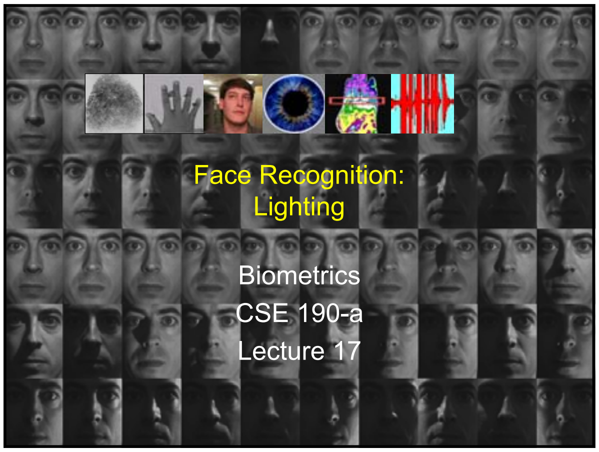


Why is Face Recognition Hard?

Fisherfaces: Class specific linear projection

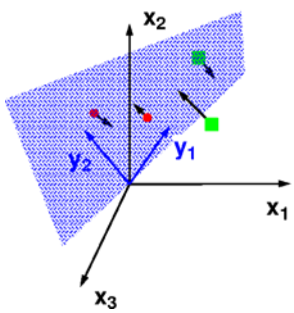
P. Belhumeur, J. Hespanha, D. Kriegman, *Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection*, PAMI, July 1997, pp. 711--720.

- An n-pixel image $\mathbf{x} \in \mathbf{R}^n$ can be projected to a low-dimensional feature space $\mathbf{y} \in \mathbf{R}^m$ by

$$\mathbf{y} = W\mathbf{x}$$

where W is an n by m matrix.

- Recognition is performed using nearest neighbor in $\mathbf{R}^m$.
- How do we choose a good W?



PCA & Fisher's Linear Discriminant

- Between-class scatter

$$S_B = \sum_{i=1}^{c} |\chi_i| (\mu_i - \mu)(\mu_i - \mu)^T$$

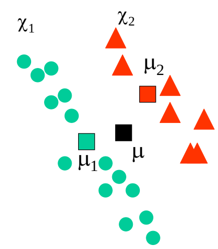
- Within-class scatter

$$S_W = \sum_{i=1}^{c} \sum_{x_k \in \chi_i} (x_k - \mu_i)(x_k - \mu_i)^T$$

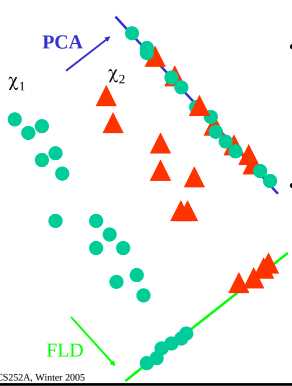
- Total scatter

$$S_T = \sum_{i=1}^{c} \sum_{x_k \in \chi_i} (x_k - \mu)(x_k - \mu)^T = S_B + S_W$$

- Where
  - c is the number of classes
  - μ_i is the mean of class χ_i
  - $|\chi_i|$ is number of samples of χ_i.



PCA & Fisher's Linear Discriminant



- PCA (Eigenfaces)

$$W_{PCA} = \arg\max_W \left| W^T S_T W \right|$$

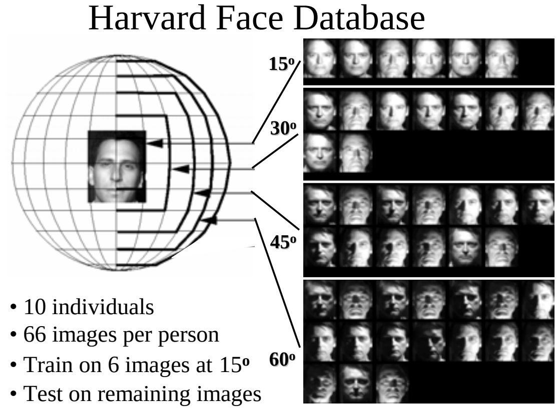
Maximizes projected total scatter

- Fisher's Linear Discriminant

$$W_{fld} = \arg\max_W \frac{\left| W^T S_B W \right|}{\left| W^T S_W W \right|}$$

