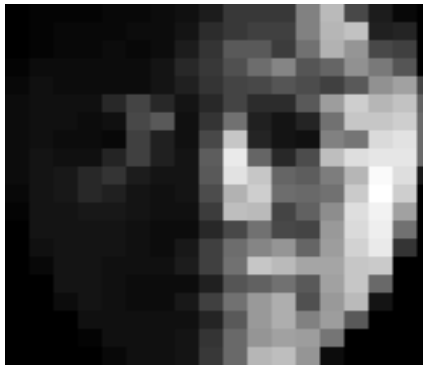
I.e. the data is Grassmann separable w.r.t. the first principal angle.

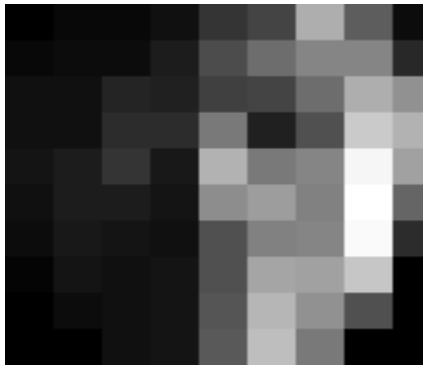
The difference between 10 and 22080 is huge.

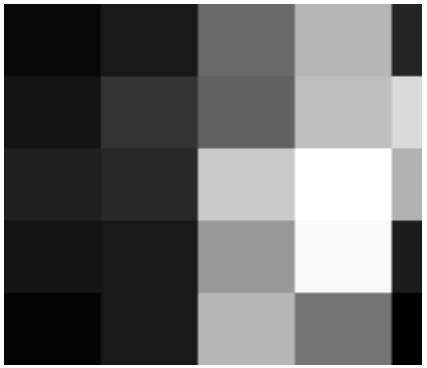
This suggests that a generic projection may be close to an isometry. In other words, a general mapping could be expected to roughly preserve interrelationships between angles.

In the next experiment, we describe what happens if we consider a rather special linear map derived from the 2-d Haar wavelet.

Essentially, all that is being done is to replace a region of an image with its average pixel value.





Application of the 2D Haar wavelet transform.

The resulting LL image is displayed at each level.

This transform mimics a low resolution camera.

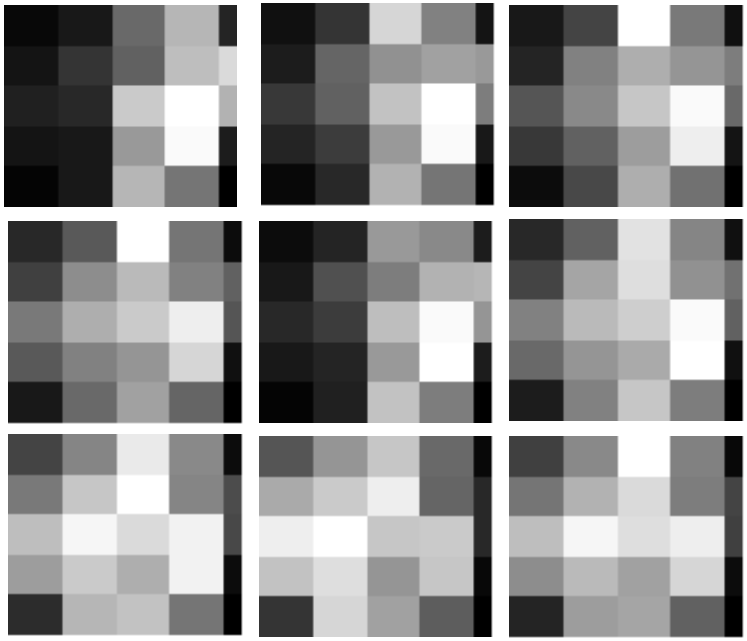
25 pixels

Even under drastic resolution reduction, the CMU-PIE data base is still Grassmann separable.

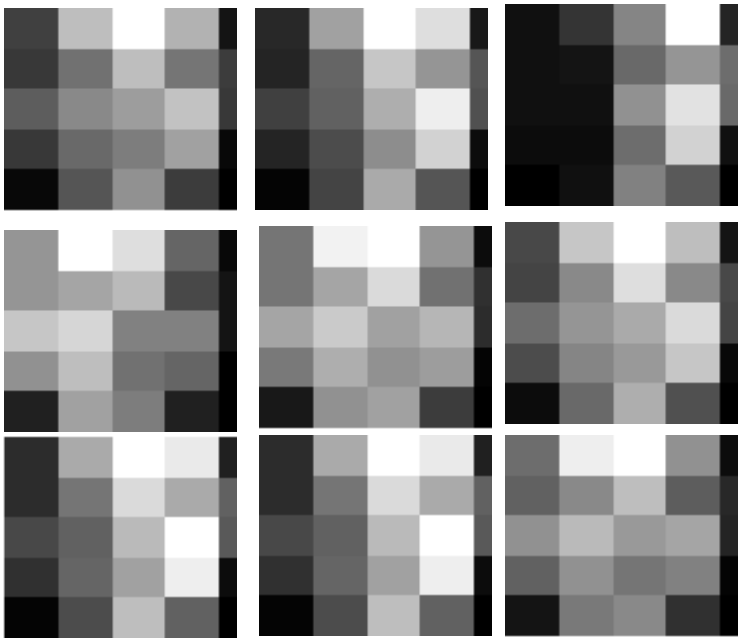
In other words, it is still possible to match the reduced resolution illumination spaces.

