視覚情報処理論
(Visual Information Processing)
開講所属: 学際情報学府
水(Wed)5 [16:50-18:35]
Schedule

- 9/26  Introduction (Prof. Oishi)
- 10/3  Patch-based Object Recognition (1) (Dr. Kagesawa)
- 10/10 Patch-based Object Recognition (2) (Dr. Kagesawa)
- 10/17 Computer Vision basics (1)(Prof. Oishi)
- 10/24 Computer Vision basics (2)(Prof. Oishi)
- 10/31 Image and Video Inpainting (1) (Dr. Roxas) (※in English)
- 11/7  Image and Video Inpainting (2) (Dr. Roxas) (※in English)
- 11/14 (Cancelled)
- 11/21 Vision for Robotics Applications (1) (Dr. Sato)
- 11/28 Vision for Robotics Applications (2) (Dr. Sato)
- 12/5  3D Data Visualization (1) (Dr. Okamoto)
- 12/12 3D Data Visualization (2) (Dr. Okamoto)
- 12/19 3D Data Processing (1) (Prof. Oishi)
- 1/9  3D Data Processing (2) (Prof. Oishi)
Computer Vision Paradigm (Marr)

Object oriented

- 3D Model
- 3D representation

Observer oriented

- 2.5D Image
- Integration

- Brightness
- Texture
- Line drawing
- Stereo
- Motion

3D Feature Extraction (shape-from-x)

2D Image
3D Measurement method (non-contact)

- Passive
  - Silhouette
  - Stereo

- Active
  - Structured light
  - Laser range sensor
Stereo

■ Fixed Camera System
  • Binocular Stereo
  • Multi-Baseline Stereo

■ Moving Camera System
  • Motion Stereo
  • Multi-View Stereo
Binocular Stereo
Binocular Stereo
Binocular stereo

A single image is ambiguous, but another image taken from a different direction gives the unique 3D point.
Epipolar constraints

- Epipolar line
- Possible line of sight
- Base line
- Epipolar plane
- One image point
- Corresponding points lie on the Epipolar lines
Epipoles

- Intersections of baseline with image planes
- Projection of the optical center in another image
- Vanishing points of camera motion direction
Examples of epipolar lines
Examples of epipolar lines
Examples of epipolar lines
Rectification
Panoramic View Interpolation

Left panorama

Right panorama

Input panorama pair

Rectified panorama pair
Panorama rectification

$\Omega_0(\theta_e, \phi_e) \rightarrow \Omega_1(\theta_e, \phi_e, \alpha, \beta, \gamma)$
Terminology

A 3D point

left image point

left image plane

center

right image plane

right image point

World coordinate system

t focal length

base line length
Pinhole Camera

Fig. 131.—How Light and a Pinhole Form an Image.
Perspective Projection

\[
f : \text{focal length}
\]

\[
\frac{u}{f} = \frac{X}{-Z}, \quad (u, v) = \left( f \frac{X}{-Z}, \ f \frac{Y}{-Z} \right)
\]
Basic binocular stereo equation

\[ \frac{d + x}{-z} = \frac{x''}{f} \]

\[ x'' = -\frac{f}{z}(x + d) \]

\[ z = \frac{-2df}{(x'' - x')} \]

\[ x'' - x' = -\frac{f}{z}(x + d - x + d) = -\frac{f}{z}2d \]

\[ \frac{d - x}{-z} = \frac{-x'}{f} \]

\[ x' = -\frac{f}{z}(x - d) \]

\( d \): base line length

\( z \): disparity
Features for matching

a. Brightness

b. Edges

c. Interest points
Classification of stereo methods

- Features for matching
  - brightness value
  - feature point
  - edge
  - region

- Strategies for matching
  - brute-force
  - coarse-to-fine
  - relaxation
  - dynamic programming

- Constraints for matching
  - epipolar lines
  - continuity
Block-Matching Stereo

problem
a. trade-off of window size and resolution
b. dull peak
Cost Function

A) SAD (sum. of absolute difference)
\[
d = \min_d F = \min_d \sum_w \| I_l(x, y) - I_r(x - d, y) \|
\]

B) SSD (sum. of squared difference)
\[
d = \min_d F = \min_d \sum_w (I_l(x, y) - I_r(x - d, y))^2
\]

C) Correlation
\[
d = \min_d F = \min_d -\frac{Cov(I_l(x, y), I_r(x - d, y))}{\sqrt{Var(I_l(x, y)) Var(I_r(x - d, y))}}
\]
Coarse to fine
Coarse to fine

Matching

Expand

Matching

Expand

Matching

Matching
Disparity estimation

How to estimate the disparities?

