Last class

Correspondence
Review
Final exam (example questions)
1. Detect interest points (corner detector)
2. Estimate scale and orientation of each point
3. Find candidate matches with gradient descriptor
3. Use RANSAC to find a transformation with lots of inliers
4. Fit transformation (with least squares) to set of inliers
5. Align images to transformation
Planar instance object recognition

Very useful to recognizing products
Would interest point + patch matching work for these products?
How can we use for 3D alignment?
Recall: epipolar geometry

1. Find candidate correspondences by matching interest points
2. Select 7 point pairs
3. Estimate epipolar geometry (R, t) with least squares
4. Apply R, t to all points and count the number that fall within some distance of epipolar line
   repeat
   Once we have camera positions and correspondences, how do we get 3D points?
Hence, we can use Ransac with epipolar geometry to estimate camera positions + 3D points
Recall an annoying issue..

What to do when ray’s don’t intersect?
Minimize reprojection error

\[ \min_X f(X) = \| x_1 - Proj(X, C_1) \|^2 + \| x_2 - Proj(X, C_2) \|^2 \]

Perspective projection equations where \( C_1 = (R_1, t_1) \) of camera 1

For this equation, its easier to parameterize camera position in absolute coordinates instead of relative ones.
Extensions

\[ \min_X f(X) = \|x_1 - Proj(X, C_1)\|^2 + \|x_2 - Proj(X, C_2)\|^2 \]

How to generalize to multiple images?

What if (some) camera parameters are unknown?
Bundle adjustment

Minimize reprojection error over multiple cameras and 3D points

\[
\min_{X_1, X_2, \ldots, C_1, C_2, \ldots} \sum_{ij} \Vert x_{ij} - \text{Proj}(X_i, C_j) \Vert^2
\]
Estimate both camera movements and 3D points from successive video frames

Allows us to build 3D maps on-the-fly!
Crucial task in robotics

SFM “Structure from Motion”
SLAM “Simultaneous Localization and Mapping”

http://webdiis.unizar.es/~raulmur/orbslam/
Common task: fitting multiple geometric shapes to images

Can we use Ransac?

Instead of sampling, can we do a “brute-force” search?
Last class

Correspondence
Review
Final exam (example questions)

NOTE: I will use a large subset of the following questions on the final!

March 20 8:00-10:00am ICS 174
A look back

- Optics
- Signal processing
- Geometry
- Graph Algorithms
- Statistics
What’s the reason for the dark spot on the wall in the top-right image? What happens when the window opening is shrunk in size? Why?
Optics

Know perspective projection equation!

Given 3D position in camera coordinates, focal length, what’s the image position?
Optics

The Dolly effect

What is changing between the frames?
Optics

Perspective vs orthographic models.
Under what conditions will images look orthographic?

Wide angle  Standard  Telephoto
Optics

Perspective vs orthographic models versus scaled orthographic
Optics

Thin lens model: be able to derive with 2 simple rules

\[
\frac{1}{s_o} + \frac{1}{s_i} = \frac{1}{f}
\]

http://graphics.stanford.edu/courses/cs178/applets/thinlens.html
http://graphics.stanford.edu/courses/cs178/applets/gaussian.html
Understanding shutters and apertures

Explain why these 3 pictures look different
A look back

Optics

Signal processing

Geometry

Graph Algorithms

Statistics
Signal processing

Linear shift-invariant systems

How is convolution derived as a consequence of LSI assumptions?

Is a median filter LSI? Why or why not?

- Robustness to outliers

Source: K. Grauman
Signal processing: convolution

A Gaussian kernel gives less weight to pixels further from the center of the window. This kernel is an approximation of a Gaussian function:

\[
\begin{array}{ccccccccc}
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 90 & 90 & 90 & 90 & 0 & 0 \\
 0 & 0 & 0 & 90 & 90 & 90 & 90 & 0 & 0 \\
 0 & 0 & 0 & 90 & 90 & 90 & 90 & 0 & 0 \\
 0 & 0 & 0 & 90 & 90 & 90 & 90 & 0 & 0 \\
 0 & 0 & 0 & 90 & 0 & 90 & 90 & 0 & 0 \\
 0 & 0 & 0 & 90 & 90 & 90 & 90 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 90 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\end{array}
\]

Be able to compute result of a small example. Know different between correlation and convolution.
Convolution versus correlation

Original

Shifted left
By 1 pixel

Was the above result obtained by convolution or correlation?

Source: D. Lowe
Describe how one can obtain a sharpened image as a difference of two filtered images. Give examples of the two filters.

Show how the same result can be derived from a single filtering operation.
3. Fill in the blanks:

a) \_ = D * B
b) A = \_ * \_
c) F = D * \_
d) \_ = D * D

(assuming * is correlation, not convolution)
A look back

Optics

Signal processing

Geometry

Optics

Graph Algorithms

Statistics
Image Transformations

image filtering: change *range* of image

\[ g[x] = T(f[x]) \]

image warping: change *domain* of image

\[ g[x] = f(T[x]) \]
Why do we need homogenous coordinates?
What kind of transformation maps one triangle onto another?
How can it be computed from \((A, B, C, A', B', C')\)?
Suppose that the image of a face is scaled uniformly by factor $s$ and translated in either direction by $t_x$ and $t_y$.