However, the separation gap between matching and non-matching subspaces drops from 16° to 14°, 8°, and 0.17° when subspaces are realized as points in $G(10, 22080)$, $G(10, 360)$, $G(10, 90)$, and $G(10, 25)$, respectively.

Low resolution illumination spaces



Subject A



Subject B

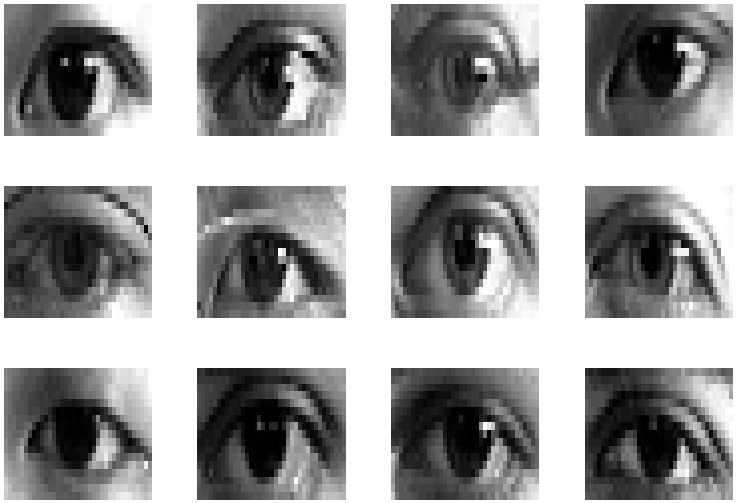
Two other special types of projections are patch projections and dust projections.

A patch projection will be taken to mean restriction of each image to a collection of patches within the image.

A dust projection will be taken to mean restriction of each image to a fixed collection of random pixels within the image.

Both patch projections and dust projections are Grassmann separable provided enough pixels are used.

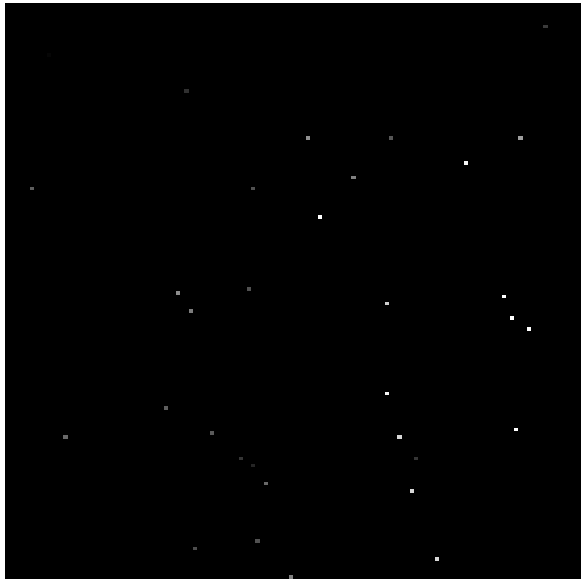
Right eye patches



Lip patches

Sampling of a strip of pixels

Sampling of a random set of pixels



Data bundles

Imagine you have a data set involving two (or more) variations of state.

In this setting there are natural maps obtained by fixing one or more of the states.

These maps break up the data set into a fiber bundle (which we call a data bundle).

Consider pictures of an object as it is spinning on a turntable.

If the pictures are gray scale, then the collection of images in $\mathbb{R}^{rc\times 1}$ correspond to the image of a map of a circle S^1.

If the pictures are color with each image stored as an $\mathbb{R}^{r\times c\times 3}$ matrix then we can map each image to three points in $\mathbb{R}^{rc\times 1}$.

There is a natural 3 to 1 map of this data to a circle.

Alternatively, we can map each color image to the span of the three points.

In this manner we obtain a map of S^1 into $G(3, rc)$.

Now suppose an illumination condition is fixed but digital images of an object are collected from all possible orientations of a camera subject to the condition that the camera is 2 meters from the object and is pointed towards the center of the object.

The illumination condition and the object give a map of $SO(3)$ into pixel space.

Roughly speaking, this map is a circle bundle over a sphere.

Finally, suppose we allow all possible illumination conditions and we allow all possible orientations of the camera.

Over each point of $SO(3)$ we have an illumination space.

In other words, we have an illumination bundle over $SO(3)$.

Fixing an illumination condition corresponds to taking a section of this bundle.

We can also consider the map of $SO(3)$ into a Grassmann variety obtained by mapping a point of $SO(3)$ to a corresponding illumination space.

Thank You!