Minimize some cost function $F(I_l, I_r)$ along the epipolar line

H. Hirschnuller, "Improvements in Real-Time Correlation-Based Stereo Vision", IEEE Workshop on Stereo and Multi-Baseline Vision, 2001
Fatting Effect on Object Boundary

No single window fits at discontinuity

⇒ Fatting effect of the object
Accurate Estimation on Object Boundary

- **Shiftable Window**

\[
c_0 = SAD(x, y, d) \\
c_1 = SAD(x - \frac{w}{2}, y - \frac{h}{2}, d) \\
c_2 = SAD(x + \frac{w}{2}, y - \frac{h}{2}, d) \\
c_3 = SAD(x - \frac{w}{2}, y + \frac{h}{2}, d) \\
c_4 = SAD(x + \frac{w}{2}, y + \frac{h}{2}, d)
\]

\[c' = \text{Min}\{c_1, c_2, c_3, c_4\} \]
\[c'' = \text{Min}\{\{c_1, c_2, c_3, c_4\} - c'\} \]

\[\text{disparity} = \arg \min_{d} c_0 + c' + c''\]
Consistency check

- Check if two independent disparity estimation coincide
  - Left $\Rightarrow$ Right search
  - Right $\Rightarrow$ Left search

- Inconsistent disparities are considered as a false match
Result

- SAD with 11x11 window
- Shiftable window + consistency checking

Left image

Disparity map
Constraints

- Epipolar line constraint
- Uniqueness constraint
  - each point in a image has only one depth value

O.K. No.

- Continuity constraint
  - each point is almost sure to have a depth value near the values of neighbors

O.K. No.
Uniqueness constraint

**Uniqueness constraint prohibits two or more matching points on one horizontal or vertical lines**
Continuity constraint

- Continuity constraint attracts more matching on a diagonal line

---

(E-A) (E-B) (E-C)

D      E      F

Same depth

Attract

Prohibit

Attract
Pixel-based Stereo Matching via Graph-cut

Labeling $f$ means disparity assignment

$$E(f) = E_{data}(f) + E_{smooth}(f)$$

$$E_{data} = \sum_{image} (I_l(p_x, p_y) - I_r(p_x - f, p_y))^2$$

$$E_{smooth} = \sum_{q \in N_p} (f_p - f_q)^2$$

minimizing the cost function by iterative graph cut ($\alpha$-expansion)
Graph-cut

- Solve graph partition problem in globally optimal way
  - Formulate the problem in energy minimization framework
  - Design a graph such that the sum of cut edges equals to the total energy
  - Find a “Cut” that minimizes the energy

\[
\min E = \sum_{e_{i,j} \in C} V_{i,j}
\]
Example of Graph-cut

- Image segmentation

\[
\min_C E = \sum_{\{e_{i,j}\} \in C} V_{i,j}
\]

**Graph partition**

- **Graph**={N,e}
- **Ni**: Graph node
- **eij**: Edge connecting nodes
- **C**: Cut

**Segmentation**

- **Image**={pixel}
- **Vij**: Similarity between neighboring pixels
- **Foreground/Background boundary**
Solution to a Graph-cut Problem

Min-Cut/Max-Flow algorithm
- Given source (s) and sink nodes (t)
- Define capacity on each edge $C_{i,j} = V_{i,j}$
- Find the maximum flow from $s \Rightarrow t$, satisfying capacity constraints, and cut the bottleneck

