If Sherlock Holmes Were a Mathematician...

Elin Smith

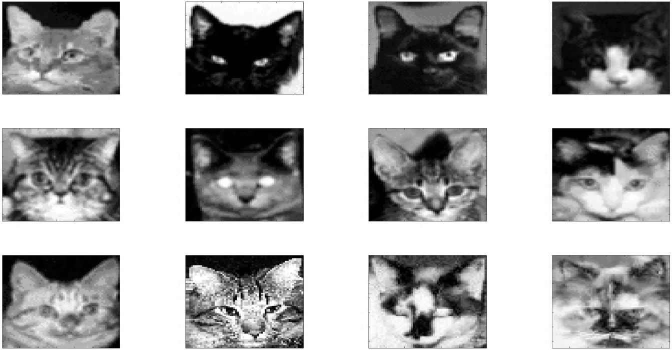
Colorado State University

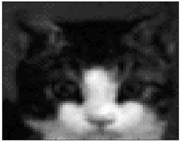
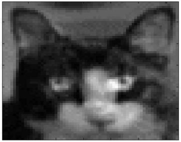
February 12, 2010

'Data! Data! Data!' he cried impatiently. 'I can't make bricks without clay.'

- (Sherlock Holmes) Sir Arthur Conan Doyle
The Adventure of the Copper Beeches

Data Set 1: 99 Images of Cat Faces



$\frac{1}{2}\cdot$  $+\ \frac{1}{2}\cdot$  $=$ 

Creating a vector from a matrix:

$$\begin{array}{|c|c|c|c|}\hline I_{11} & I_{12} & \cdots & I_{1\eta} \\ \hline I_{21} & I_{22} & \cdots & I_{2\eta} \\ \hline \vdots & \vdots & & \vdots \\ \hline I_{\eta 1} & I_{\eta 2} & \cdots & I_{\eta\eta} \\ \hline \end{array} \longmapsto \begin{array}{|c|}\hline I_{11} \\ \hline \vdots \\ \hline I_{1\eta} \\ \hline I_{21} \\ \hline \vdots \\ \hline I_{2\eta} \\ \hline \vdots \\ \hline I_{\eta 1} \\ \hline \vdots \\ \hline I_{\eta\eta} \\ \hline \end{array} = \mathbf{x}^{(i)}$$

Data matrix X:

$$\left(\begin{array}{c|c|c|c} \mathbf{x}^{(1)} & \mathbf{x}^{(2)} & \cdots & \mathbf{x}^{(n)} \end{array} \right),$$

where each $\mathbf{x}^{(i)} \in \mathbb{R}^m$, for m equal to the number of pixels per image.

Standard Spanning Set for Data Set X:

$$\begin{array}{|c|}\hline 1\\\hline 0\\\hline 0\\\hline \vdots\\\hline 0\\\hline\end{array}\;, \quad \begin{array}{|c|}\hline 0\\\hline 1\\\hline 0\\\hline \vdots\\\hline 0\\\hline\end{array}\;, \quad \cdots \;, \quad \begin{array}{|c|}\hline 0\\\hline 0\\\hline \vdots\\\hline 0\\\hline 1\\\hline\end{array}$$

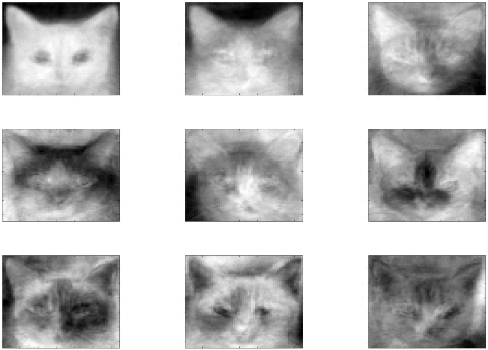
Standard Spanning Set for Data Set X:

$$\begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \ldots \quad, \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$$

A better set and a lower dimension?

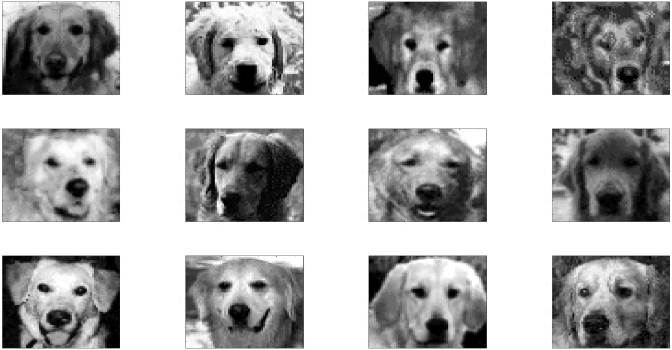
Using Principal Component Analysis, we get an ordered basis of 'eigenpictures':

Using Principal Component Analysis, we get an ordered basis of 'eigenpictures':



Only need 34 of these basis 'vectors' to capture most of the energy of the data set.