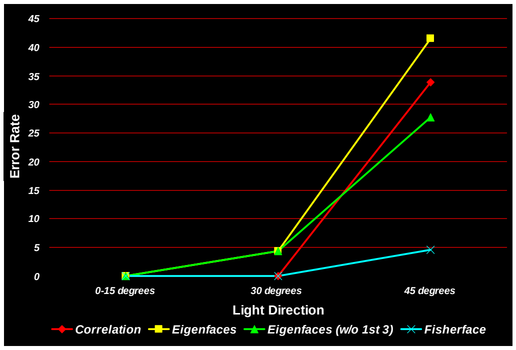
Maximizes ratio of projected between-class to projected within-class scatter

Harvard Face Database



- 10 individuals
- 66 images per person
- Train on 6 images at 15º
- Test on remaining images

Recognition Results: Lighting Extrapolation



(2D) Model-based

– Active Appearance Model

- Model Construction (linear)

labeled image landmarks shape-free texture

Active Appearance Model (AAM)

- Shape model

$$s = (x_1, y_1, \ldots, x_n, y_n)^T \qquad s = \bar{s} + P_s b_s$$

- Appearance model

$$g = (I_1, \ldots, I_m)^T \qquad g = \bar{g} + P_g b_g$$

- Combined model

$$b = \begin{pmatrix} W_s b_s \\ b_g \end{pmatrix} = \begin{pmatrix} W_s P_s^T (s - \bar{s}) \\ P_g^T (g - \bar{g}) \end{pmatrix} \xrightarrow{\text{PCA}} b = Qc$$

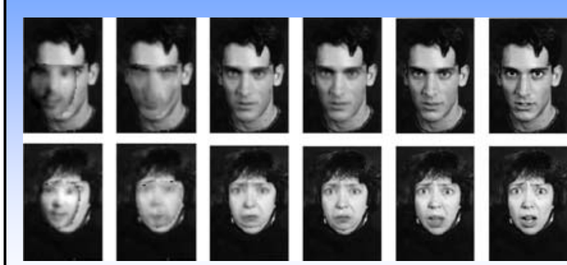
AAM

- Model fitting
  - Minimize the objective function (gray level difference between the given image and the stored model

$$\Delta = |\delta I|^2$$

  - Searching by learning
    - Annotated model (true model parameters)
    - Relation: known model displacements ↔ observed difference vector
    - Use multivariate multiple regression to learn the relation and predict the displacement during searching

AAM



Initial 3 its 8 its 11 its Converged Original

The challenge caused by lighting variability

Same Person

or

Different People

Same Person
or
Different People

Same Person
or
Different People

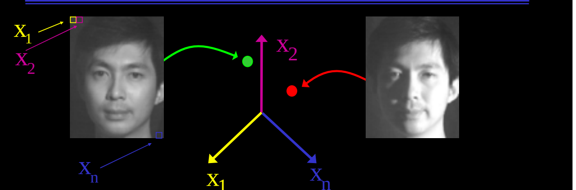
Illumination & Image Set

- Lack of illumination invariants
  [Chen, Jacobs, Belhumeur 98]
- Set of images of Lambertian surface w/o shadowing is 3-D linear subspace
  [Moses 93], [Nayar, Murase 96], [Shashua 97]
- Empirical evidence that set of images of object is well-approximated by a low-dimensional linear subspace
  [Hallinan 94], [Epstein, Hallinan, Yuille 95]
- Illumination cones
  - [Belhumeur, Kriegman 98]
- Spherical harmonics lighting & images
  - [Basri, Jacobs 01], [Ramamoorthi, Hanrahan 01]
- Analytic PCA of image over lighting
  - [Ramamoorthi 02]

Some Background Issues

1. What is the image space?
2. What is lighting?
3. What is reflectance?

The Space of Images

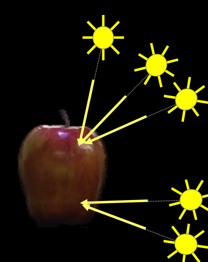


- Consider an n-pixel image to be a point in an n-dimensional space, $\mathbf{x} \in \mathbf{R}^n$.
- Each pixel value is a coordinate of **x**.
- Many results will apply to linear transformations of image space (e.g. filtered images)
- Other image representations (e.g. Cayley-Klein spaces, See Koenderink's "pixel f#@king paper")

Lighting

Generally, arbitrary lighting can be viewed as a non-negative function on a 4-D space.