Multi-label Problem

• Find labeling $f$ that minimizes the energy

$$E(f) = E_{\text{smooth}}(f) + E_{\text{data}}(f)$$

$E_{\text{smooth}}(f)$ measures the extent to which $f$ is not piecewise smooth

$E_{\text{data}}(f)$ measures the disagreement between $f$ and the observed data

$$E(f) = \sum_{\{p,q\} \in N} V_{p,q}(f_p, f_q) + \sum_{p \in P} D_p(f_p)$$

$D_p$ Measures how well label $f_p$ fits pixel $p$ given the observed data

$V_{p,q}$ Smoothness penalty between adjacent (N) pixels

Multi-label Solution via Graph-cut

- Iterative graph-cut approach
- 2 types of move algorithm are proposed
  - $\alpha\beta$-swap
  - $\alpha$-expansion

$$\min E \text{ under cond.} \quad f_{r \notin \{\alpha, \beta\}} \text{ is preserved}$$

$$\min E \text{ under cond.} \quad f_{r \notin \{\alpha\}} \text{ can be changed to } \alpha$$
**αβ-swap Algorithm**

1. Start with an arbitrary labeling $f$
2. Success:= 0
3. For each pair of labels $\{\alpha, \beta\} \subset L$
   a. Find $f' = \arg \min E(f')$ among $f'$ within one $\alpha\beta$-swap of $f$
   b. If $E(f') < E(f)$ then $f' := f$ and success:=1
4. If success=1 goto 2
5. Return $f$
\( \alpha \)-expansion Algorithm

1. Start with an arbitrary labeling \( f \)
2. Success:= 0
3. For each label \( \alpha \in L \)
   a. Find \( f' = \text{arg min } E(f') \) among \( f' \) within one \( \alpha \)-expansion of \( f \)
   b. If \( E(f') < E(f) \) then \( f' := f \) and success:=1
4. If success=1 goto 2
5. Return \( f \)
Graph-cut with Occlusions

- Occluded pixels are handled explicitly in the graph
- Find a subset of A
  \[ A = \{ (p, q) \mid p_y = q_y \text{ and } 0 \leq q_x - p_x \leq k \} \]
- Find a configuration f such that
  \[ a \in A \quad f_a = 1 \quad \text{if the pixels } (p,q) \text{ correspond} \]
  \[ f_a = 0 \quad \text{otherwise} \]

\[ p \text{ is an occluded pixel} \quad \text{if } f_a = 0 \text{ for } a = (p, q) \in A \]

Energy Function to Minimize

\[ E(f) = E_{\text{data}}(f) + E_{\text{occlusion}}(f) + E_{\text{smooth}}(f) \]

\[ E_{\text{data}}(f) = (I_l(p_x, p_y) - I_r(p_x - d, p_y))^2 \]

\[ E_{\text{occlusion}}(f) = \sum_{p \in P} C_p \cdot T(\lceil N_p(f) \rceil = 0) \]

Occlusion penalty

\[ E_{\text{smooth}}(f) = \sum_{\{a1,a2\} \in N} V_{a1,a2} \cdot T(f(a1) \neq f(a2)) \]

\[ V_{a1,a2} = \begin{cases} 
3\lambda & \text{if } \max(|I_l(p) - I_l(r)|, |I_r(q) - I_r(s)|) < \beta \\
\lambda & \text{otherwise}
\end{cases} \]
Result

Left Image

Ground Truth

L1 correlation

Graph-cut without Occ.

Graph-cut with Occ.
DP (Dynamic Programming) Stereo

Path Search

- Matching problem can be considered as a path search problem

- Define a cost at each candidate of path segment based on some ad-hoc function
Dynamic programming

We can formalize the path finding problem as the following iterative formula:

\[ D(M) = \min_{\{k\}} \left\{ d(M; k) + D(k) \right\} \]

where:
- \( D(M) \) is the optimum cost to reach \( K \)
- \( d(M; k) \) is the cost between \( M \) and \( K \)
- \( D(k) \) is the optimum cost to reach \( K \)

Optimum costs are known.

For initial conditions:

\[ D(0) = \min \left\{ d(0;3) + D(3), d(0;2) + D(2), d(0;1) + D(1) \right\} \]
stereo pair
edges
deepth
4-move, 4-plane DP

- Occluded pixels are handled explicitly in 4-move, 4-plane representation

- Disparity map is calculated under DP Matching (global energy minimization)

**Conventional 3-move DP**

Occluded path and visible path cannot be distinguished in this representation.
4-move DP

Occluded path and visible path are handled separately

Occluded move (r)

Matched move (r)

Occluded move (l)

Matched move (l)