We can use this basis to compare data sets:
Data Set 2: 99 Images of Dog faces



How much of a given dog face looks like a cat?

How much of a given dog face looks like a cat?

$=$

$+$

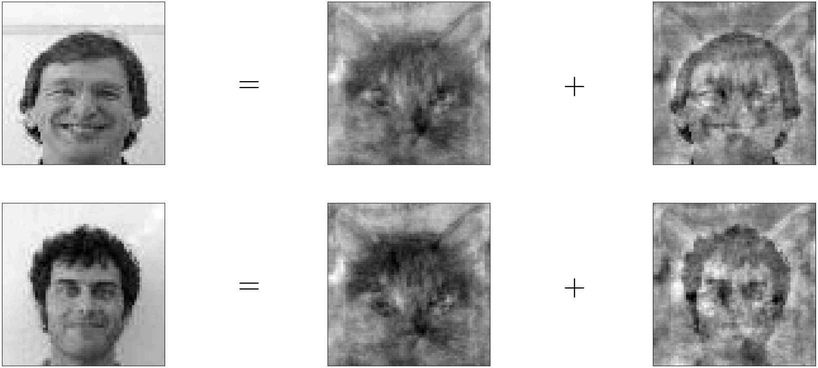
How much of a given dog face looks like a cat?

$=$

$+$

How much do Chris and Renzo look like a cat?

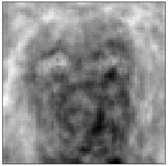
How much do Chris and Renzo look like a cat?

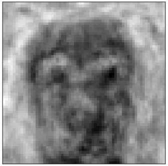


How much do Jorge and William look like a dog?

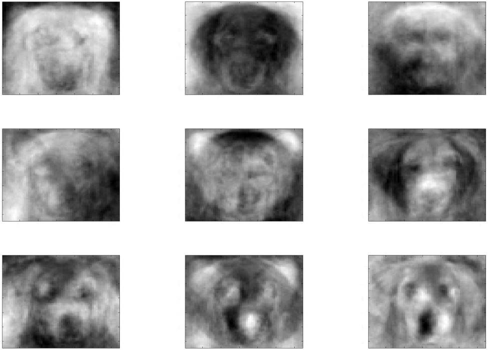
How much do Jorge and William look like a dog?

 $=$  $+$ 

 $=$  $+$ 

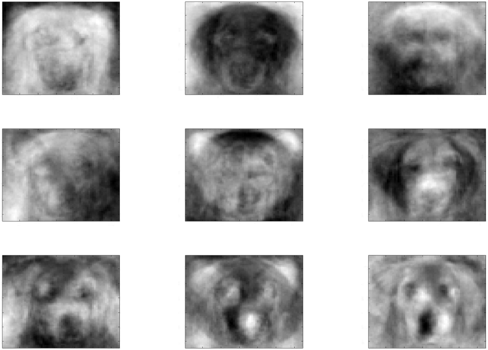
Classification?

Basis of eigenpictures for dogs:



Classification?

Basis of eigenpictures for dogs:



What is PCA and how do we find such a basis?

Theorem (Singular Value Decomposition)
Let $X \in \mathbb{R}^{m \times n}$, and let $\ell = \min\{m, n\}$. Then there exist real orthogonal matrices U and V such that

$$X = U \Sigma V^T,$$

where U has size $m \times m$, V has size $n \times n$, and Σ is an $m \times n$ diagonal matrix.

The entries of $\Sigma = \mathrm{diag}\left(\sigma^{(1)}, \ldots, \sigma^{(\ell)}\right)$ are ordered by $\sigma^{(1)} \geq \cdots \geq \sigma^{(\ell)} \geq 0$.

Given a matrix X, we define the **covariance matrix** C to be

$$C = XX^T.$$

Given a matrix X, we define the **covariance matrix** C to be

$$C = XX^T.$$

It can be shown that

$$XX^T\mathbf{u}^{(i)} = \left(\sigma^{(i)}\right)^2\mathbf{u}^{(i)}.$$

Given a matrix X, we define the **covariance matrix** C to be

$$C = XX^T.$$

It can be shown that

$$XX^T\mathbf{u}^{(i)} = \left(\sigma^{(i)}\right)^2 \mathbf{u}^{(i)}.$$

The left singular vectors of X are eigenvectors of the covariance matrix with eigenvalues equal to the singular values squared.