Typically make limiting assumptions



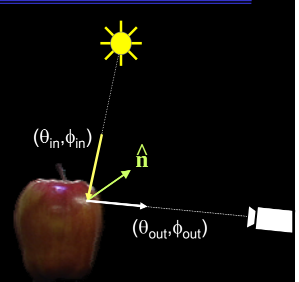
- Distant lighting (non-negative function on sphere)
- Point light sources (delta function)
- Diffuse lighting (constant function)

BRDF

Bi-directional Reflectance Distribution Function

$\rho(\theta_{in}, \phi_{in} ; \theta_{out}, \phi_{out})$

- Ratio of incident irradiance to emitted radiance
- Function of
  - Incoming light direction: θ_{in} , ϕ_{in}
  - Outgoing light direction: θ_{out} , ϕ_{out}



$$\rho(\underline{x};\theta_{in},\phi_{in};\theta_{out},\phi_{out}) = \frac{L_o(\underline{x};\theta_{out},\phi_{out})}{L_i(\underline{x};\theta_{in},\phi_{in})\cos\phi_{in}\,d\omega}$$

The Illumination Invariance Question

Let $I(O,L)$ be the image of object O under lighting L.

Does there exist a recognition function $c(I_1,I_2)$ of two images I_1 and I_2 such that

$$\forall O, L_1, L_2 \quad C(I(O,L_1), I(O, L_2)) = 0$$

and

$$\exists O_i, O_j, L_1, L_2 \quad C(I(O_i ,L_1), I(O_j, L_2)) \neq 0$$

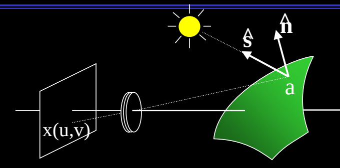
NO!

Assumptions

- Lambertian reflectance functions.
- Objects have convex shape.
- Light sources at infinity.
- Orthographic projection.

- Note: many of these can be relaxed….

Lambertian Surface



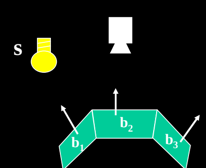
At image location (u,v), the intensity of a pixel x(u,v) is:

$$x(u,v) = [a(u,v)\,\hat{n}(u,v)] \bullet [s_0\hat{s}\,] = \mathbf{b}(u,v) \bullet \mathbf{s}$$

where

- a(u,v) is the albedo of the surface projecting to (u,v).
- **n**(u,v) is the direction of the surface normal.
- s_0 is the light source intensity.
- **s** is the direction to the light source.

Model for Image Formation



Lambertian Assumption with shadowing:

$$\mathbf{x} = \max(\mathbf{B}\,\mathbf{s}, 0) \qquad \mathbf{B} = \begin{bmatrix} -\,\mathbf{b}^T_1\,- \\ -\,\mathbf{b}^T_2\,- \\ \ldots\ldots \\ -\,\mathbf{b}^T_n\,- \end{bmatrix}_{n \times 3}$$

where

- **x** is an *n*-pixel image vector
- **B** is a matrix whose rows are unit normals scaled by the albedos
- $\mathbf{s} \in \mathbf{R}^3$ is vector of the light source direction scaled by intensity

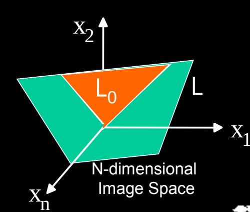
3-D Linear subspace

The set of images of a Lambertian surface with no shadowing is a subset of 3-D linear subspace.