True matching path

Approximated matching path

Occluded path and visible path are handled separately
Design of Move Transition

Move Transition

Lo: Left Occluded Move
Lm: Left Matched Move
Ro: Right Occluded Move
Rm: Right Matched Move

Matching Cost

Normalized Sum of Squared Difference

\[ M(l, r) = \frac{1}{2} \frac{\sum [(l_{pl}^l - \bar{l}_{pl}^l) - (l_{pr}^r - \bar{l}_{pr}^r)]^2}{\sum (l_{pl}^l - \bar{l}_{pl}^l)^2 + \sum (l_{pr}^r - \bar{l}_{pr}^r)^2} \]
4-move, 4-plane DP

- Each node should hold 4 accumulation costs separately for each move
  - 4-plane model
Result

Input Images

3-move DP

4-move, 4-plane DP

[Images of results with color-coded areas indicating occluded pixels]
Comparison in Severe Situation

Input

left right

Occluded Pixels

Texture-less region

Result

Block Matching Graph-cut with occlusions 4-move, 4-plane DP
Multiple Baseline Stereo
Multiple Baseline Stereo
Moravec’s cart

Slide stereo

Motion stereo

Slider stereo (9 eyes stereo)

\[ _9C_2 = 36 \text{ stereo pairs!!!} \]

- Each stereo has an uncertainty measure
- Uncertainty = 1 / base-line

Long base line

Small uncertainty
Estimated distance

σ: uncertainty measure

Area: confidence measure

$\sigma = 36$ curves

Figure 7-2: A typical ranging. The nine pictures are from a slider scan. The interest operator chose the marked feature in the central image, and the correlator found it in the other eight. The small curves at bottom are distance measurements of the feature made from pairs of the images. The large beaded curve is the sum of the measurements over all 36 pairings. The horizontal scale is linear in inverse distance.
Matching Ambiguity

Figure 2: An example scene. The grid pattern in the background has ambiguity of matching.

Evaluation function

- Short Baseline

- Large Baseline
Fig. 5. SSD values versus inverse distance: (a) $B = b$; (b) $B = 2b$; (c) $B = 3b$; (d) $B = 4b$; (e) $B = 5b$; (f) $B = 6b$; (g) $B = 7b$; (h) $B = 8b$.

The horizontal axis is normalized such that $bb'F = 1$.

Fig. 7. Combining multiple baseline stereo pairs.
Motion Stereo
Motion Stereo

StereoScan: Dense 3D Reconstruction in Real-time
[A. Geiger et al. Intelligent Vehicles Symposium (IV), 2011]
Benchmark

Kitti Dataset
http://www.cvlibs.net/datasets/kitti/
Multi-view Stereo
Multi-view Stereo

- Dense multi-view stereo
  Patch-based multi-view stereo
  [Furukawa and Ponce, CVPR 2007]
Structure-from-motion theorem

Given three distinct orthographic projection of four non-coplanar points in a rigid configuration, the structure compatible with the three views are uniquely determined.

[Ullman]
Shape and motion without depth

Observation Matrix

\[
\begin{bmatrix}
  x_{11} & x_{12} & x_{13} & x_{14} & x_{15} \\
  x_{21} & x_{22} & x_{23} & x_{24} & x_{25} \\
  x_{31} & x_{32} & x_{33} & x_{34} & x_{35} \\
  y_{11} & y_{12} & y_{13} & y_{14} & y_{15} \\
  y_{21} & y_{22} & y_{23} & y_{24} & y_{25} \\
  y_{31} & y_{32} & y_{33} & y_{34} & y_{35}
\end{bmatrix}
\]

Rank = 3

motion matrix

shape matrix

\[
x_{11} = l_1 \cdot X_1
\]

\[
= (i_{1x} \ i_{1y} \ i_{1z}) \cdot (x_1 \ y_1 \ z_1)
\]

Singular Value Decomposition

Tomasi & Kanade
Shape from Shading
Why does it look sphere?
Incident light intensity

Irradiance

- amount of light falling on a surface falling energy measured by a unit surface area

\[ E = \frac{d\Phi}{dS} \text{ [W/m}^2\text{]} \]
Emitting light intensity

Radiance

- amount of light radiated from a surface emitting energy measured from a unit for shorted light source surface area to a unit solid angle

\[
L = \frac{d\Phi}{d\omega \cdot dS \cdot \cos \theta} \text{[W/sr \cdot m}^2]\]
Reflection geometry