Let $\text{rank}(X) = r$. Then column i of X is given by

$$\mathbf{x}^{(i)} = \sum_{j=1}^{r} \sigma^{(j)} \mathbf{v}_i^{(j)} \mathbf{u}^{(j)}.$$

Let $\text{rank}(X) = r$. Then column i of X is given by

$$\mathbf{x}^{(i)} = \sum_{j=1}^{r} \sigma^{(j)} \mathbf{v}_i^{(j)} \mathbf{u}^{(j)}.$$

The set of the first r left singular vectors form a basis for the column space of the matrix X.

Principal Component Analysis:

Given a data set $\left(\begin{array}{c|c|c|c} \mathbf{x}^{(1)} & \mathbf{x}^{(2)} & \cdots & \mathbf{x}^{(n)} \end{array} \right)$,

find the optimal ordered orthonormal basis $\mathcal{B}$ for X.

What does it mean to be optimal?

Suppose our data set X is $n-$dimensional.
Let $\mathcal{B} = \left\{\phi^{(1)}, \ldots, \phi^{(n)}\right\}$ be an ordered orthonormal basis for X.

What does it mean to be optimal?

Suppose our data set X is $n-$dimensional.
Let $\mathcal{B} = \left\{\phi^{(1)}, \ldots, \phi^{(n)}\right\}$ be an ordered orthonormal basis for X.

Any point $\mathbf{x} \in X$ can be expressed in terms of this basis:
$\mathbf{x} = \alpha_1 \phi^{(1)} + \cdots + \alpha_n \phi^{(n)}$.

What does it mean to be optimal?

Suppose our data set X is $n-$dimensional.
Let $\mathcal{B} = \left\{\phi^{(1)}, \ldots, \phi^{(n)}\right\}$ be an ordered orthonormal basis for X.

Any point $\mathbf{x} \in X$ can be expressed in terms of this basis:
$\mathbf{x} = \alpha_1 \phi^{(1)} + \cdots + \alpha_n \phi^{(n)}$.

The $d-$term truncation of this expansion for $\mathbf{x}$ is given by
$\mathbf{x}_d = \alpha_1 \phi^{(1)} + \cdots + \alpha_d \phi^{(d)}$.

What does it mean to be optimal?

Suppose our data set X is $n-$dimensional.
Let $\mathcal{B} = \left\{\phi^{(1)}, \ldots, \phi^{(n)}\right\}$ be an ordered orthonormal basis for X.

Any point $\mathbf{x} \in X$ can be expressed in terms of this basis:
$\mathbf{x} = \alpha_1 \phi^{(1)} + \cdots + \alpha_n \phi^{(n)}$.

The $d-$term truncation of this expansion for $\mathbf{x}$ is given by
$\mathbf{x}_d = \alpha_1 \phi^{(1)} + \cdots + \alpha_d \phi^{(d)}$.

The mean squared truncation error is given by $\epsilon = \left\langle \|\mathbf{x} - \mathbf{x}_d\|^2 \right\rangle$.

What does it mean to be optimal?

Suppose our data set X is $n-$dimensional.
Let $\mathcal{B} = \left\{\phi^{(1)}, \ldots, \phi^{(n)}\right\}$ be an ordered orthonormal basis for X.

Any point $\mathbf{x} \in X$ can be expressed in terms of this basis:
$\mathbf{x} = \alpha_1 \phi^{(1)} + \cdots + \alpha_n \phi^{(n)}$.

The $d-$term truncation of this expansion for $\mathbf{x}$ is given by
$\mathbf{x}_d = \alpha_1 \phi^{(1)} + \cdots + \alpha_d \phi^{(d)}$.

The mean squared truncation error is given by $\epsilon = \left\langle \|\mathbf{x} - \mathbf{x}_d\|^2 \right\rangle$.

We define $\mathcal{B}$ to be an optimal basis if, among all orthonormal bases, $\mathcal{B}$ minimizes the mean squared truncation error, ϵ.

Properties of the Principal Component Basis:

The basis can be computed by finding the left singular vectors of X.

Properties of the Principal Component Basis:

The basis can be computed by finding the left singular vectors of X. (Hence the term eigenpictures!)

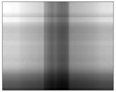
Properties of the Principal Component Basis:
The basis can be computed by finding the left singular vectors of X. (Hence the term eigenpictures!)

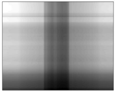
Any truncation of $\mathcal{B}$ captures more statistical variance than any other basis of the same dimension.