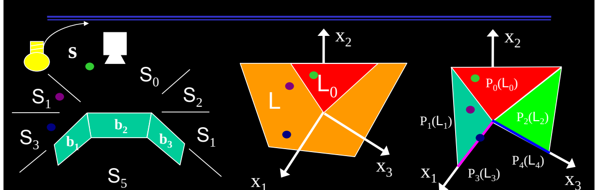
[Moses 93], [Nayar, Murase 96], [Shashua 97]

$$L = \{\mathbf{x} \mid \mathbf{x} = \mathbf{B}\mathbf{s},\ \forall \mathbf{s} \in R^3\}$$

where B is a *n* by 3 matrix whose rows are product of the surface normal and Lambertian albedo



Set of Images from a Single Light Source



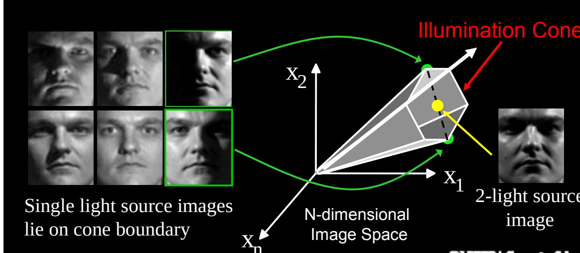
- Let L_i denote the intersection of L with an orthant *i* of R^n.
- Let $P_i(L_i)$ be the projection of L_i onto a "wall" of the positive orthant given by max(**x**, **0**).

Then, the set of images of an object produced by a single light source is:

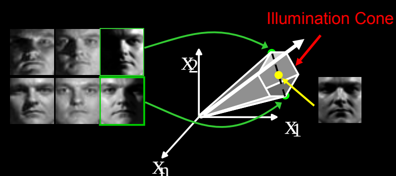
$$U = \bigcup_{i=0}^{M} P_i(L_i)$$

The Illumination Cone

Theorem: *The set of images of any object in fixed posed, but under all lighting conditions, is a* **convex cone** *in the image space.* (Belhumeur and Kriegman, IJCV, 98)



Some natural ideas & questions

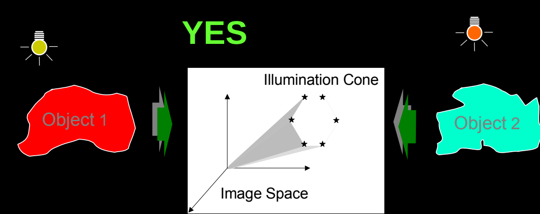


- Recogntion: Is a test image within an object's cone?
- Can the cones of two different objects intersect?
- Can two different objects have the same cone?
- How big is the cone?

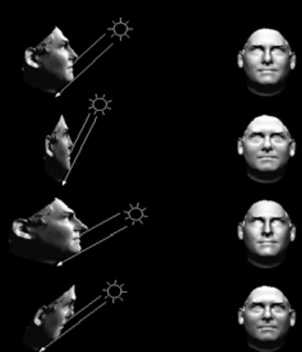
Do Ambiguities Exist?

Can two objects of differing shapes produce the same illumination cone?

YES



Generalized Bas-Relief Transformations



Objects differing by a GBR have the same illumination cone.

Without knowledge of light source location, one can only recover surfaces up to GBR transformations.

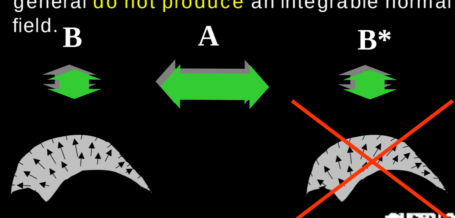
Do Ambiguities Exist? Yes

- Cone is determined by linear subspace **L**
- The columns of **B** span **L**
- For any **A** ∈ **GL(3), B* = BA** also spans **L** , **i.e. X = B*S* = B AA^{-1}S**
- If **B*** corresponds to integrable vector field, then then there is a surface having the same cone.

Surface Integrability

In general, **B*** does not have a corresponding surface.

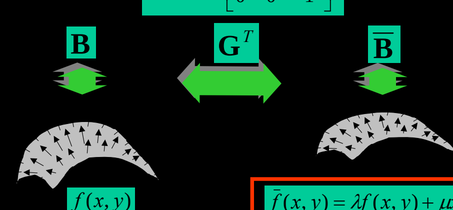
Linear transformations of the surface normals in general do not produce an integrable normal field.