• Irradiance at a pixel depends on
  • **Illumination**
  • **Material**
  • **Geometry**

  • under the same illuminate condition, we observe irradiance difference on the same material surface
  • there is a relationship between pixel irradiance and geometry

• Reflectance geometry

\[ L = \text{illumination} \]
\[ N = \text{normal} \]
\[ V = \text{viewer} \]

\[ L \cdot \mathbf{N} = \text{i} \]
\[ V \cdot \mathbf{N} = \text{e} \]
\[ L \cdot V = \text{g} \]

\[ \cos i = N \cdot L \]
\[ \cos e = N \cdot V \]
\[ \cos g = L \cdot V \]
Gradient space

- Reflection functions are defined in the local coordinate system (e, i, g).
- For our development, we will redefine the reflectance geometry in the gradient space.

\[
p = \frac{\partial Z}{\partial X}, q = \frac{\partial Z}{\partial Y}, \quad \tilde{n} = \frac{(p, q, 1)}{\sqrt{p^2 + q^2 + 1}}
\]

Viewers is always on the Z axis.
Surface and body reflection

- Surface reflection = gloss, highlights very directional (specular)
- Body reflection = object color all direction (diffuse)
  - Plastic, paint have both
  - Metal has only surface reflection
Model for body reflection

- Diffuse---scatters in all directions
- Common approximation:
  - equal in all directions
  - “lambertian” Lambertian’s cosine law
  - “perfectly diffuse reflector”
- Reflectance = constant * geometric factor
  - $f(i, e, g) = K_b \times \cos i$
- Why $\cos i$?

Angle of incidence affects “density” of illumination. (Irradiance)

Irradiance = light/area
light = 1
area = 1/\cos i
Irradiance = \cos i
Calculating a reflection map (Lambertian)

- for each \((p, q)\), \(N = (p, q, 1)\)
- light source direction, \(S = (p_s, q_s, 1)\)

\[
R(p, q, p_s, q_s) = \cos i = N \cdot L = \frac{p \cdot p_s + q \cdot q_s + 1}{\sqrt{p^2 + q^2 + 1} \sqrt{p_s^2 + q_s^2 + 1}}
\]

\((p_s, q_s) = (0, 0)\)

\[
R(p, q) = \frac{1}{\sqrt{p^2 + q^2 + 1}}
\]

iso-brightness contour

\[0.5\]

\[0.8\]

\[0.9\]
Reflectance map (continue)

Lambertian

\[ R(p, q, p_s, q_s) = \cos \beta = N \cdot L \]

\[ = \frac{p \cdot p_s + q \cdot q_s + 1}{\sqrt{p^2 + q^2 + 1} \sqrt{p_s^2 + q_s^2 + 1}} \]

Self-shadow line
Surface reflection

- Metals have the only surface reflection
- Dielectrics (plastics, paint) have the surface reflection as well as the body reflection
- Simplest approximation: perfect mirror

reflection is specular direction, $S'$$ S'$ is coplanar with $S,N$$ SN = i = NS'$: opposite sides
Phong's model
tells amount of light at each angle

\[ f\text{-surface}(\text{i,e,g}) = K_s \times \cos \alpha \]

calculate angle between S' and V ---\( \alpha \)

typical : \( n = 10 \) to 500

heuristic model
Torrance and Sparrow model

Geometrical optics
- a collection of planar mirror-like facets
- surface reflection caused only by these microfacets
- their sizes are much larger than wave length

facet slopes to be normally distributed

\[ p_\alpha(\alpha) = c \cdot \exp\left(-\frac{\alpha^2}{\sigma_\alpha^2}\right) \]
\[ \langle \alpha \rangle = 0 \]
V-shaped valleys
Torrance-Sparrow model

\[ K_d \frac{\cos i}{\cos e} g(i, e) \exp \left( -\frac{\alpha^2}{\sigma^2} \right) \frac{4}{\cos i} \]

Material \hspace{1cm} Light direction

Geometry attenuation \hspace{1cm} Facet distribution

\[ K_d \ g(i, e) \frac{1}{\cos e} \exp \left( -\frac{\alpha^2}{\sigma^2} \right) \]
Calculating reflectance map
specular lobe + diffuse lobe