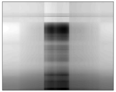
Properties of the Principal Component Basis:
The basis can be computed by finding the left singular vectors of X. (Hence the term eigenpictures!)

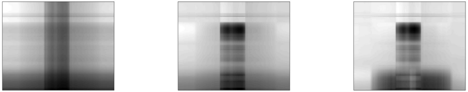
Any truncation of $\mathcal{B}$ captures more statistical variance than any other basis of the same dimension.

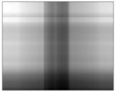
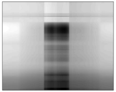
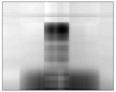
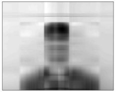
A reconstruction of X using any $d-$term truncation of $\mathcal{B}$ gives the best rank d approximation of X.

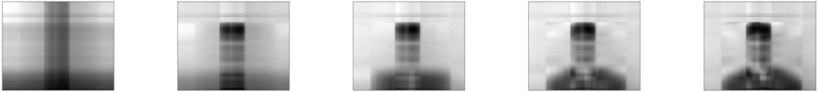


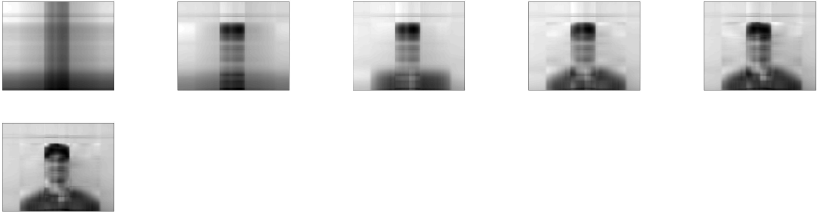