GBR Transformation

Only Generalized Bas-Relief transformations satisfy the integrability constraint:

$$\mathbf{A} = \mathbf{G}^T = \begin{bmatrix} \lambda & 0 & -\mu \\ 0 & \lambda & -\nu \\ 0 & 0 & 1 \end{bmatrix}^T$$



$$\bar{f}(x,y) = \lambda f(x,y) + \mu x + \nu y$$

What about cast shadows for nonconvex objects?

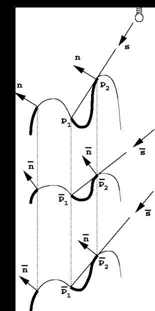
P.P. Reubens in Opticorum Libri Sex, 1613

GBR Preserves Shadows

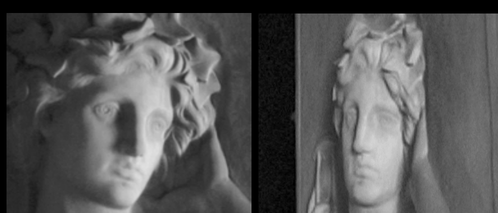
Given a surface *f* and a GBR transformed surface *f'* then for every light source **s** which illuminates *f* there exists a light source **s'** which illuminates *f'* such that the **attached** and **cast shadows** are identical.

GBR is the **only** transform that preserves shadows.

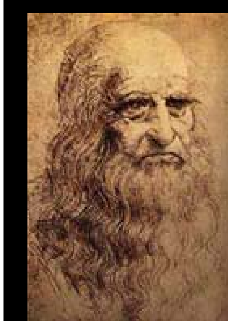
[Kriegman, Belhumeur 2001]



Bas-Relief Sculpture



Codex Urbinas

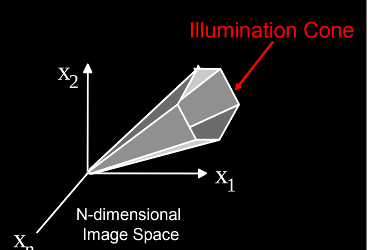


As far as light and shade are concerned low relief fails both as sculpture and as painting, because *the shadows correspond to the low nature of the relief,* ***as for example in the shadows of foreshortened objects, which will not exhibit the depth of those in painting or in sculpture in the round.***

Leonardo da Vinci
Treatise on Painting (Kemp)

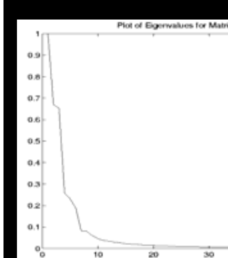
The Illumination Cone

Thm: The span of the extreme rays of the illumination cone is equal to the number of distinct surface normals – i.e., as high as n. (Belhumeur and Kriegman, IJCV,'98)



Shape of the Illumination Cone

Observation: The illumination cone is **flat** with most of its volume concentrated near a **low-dimensional** linear subspace



	Ball	Face	Phone	Parrot
#1	48.2	53.7	67.9	42.8
#3	94.4	90.2	88.2	76.3
#5	97.9	93.5	94.1	84.7
#7	99.1	95.3	96.3	88.5
#9	99.5	96.3	97.2	90.7

[Epstein, Hallinan, Yuille 95]

Dimension: 5 ± 2D

Expressing lighting with spherical harmonics

(Basri, Jacobs'01; Ramamoorthi, Hanrahan'01)

- Express lighting as a function on the sphere in terms of basis functions $Y_{lm}(\theta,\Phi)$

$$f(\theta,\phi)=\sum_{l=0}^{\infty}\sum_{m=-l}^{l} f_{l,m}Y_{l,m}(\theta,\phi)$$

- Spherical Harmonics: A set of orthonormal basis functions defined on the unit sphere.

