Illumination Cones and Uncalibrated Photometric Stereo

Computer Vision I
CSE252A
Lecture 9
HW 2 Assigned: Warping and Mosaic
Vignetting correction
What is the set of images of an object under all possible lighting conditions?

In answering this question, we’ll arrive at a method for reconstructing surface shape with unknown lighting.
The Space of Images

Consider an n-pixel image to be a point in an n-dimensional space, \( \mathbf{x} \in \mathbb{R}^n \).
Each pixel value is a coordinate of \( \mathbf{x} \).
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”)
Assumptions

For discussion, we assume:

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

• Note: many of these can be relaxed....
At image location \((u,v)\), the intensity of a pixel \(x(u,v)\) is:

\[
x(u,v) = a(u,v) \cdot n(u,v) \cdot s_0 \cdot s
\]

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.
Lambertian Assumption with shadowing:

\[ x = \text{max}(B s, 0) \]

where

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

\[ B = \begin{bmatrix} -b_{1}^T & -b_{2}^T & \ldots & -b_{n}^T \end{bmatrix} \] \( n \times 3 \)
The set of images of a Lambertian surface with no shadowing is a subset of 3-D linear subspace.

\[ L = \{ x \mid x = Bs, \forall s \in \mathbb{R}^3 \} \]

where B is a \( n \) by 3 matrix whose rows are product of the surface normal and Lambertian albedo.
How do you construct subspace?

\[
\begin{bmatrix}
\mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3
\end{bmatrix}
\xrightarrow{\text{Any three images w/o shadows taken under different lighting span } L}
\begin{bmatrix}
\mathbf{B}
\end{bmatrix}
\]

- Any three images w/o shadows taken under different lighting span L
- Not orthogonal
- Orthogonalize with Gram-Schmidt
How do you construct subspace?

\[
\begin{bmatrix}
\end{bmatrix}
\rightarrow
\begin{bmatrix}
\end{bmatrix}
\]

\[
\begin{bmatrix}
\mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 & \mathbf{x}_4 & \mathbf{x}_5 \\
\end{bmatrix}
\rightarrow
\mathbf{B}
\]

With more than three images, perform least squares estimation of \( \mathbf{B} \) using Singular Value Decomposition (SVD)
Matrix Decompositions

- Definition: The factorization of a matrix $M$ into two or more matrices $M_1, M_2, \ldots, M_n$, such that $M = M_1M_2\ldots M_n$.

- Many decompositions exist...
  - **QR Decomposition**
  - **LU Decomposition**
  - **LDU Decomposition**
  - Etc.
Singular Value Decomposition

Excellent ref: ‘Matrix Computations,” Golub, Van Loan

- Any m by n matrix $A$ may be factored such that
  $$A = U \Sigma V^T$$
  $$[m \times n] = [m \times m][m \times n][n \times n]$$

- $U$: m by m, orthogonal matrix
  - Columns of $U$ are the eigenvectors of $AA^T$

- $V$: n by n, orthogonal matrix,
  - columns are the eigenvectors of $A^TA$

- $\Sigma$: m by n, diagonal with non-negative entries ($\sigma_1$, $\sigma_2$, ..., $\sigma_s$) with $s = \min(m,n)$ are called the called the singular values
  - Singular values are the square roots of eigenvalues of both $AA^T$ and $A^TA$
  - Result of SVD algorithm: $\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_s$
SVD Properties

• In Matlab `[u s v] = svd(A), and you can verify that: A = u*s*v'` 

• `r = Rank(A) = # of non-zero singular values.` 
• `U, V` give us orthonormal bases for the subspaces of `A`:
  - `1st r columns of U`: Column space of `A`
  - `Last m - r columns of U`: Left nullspace of `A`
  - `1st r columns of V`: Row space of `A`
  - `Last n - r columns of V`: Nullspace of `A`

• For `d <= r`, the first `d` column of `U` provide the best `d`-dimensional basis for columns of `A` in least squares sense.
Any m by n matrix $A$ may be factored such that
\[ A = U \Sigma V^T \]
\[[m \times n] = [m \times n][n \times n][n \times n]\]

If $m > n$, then one can view $\Sigma$ as:

Where $\Sigma' = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_s)$ with $s = \min(m, n)$, and lower matrix is $(n-m \times m)$ of zeros.