$L = (p_s, q_s, 1)$
$N = (p, q, 1)$
$V = (0,0,1)$

then $S = \left( \frac{\sqrt{1 + p_s^2 + q_s^2}}{p_s^2 + q_s^2} - 1 \right) p_s, \left( \frac{\sqrt{1 + p_s^2 + q_s^2}}{p_s^2 + q_s^2} - 1 \right) q_s, 1$}

$R(p, q, p_s, q_s) =$

$K_b \frac{p \cdot p_s + q \cdot q_s + 1}{\sqrt{p^2 + q^2} + 1/\sqrt{p_s^2 + q_s^2} + 1} + K_s \exp\left(-\left(\frac{a}{m}\right)^2\right)/\cos e$

$\cos a = S \cdot V$
$\cos e = N \cdot V$
Shape-from-shading

recover object shape (orientation)
from image irradiance (brightness)

$E(x,y) = R(p,q)$ -- image irradiance equation

gives one constraint on the gradient space at each pixel

--- > ill-posed problem (cannot solve !!!!!)
Photomotronic stereo

• One image irradiant equation gives only one constraint
  --- > use multiple equations at each pixel.
• Take multiple images from the same points under different light source directions
• Recall different light source directions give different reflectance map
• At each pixel, multiple irradiance values
Photometric Stereo
Occluding boundary

- Surface orientations on occluding boundaries are known from the shape of silhouette.
- These surface orientations cannot be represented by the gradient space. $(p,q)$ becomes infinite.
- We will use the stereographic plane, $(f,g)$. On $(f,g)$ plane, occluding boundaries lie on the unit circle.

**Boundary condition**
Stereographic Projection

\[ f = \frac{2p}{1 + \sqrt{1 + p^2 + q^2}}, \quad g = \frac{2q}{1 + \sqrt{1 + p^2 + q^2}} \]

occluding boundaries lie on the unit circle

gradient space

gaussian sphere
Relaxation method

1. Image irradiance equation on \((f, g)\) space on the \((f, g)\) space, we can also define a reflectance map.

\[
E(x, y) = R(f(x, y), g(x, y))
\]

2. Smoothness constraint. Neighboring points have roughly the same surface orientation.

\[
\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2 + \left(\frac{\partial g}{\partial x}\right)^2 + \left(\frac{\partial g}{\partial y}\right)^2 = 0
\]
Relaxation method

3. Set up a minimization problem.

\[ E = \iint (E(x, y) - R(f(x, y), g(x, y)))^2 \]
\[ + \lambda \left\{ \left( \frac{\partial f}{\partial x} \right)^2 + \left( \frac{\partial f}{\partial y} \right)^2 + \left( \frac{\partial g}{\partial x} \right)^2 + \left( \frac{\partial g}{\partial y} \right)^2 \right\} dxdy \rightarrow \text{min} \]

4. Using the calculus of variations get iterative formula.

\[ f^{n+1}(x, y) = \frac{1}{4} \left\{ f^n(x+1, y) + f^n(x-1, y) + f^n(x, y+1) + f^n(x, y-1) \right\} \]
\[ + \lambda \left\{ E(x, y) - R(f^n(x, y), g^n(x, y)) \right\} \]

\[ g^{n+1}(x, y) = \frac{1}{4} \left\{ g^n(x+1, y) + g^n(x-1, y) + g^n(x, y+1) + g^n(x, y-1) \right\} \]
\[ + \lambda \left\{ E(x, y) - R(f^n(x, y), g^n(x, y)) \right\} \]
The brightness image is occluding the boundary. The needle map is used to generate the depth map. The equations for the $n+1$th solution are:

$$f^{n+1}(x, y) = \frac{1}{4}\{f^n(x-1, y) + \ldots\} + \lambda\{\ldots\}$$

$$g^{n+1}(x, y) = \frac{1}{4}\{g^n(x-1, y) + \ldots\} + \lambda\{\ldots\}$$

Ikeuchi & Horn 81
Photometric Stereo
Summary

■ Stereo
  • Binocular Stereo (Epipolar constraint, Graph-cut, DP matching)
  • Motion Stereo
  • Multi-view Stereo

■ Brightness analysis (Shape from shading, Photometric stereo)