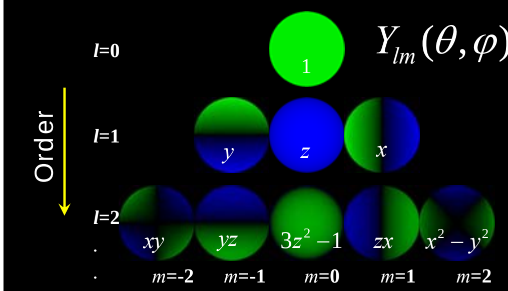
$$Y_{lm}(\theta,\phi)=N_{lm}P_l^{|m|}(\cos(\theta))e^{im\phi}$$

l, m are indices, $l \geq 0$ and $-l \leq m \leq l$

N_{lm} : normalization factor

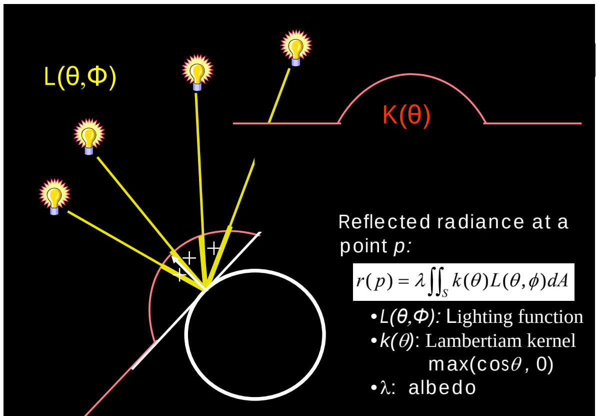
$P_l^{|m|}$: Legendre functions

Spherical Harmonics I



(Borrowed from: Ramamoorthi, Hanrahan, SIGGRAPH'01)

Green: Positive
Blue: Negative



Reflected radiance at a point *p*:

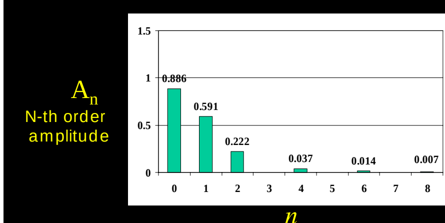
$$r(p)=\lambda\iint_S k(\theta)L(\theta,\phi)dA$$

- *L(θ,Φ)*: Lighting function
- *k(θ)*: Lambertian kernel max(cos*θ*, 0)
- λ: albedo

[Taken from David Jacobs]

Lambertian Kernel

- Express Kernel in spherical Harmonics
- Funk-Hecke convolution theorem
  - Convolution can be computed as product of coefficients of spherical Harmonics basis functions



Reflectance functions near low-dimensional linear subspace

$$r=k*l=\sum_{n=0}^{\infty}\sum_{m=-n}^{n}(K_{nm}L_{nm})h_{nm}$$

$$\approx\sum_{n=0}^{2}\sum_{m=-n}^{n}(K_{nm}L_{nm})h_{nm}$$

Yields 9D linear subspace.

Harmonic Images

- Images formed under lighting conditions specified by the first nine spherical harmonics
- Basis for a 9-D subspace

How accurate is approximation?

- Accuracy depends on lighting.
- For point source: 9D space captures 99.2% of variance
- For *any* lighting: 9D space captures >98% of variance.

Analytic PCA

[Ramamoorthi 02]

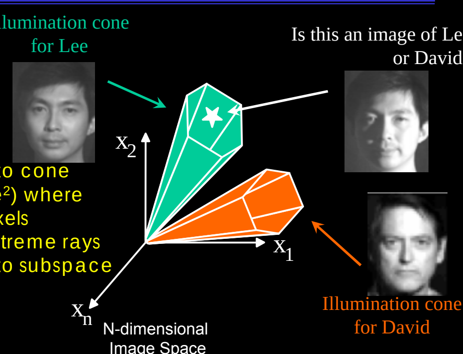
- Convex, Lambertian surface, attached shadows.
- Analytic construction of covariance matrix given normals, albedo
- Lighting probability distribution– single point source whose location is uniformly distributed over the sphere
- Constructed in terms of spherical harmonic basis functions