Alternatively, you can write:
\[ A = U' \Sigma' V^T \]

In Matlab, thin SVD is: $[U \; S \; V] = \text{svds}(A)$
Application: Pseudoinverse

- Given $y = Ax$, $x = A^+ y$
- For square $A$, $A^+ = A^{-1}$
- For any $A$...
  \[
  A^+ = V \Sigma^{-1} U^T
  \]
- $A^+$ is called the pseudoinverse of $A$.
- $x = A^+ y$ is the least-squares solution of $y = Ax$.
- Alternative to previous solution.
Estimating B with SVD

1. Construct data matrix
   \[ D = \begin{bmatrix} x_1 & x_2 & x_3 & \ldots & x_n \end{bmatrix} \]

2. \[ [U \ S \ V] = \text{svds}(D) \]
   - If data had no noise, then \( \text{rank}(D) = 3 \), and the first three singular values \( S \) would be positive and rest would be zero.
   - Take first three column of \( u \) as \( B \).
Still Life

Original Images

Basis Images
Rendering Images: $\sum_i \max(B_{s_i}, 0)$

1 Light

2 Lights

3 Lights
Face Basis

Original Images

Basis Images
Movie with Attached Shadows

Single Light Source  Face Movie
3-D Linear subspace

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

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

\[ L = \{ x \mid x = Bs, \forall s \in \mathbb{R}^3 \} \]

where B is an \( n \) by 3 matrix whose rows are product of the surface normal and Lambertian albedo
Let $L_i$ denote the intersection of $L$ with an orthant $i$ of $\mathbb{R}^n$.

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 image $\mathbf{x}$ produced by multiple light sources is

$$\mathbf{x} = \sum_i \max(\mathbf{B} \mathbf{s}_i, 0)$$

where

- $\mathbf{x}$ is an $n$-pixel image vector.
- $\mathbf{B}$ is a matrix whose rows are unit normals scaled by the albedo.
- $\mathbf{s}_i$ is the direction and strength of the light source $i$. 

Lambertian, Shadows and Multiple Lights
• With two lights on, resulting image along line segment between single source images: superposition of images, non-negative lighting
• For all numbers of sources, and strengths, rest is convex hull of $U$. 
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

• Can the cones of two different objects intersect?
• Can two different objects have the same cone?
• How big is the cone?
• How can cone be used for recognition?
Do Ambiguities Exist?

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

YES
Do Ambiguities Exist? Yes

• Cone is determined by linear subspace $L$
• The columns of $B$ span $L$
• For any $A \in \text{GL}(3)$, $B^* = BA$ also spans $L$.
• For any image of $B$ produced with light source $S$, the same image can be produced by lighting $B^*$ with $S^* = A^{-1}S$ because
  \[ X = B^*S^* = B AA^{-1}S = BS \]
• When we estimate $B$ using SVD, the rows are NOT generally normal * albedo.
In general, $\mathbf{B}^*$ 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:

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

\[
f(x, y) = \lambda f(x, y) + \mu x + \nu y
\]
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.
Uncalibrated photometric stereo

1. Take $n$ images as input, perform SVD to compute $B^*$. 
2. Find some $A$ such that $B^*A$ is close to integrable. 
3. Integrate resulting gradient field to obtain height function $f^*(x,y)$. 

Comments: 
- $f^*(x,y)$ differs from $f(x,y)$ by a GBR. 
- Can use specularities to resolve GBR for non-Lambertian surface.
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
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 \( n \).

The number of extreme rays of the cone is \( n(n-1)+2 \)

(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.

<table>
<thead>
<tr>
<th></th>
<th>Ball</th>
<th>Face</th>
<th>Phone</th>
<th>Parrot</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>48.2</td>
<td>53.7</td>
<td>67.9</td>
<td>42.8</td>
</tr>
<tr>
<td>#3</td>
<td>94.4</td>
<td>90.2</td>
<td>88.2</td>
<td>76.3</td>
</tr>
<tr>
<td>#5</td>
<td>97.9</td>
<td>93.5</td>
<td>94.1</td>
<td>84.7</td>
</tr>
<tr>
<td>#7</td>
<td>99.1</td>
<td>95.3</td>
<td>96.3</td>
<td>88.5</td>
</tr>
<tr>
<td>#9</td>
<td>99.5</td>
<td>96.3</td>
<td>97.2</td>
<td>90.7</td>
</tr>
</tbody>
</table>

[ Epstein, Hallinan, Yuille 95]

Dimension: $5 \pm 2D$
Recent results

• Illumination cone is well captured by nine dimensions for a convex Lambertian surface.
  – Spherical Harmonic representation of lighting & BRDF.
Coke vs. Pepsi: Albedo & Cone Shape

Consider two objects with the same geometry, but different albedo patterns. Their 3-D linear subspaces can be expressed as:

\[ B_1 = R_1 N \]
\[ B_2 = R_2 N \]

where \( R_i \) is an \( n \) by \( n \) diagonal matrix of albedos (i.e. positive numbers)

Proposition: If \( \mathcal{C}_1 \) is the illumination cone for an object defined by \( B_1 = R_1 N \) and \( \mathcal{C}_2 \) is the cone for an object defined by \( B_2 = R_2 N \), then:

\[
\mathcal{C}_1 = \left\{ R_1 R_2^{-1} x : x \in \mathcal{C}_2 \right\}
\]

\[
\mathcal{C}_2 = \left\{ R_2 R_1^{-1} x : x \in \mathcal{C}_1 \right\}
\]
Where’d the moguls go?

• When a convex Lambertian surface is illuminated by perfectly diffuse lighting, the resulting image is directly proportional to the albedo.

• For a convex object, the n-dimensional vector of albedos (and image) is contained within the object’s cone.