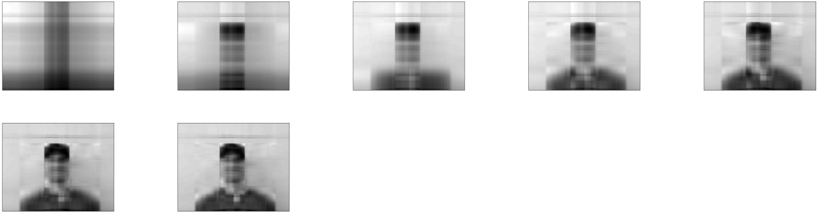


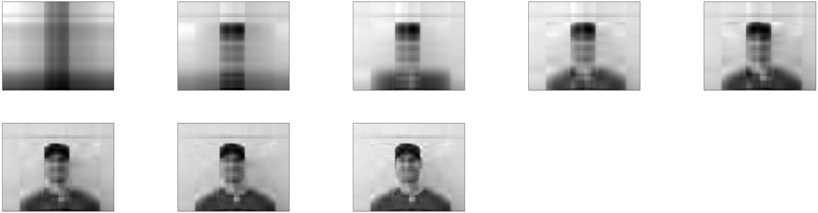


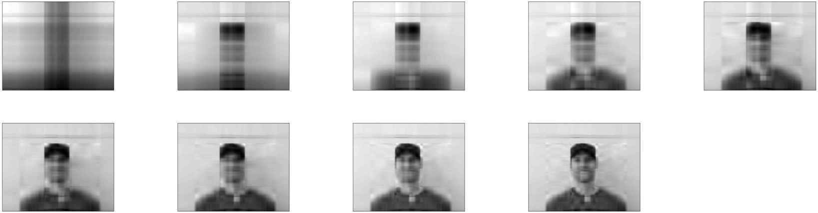





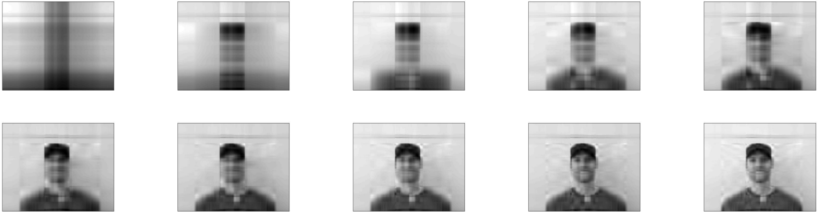


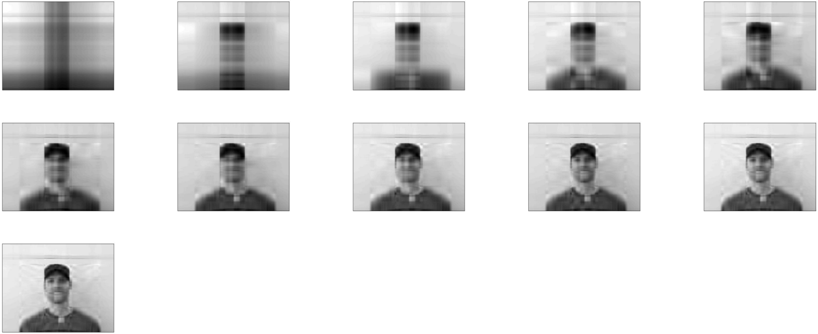


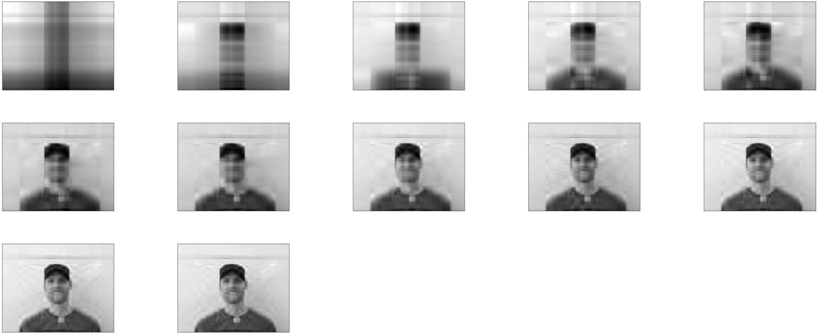


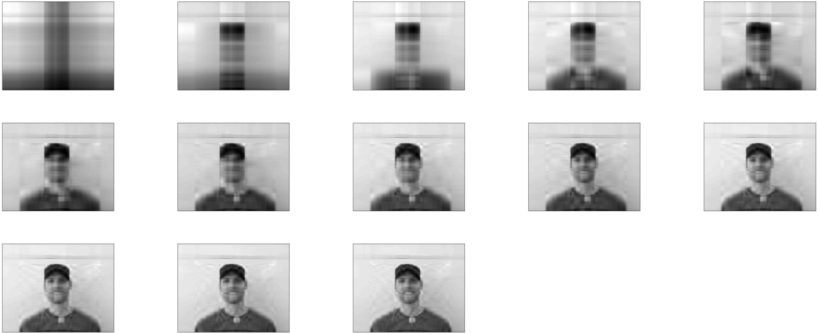


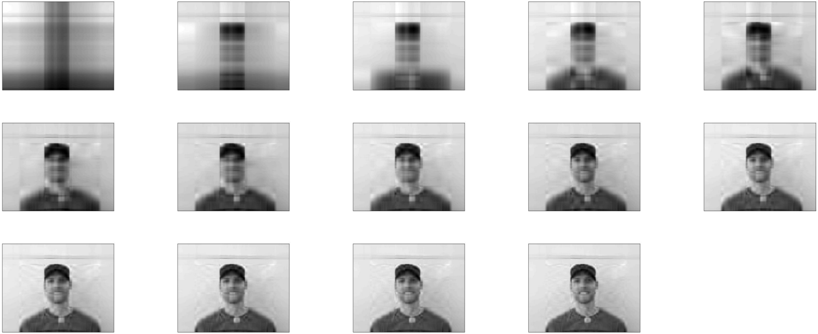


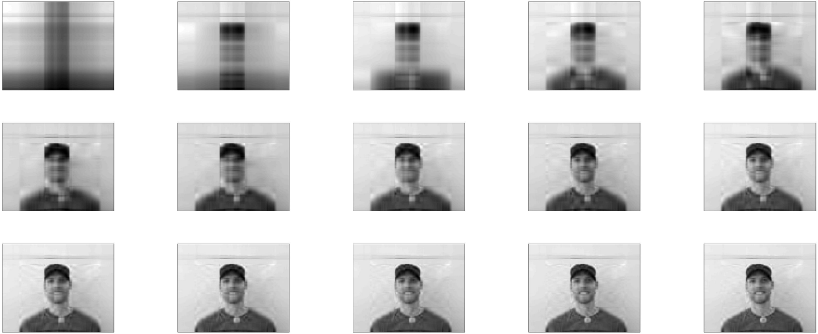


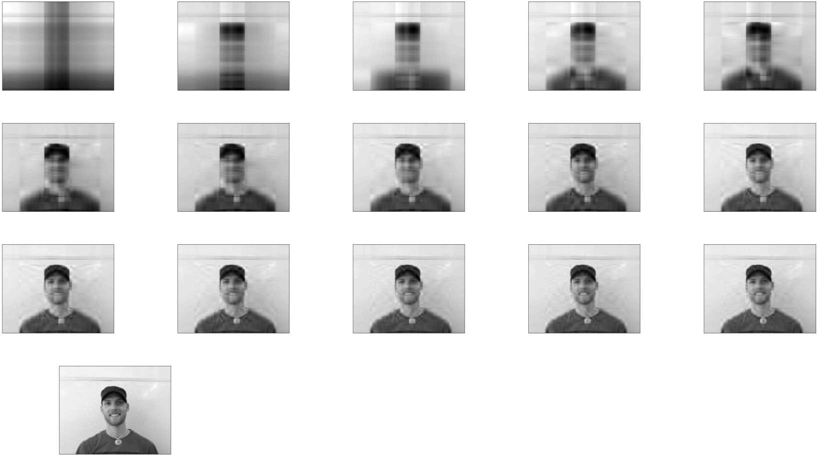


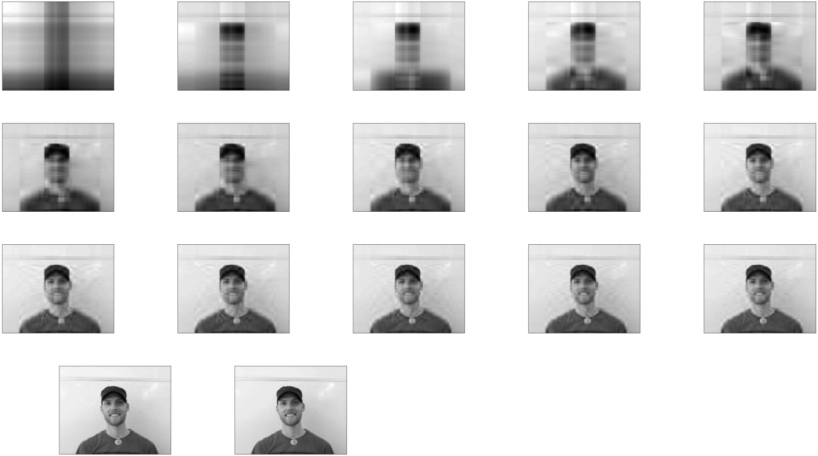












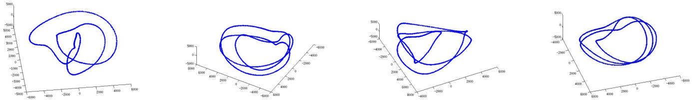
Data Set 3: Images of a Rotating Object

$$
\begin{aligned}
X &= U\Sigma V^T \\
\hat{U} &= \left(\begin{array}{c|c|c} \mathbf{u}^{(1)} & \mathbf{u}^{(2)} & \mathbf{u}^{(3)} \end{array} \right)