Looks Like Hallinan

Results

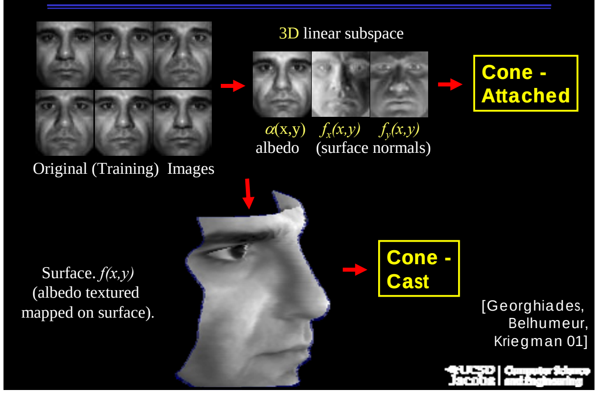
Object	# of Eigenvectors	Empirical Epstein et al	Analytic PCA
Sphere/ Basketball	3	94	91
Sphere/ Basketball	5	98	96
Face	3	90	91
Face	5	94	97

Illumination Cones: Recognition Method



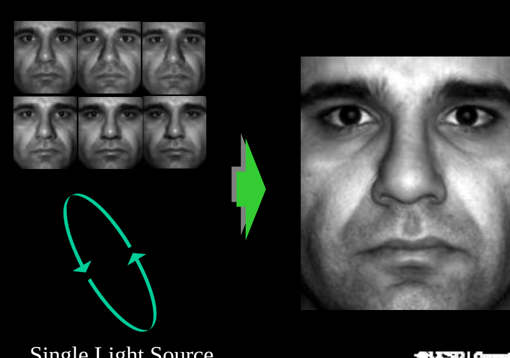
- Distance to cone
- Cost O(ne^2) where
  - n: # pixels
  - e: # extreme rays
- Distance to subspace

Generating the Illumination Cone



[Georghiades, Belhumeur, Kriegman 01]

Predicting Lighting Variation



Face databases

- FERET -14,000 images, 1200 people
- Yale Face Database B (and C)
  - Pose and illumination 10 (38) people
  - 5,760 images (21,888 images)
- PIE Database
  - Pose, Illumination, expression
  - 69 people, 23 lights,
- FRVT 2002 (Facial Recognition Vendor Test)
  - > 140,000 images – not yet available

Yale Face Database B

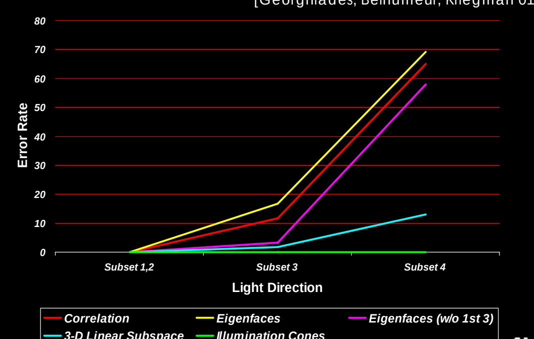
64 Lighting Conditions
9 Poses
=> 576 Images per Person

Face Recognition: Test Subsets

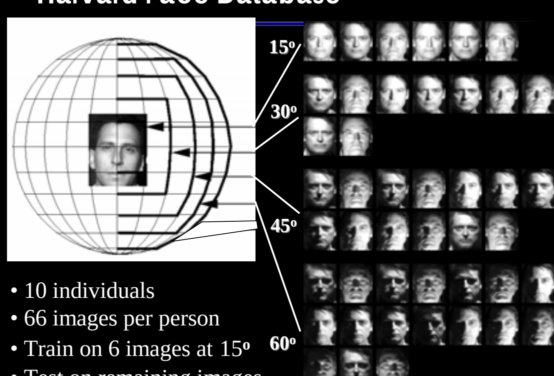
Test images divided into 4 subsets depending on illumination.

Geodesic Dome Database - Frontal Pose

[Georghiades, Belhumeur, Kriegman 01]



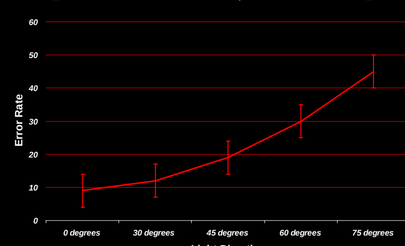
Harvard Face Database



- 10 individuals
- 66 images per person
- Train on 6 images at 15°
- Test on remaining images

Human Performance

[Colleen Jackson, Mike Tarr]



- 20 Subjects trained on subset 0.