(a) Write down the equation for a transformed point $(x', y')$ as a function of the original point $(x, y)$ in terms of $t_x$, $t_y$, and $s$. 

(b) Write the transformation in matrix form in terms of $t_x$, $t_y$, $s$, $x$, $y$, $x'$, and $y'$:

\[
\begin{bmatrix}
  x' \\
  y'
\end{bmatrix} = \begin{bmatrix}
  s & 0 & t_x \\
  0 & s & t_y
\end{bmatrix} \begin{bmatrix}
  x \\
  y
\end{bmatrix}
\]

(c) If you are given two pairs of corresponding points: $(x_1, y_1)$ to $(x_1', y_1')$ and $(x_2, y_2)$ to $(x_2', y_2')$, how do you solve for the transformation parameters $t_x$, $t_y$, $s$? Write down the system of equations in a matrix form (you don’t need to solve it):

\[
\begin{bmatrix}
  s & 0 & t_x \\
  0 & s & t_y
\end{bmatrix} \begin{bmatrix}
  x_1 \\
  y_1
\end{bmatrix} = \begin{bmatrix}
  x_1' \\
  y_1'
\end{bmatrix}
\]

\[
\begin{bmatrix}
  s & 0 & t_x \\
  0 & s & t_y
\end{bmatrix} \begin{bmatrix}
  x_2 \\
  y_2
\end{bmatrix} = \begin{bmatrix}
  x_2' \\
  y_2'
\end{bmatrix}
\]

(d) Suppose that the corresponding points may have small random localization errors.

i. Will the above constraints be satisfied exactly? Why or why not?

ii. Would having more corresponding points with the same distribution of error make the transformation more or less accurate? Why?
Special case: projective or homography transformation

What is the shape of the b/w floor pattern?

The floor (enlarged)

Automatically rectified floor

Slide from Criminisi
Explain, in words, how an artist would figure out what to paint so that the drawing would look correct from a particular viewing angle.
Homography versus epipolar geometry

\[ w \mathbf{x}_2 = H \mathbf{x}_1 \]

Maps points in one image to points in another

\[ \mathbf{x}_2^T E \mathbf{x}_1 = 0 \]

Maps points in one image to line in another
Two-view geometry

Why do we see missing data in a Kinect depth sensor?
A look back

Optics
Signal processing
Geometry
Graph Algorithms
Statistics
Graph algorithms

Be able to compute the optimal path for a small graph.

What is the constrained objective function being minimized?
Object removal: how can the previous objective function be modified to remove an object?
Video (not on exam)

• Before we wanted to find a smooth seam (1D) across an image (2D)
• Now we want a smooth sheet (2D) inside a video (3D)
Alternative formulation: mincut on a graph

- Each pixel is a node that’s connected to its 4 neighbors and a “source” and “terminal”
- Weight edges appropriately (eg, gradient magnitude)
- Find minimum cost cut that separates S and T - this is a “graphcut” problem

- Graphcuts is a very common tool in pixel labeling problems
- Can solve mincut/maxflow problem (cf, Algorithms textbook)
- With a particular choice of of weights on edges, the min cut is equivalent to DP soln
Extension to video: 3D graph of pixels

Video Cube
3D Graph Cut

Video Cube
A look back

Optics

Signal processing

Geometry

Graph Algorithms

Statistics
Statistics

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2/7</td>
<td>1/7</td>
<td>1/7</td>
<td>1/7</td>
<td>1/7</td>
<td>1/7</td>
</tr>
<tr>
<td>2</td>
<td>1/7</td>
<td>2/7</td>
<td>1/7</td>
<td>1/7</td>
<td>1/7</td>
<td>1/7</td>
</tr>
<tr>
<td>3</td>
<td>1/7</td>
<td>1/7</td>
<td>2/7</td>
<td>1/7</td>
<td>1/7</td>
<td>1/7</td>
</tr>
<tr>
<td>4</td>
<td>1/7</td>
<td>1/7</td>
<td>1/7</td>
<td>2/7</td>
<td>1/7</td>
<td>1/7</td>
</tr>
<tr>
<td>5</td>
<td>1/7</td>
<td>1/7</td>
<td>1/7</td>
<td>1/7</td>
<td>2/7</td>
<td>1/7</td>
</tr>
<tr>
<td>6</td>
<td>1/7</td>
<td>1/7</td>
<td>1/7</td>
<td>1/7</td>
<td>1/7</td>
<td>2/7</td>
</tr>
</tbody>
</table>

Joints, conditionals, independance

- \( P(R1|R2) = \frac{P(R1,R2)}{P(R2)} \)
Statistics: markov models

\[
P(\text{word1}, \text{word2}, \text{word3})
\]

\[
\text{word1} \text{ is conditionally independent of word2 iff } P(\text{word3}|\text{word1}, \text{word2}) = P(\text{word3}|\text{word2})
\]
Bayes rule for building a skin detector

Given likelihoods of skin and non-skin, and prior probabilities of skin and non-skin, what would be the predicted label of a RGB color?
A look back

Signal processing

Geometry

Graph Algorithms

Statistics

Optics
Recognition

What invariances are required for general recognition (person detector) versus instance recognition (deva-detector)?

How can one build a detector that works for frontal and side-view cars?

Know how ransac works!
Parting thoughts

You will be responsible for material from notes in class and on the slides (not text or papers)

Please give feedback on evaluation forms

Good luck on the final!