\end{aligned}
$$

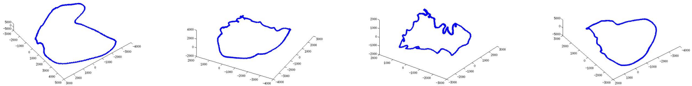
Projection:

$$
\begin{aligned}
\pi : \mathbb{R}^{262,144\times 1500} &\longrightarrow \mathbb{R}^{3\times 1500} \\
X &\longmapsto \hat{U}^T X.
\end{aligned}
$$

Volleyball: Red, Green, Blue, and Gray

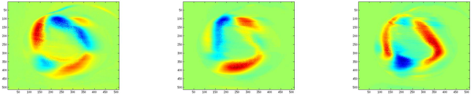


Jack O' Lantern: Red, Green, Blue, and Gray

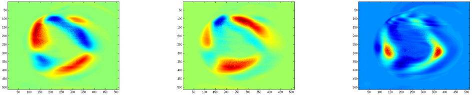


Volleyball Eigenpictures 1, 2, and 3:

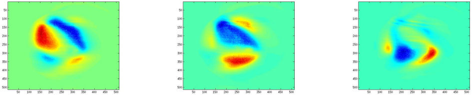
Red Filter



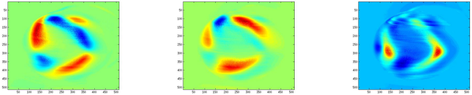
Green Filter



Blue Filter

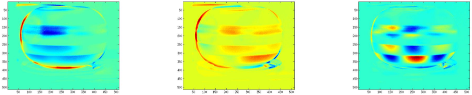


Grayscale

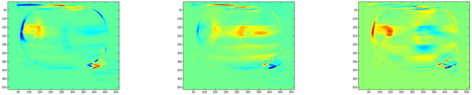


Jack o' Lantern Eigenpictures 1, 2, and 3:

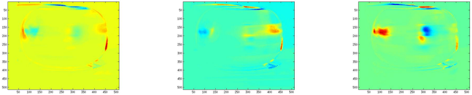
Red Filter



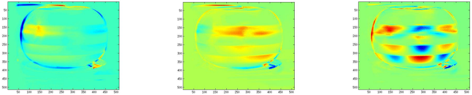
Green Filter



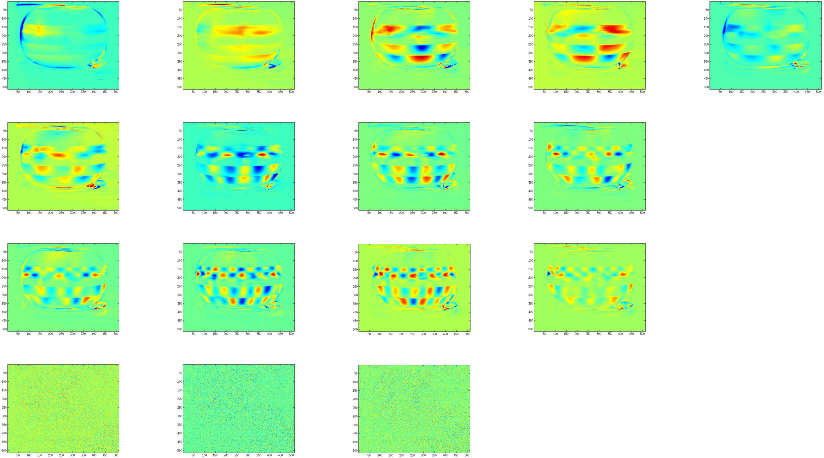
Blue Filter



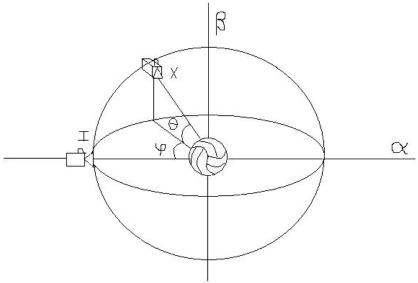
Grayscale



Jack o' Lantern Eigenpictures 1-13,1000-1002: Grayscale



Data Set 4: Capturing the Illumination Space of an Object



To each point of the sphere, we attach a vector space called the illumination space.

Consider all possible illuminations of a person.

Consider all possible illuminations of a person.

The vector space which captures the majority of the energy of this set is called the **illumination space.**

Determining Optimal Camera Distribution:
Fix a number of camera locations n and fix the camera resolution. What is the camera location distribution that optimizes capture of variance in illumination space?

Determining Optimal Camera Distribution:
Fix a number of camera locations n and fix the camera resolution. What is the camera location distribution that optimizes capture of variance in illumination space?

That is, we wish to find an optimal distribution of points with respect to some measure.

Consider the problem of finding an optimal distribution of points on the sphere based on a potential energy function.

Consider the problem of finding an optimal distribution of points on the sphere based on a potential energy function.

Let f be a decreasing and continuous function on $[0, \infty)$, and let C be a finite set of points on S^{n-1}. The **potential energy of** C is defined to be

$$\sum_{x,y \in C, x \neq y} f(|x-y|^2).$$

Consider the problem of finding an optimal distribution of points on the sphere based on a potential energy function.

Let f be a decreasing and continuous function on $[0, \infty)$, and let C be a finite set of points on S^{n-1}. The **potential energy of** C is defined to be

$$\sum_{x,y \in C, x \neq y} f(|x-y|^2).$$

Goal: Minimize the potential energy of C.

One way to find an optimal distribution of points is to start with a random configuration and let the points push away from each other.

(Loading movie...)

One way to find an optimal distribution of points is to start with a random configuration and let the points push away from each other.

(Loading movie...)

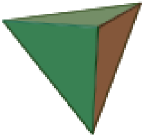
Frequently, this will result in a point configuration which gives a local minimum of our potential energy function.

Under what conditions can we guarantee a global minimum of our potential function?

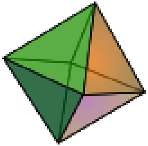
A sufficient condition for a local energy minimizer to be a global minimizer for potential energy was obtained by Cohn and Kumar.

Theorem (Cohn, Kumar 2007):
For any completely monotonic potential function, sharp configurations are global potential energy minimizers.