• For two objects with the same albedo pattern but different shape, their cones intersect in the interior.

• Two objects differing by a generalized bas relief transformation have the same cone.
A color image of a Lambertian surface without shadowing can be expressed as

\[ x_i = \int \rho_i(\lambda)(R(\lambda)N)(\tilde{s}(\lambda)\hat{s}) d\lambda \]

where:

- \( \lambda \) is the wavelength.
- \( \rho_i(\lambda) \) is pixel response of the i-th color channel.
- \( R(\lambda) \) is an \( n \) by \( n \) diagonal matrix of the spectral reflectance of the facets.
- \( N \) is the \( n \) by 3 matrix of surface normals.
- \( s(\lambda) \) is the spectrum of the light source.
- \( \hat{s} \) is the direction of the light source.
Narrow Band Cameras

For a single channel with \( \rho_i(\lambda) \) modeled as a delta function at \( \lambda_i \), the image is:

\[
\mathbf{x}_i = \rho(\lambda_i)(\mathbf{R}(\lambda_i)\mathbf{N})(\mathbf{s}(\lambda_i)\hat{s})
= \rho_i(\mathbf{R}_i\mathbf{N})(\mathbf{s}_i\hat{s})
\]

For \( c \) channels, the image \( \mathbf{x} \in \mathbb{R}^{cn} \) is

\[
\mathbf{x} = \begin{bmatrix}
\mathbf{x}_1 \\
\mathbf{x}_2 \\
\vdots \\
\mathbf{x}_c
\end{bmatrix}
\]
Narrow Band Cameras

- Overall light source directions and spectral distributions, the set of all images without shadowing is a \( \mathbb{C} + 2 \) dimensional manifold with boundary in \( \mathbb{R}^{cn} \).
- This manifold is embedded in a \( 3c \) dimensional linear subspace of \( \mathbb{R}^{cn} \).
- Any image in this \( 3c \) dimensional linear subspace can be achieved with three light sources.
- The cone in \( \mathbb{R}^{cn} \) can be constructed as the Cartesian product of \( c \) cones in \( \mathbb{R}^{n} \), one for each channel.
Color Image Formation

A color image can be expressed as

$$x_i = \int \rho_i(\lambda)(R(\lambda)N)(\tilde{s}(\lambda)\hat{s})d\lambda$$

where:

- $\lambda$ is the wavelength.
- $\rho_i(\lambda)$ is pixel response of the $i$-th color channel.
- $R(\lambda)$ is an $n$ by $n$ diagonal matrix of the spectral reflectance of the facets.
- $N$ is the $n$ by $3$ matrix of surface normals.
- $s(\lambda)$ is the spectrum of the light source.
- $\hat{s}$ is the direction of the light source.
Consider $c$ color channels (not necessarily narrow band) and a scene illuminated by sources with identical spectra.

For $c$ channels, the image $\mathbf{x} \in \mathbb{R}^{cn}$ is

$$\mathbf{x} = \left[ \mathbf{R}_1 \mid \mathbf{R}_2 \mid \cdots \mid \mathbf{R}_c \right]^\mathsf{T} \mathbf{Ns}$$
Light Sources with Identical Spectra

• The set of images without shadowing lies in a 3-D linear subspace of $\mathbb{R}^{cn}$.
• The illumination cone spans $m$ dimensions of $\mathbb{R}^{cn}$.
• By restricting the spectra of the light source, the cone is much smaller ($m$ dimensions vs. $cm$ dimensions), and the camera does not have to be narrow band.
Major Issues

- Is Euclidean metric best choice for recognition?
- Should recognition metric be posed in image space?
- All lighting conditions are not equally likely - Can and should probabilities be attached to lighting?
- Does it make sense to pose recognition in image space when pose and within class variation are considered?
Illumination Cones: Recognition Method

Is this an image of Lee or David?

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

N-dimensional Image Space

Illumination cone for Lee

Illumination cone for David
Generating the Illumination Cone

Original (Training) Images

3D linear subspace

\[ \alpha(x,y) \]
\[ f_x(x,y) \]
\[ f_y(x,y) \]
albedo
(surface normals)

Surface. \( f(x,y) \)
(albedo textured mapped on surface).

[Cone - Attached]

[Cone - Cast]

[Georghiades, Belhumeur, Kriegman 01]

CSE252A
Predicting Lighting Variation

Single Light Source
Yale Face Database B

64 Lighting Conditions
9 Poses

⇒ 576 Images per Person
Face Recognition: Test Subsets

- **Subset 1:** 0-12°
- **Subset 2:** 12-25°
- **Subset 3:** 25-50°
- **Subset 4:** 50-77°

Test images divided into 4 subsets depending on illumination.

Increasing extremity in illumination
Geodesic Dome Database - Frontal Pose

[Georghiades, Belhumeur, Kriegman 01]

Error Rate vs. Light Direction

- Subset 1,2
- Subset 3
- Subset 4

Error Rate

- Correlation
- Eigenfaces
- Eigenfaces (w/o 1st 3)
- 3-D Linear Subspace
- Illumination Cones