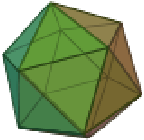
Examples of sharp configurations:



Tetrahedron



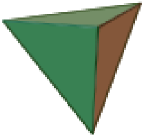
Octahedron

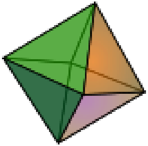


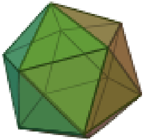
Icosahedron

http://en.wikipedia.org/wiki/Platonic_solid

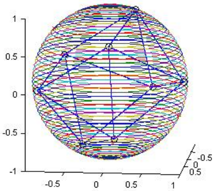
Examples of sharp configurations:


Tetrahedron


Octahedron

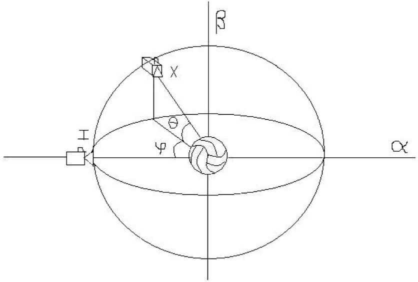

Icosahedron

http://en.wikipedia.org/wiki/Platonic_solid



Square Antiprism

Return to



.

We wish to let points push away from each other on the sphere, this time driven by 'closeness' of illumination spaces.

Given two subspaces $\mathcal{L}$ and $\mathcal{M}$ of $\mathbb{R}^n$, with $\dim(\mathcal{L}) = \ell \leq \dim(\mathcal{M}) = m$, what is the distance between $\mathcal{L}$ and $\mathcal{M}$?

Given two subspaces $\mathcal{L}$ and $\mathcal{M}$ of $\mathbb{R}^n$, with $\dim(\mathcal{L}) = \ell \leq \dim(\mathcal{M}) = m$, what is the distance between $\mathcal{L}$ and $\mathcal{M}$?

Let $Q_{\mathcal{L}}$ and $Q_{\mathcal{M}}$ be orthonormal bases for $\mathcal{L}$ and $\mathcal{M}$, respectively. The principal angles $\theta_1, \ldots, \theta_\ell$ between $\mathcal{L}$ and $\mathcal{M}$ are given by

$$\cos(\theta_i) = \sigma_i.$$

where the σ_i are the singular values of $Q_{\mathcal{M}}^T Q_{\mathcal{L}}$.

Some example metrics on the Grassmannian:

- $\sqrt{\sum_{i=1}^{\ell} \theta_i^2}$
- $\cos^{-1}\left(\prod_{i=1}^{\ell} \cos(\theta_i)\right)$
- $\sqrt{\sum_{i=1}^{\ell} \sin(\theta_i)^2}$.

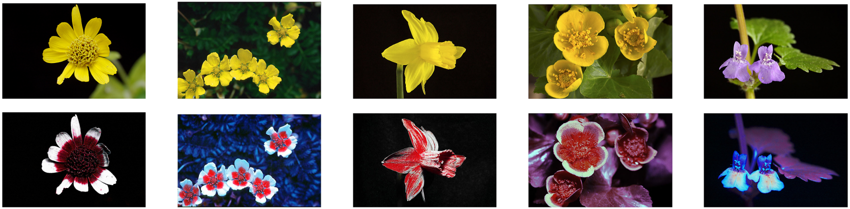
- We have a way of using a potential energy function to push points away from each other.
- We can quickly calculate principal angles between any two illumination spaces.
- We can therefore efficiently determine the distance between any two illumination spaces.
- Goal: Generalize the theorem of Cohn and Kumar's to this setting to determine if a given configuration yields a global minimum.

Variations in illumination spaces:

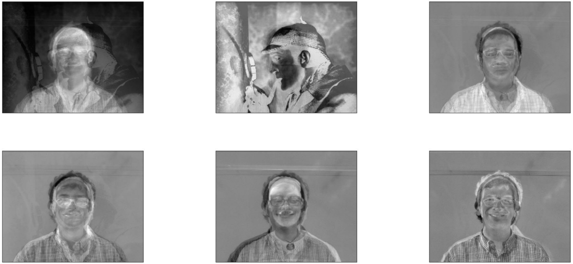


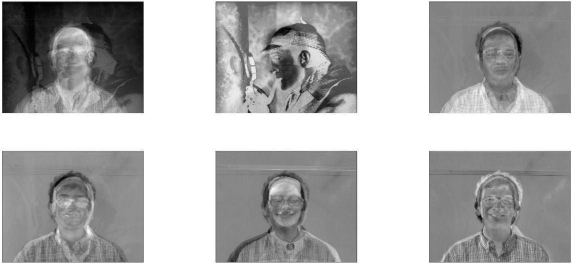


Variations in illumination spaces:



Bjørn Rørslett
www.naturfotograf.com





Thank You!