Lecture 10: Stereo

Tuesday, Oct 2

Grad student extension ideas for problem set 2

- Implement textons approach for texture recognition [Leung & Malik]
  - Possible data sources: Vistex, Curet databases
- Build a shape-based object detector using the generalized Hough transform
- Clustering approach to video shot boundary detection
- Build a deformable contour tracker

Exam

- Next Tuesday, Oct 9, in class
- Bring one handwritten 8.5 x 11”, one-sided sheet with any notes
- Closed book/laptop/calculator

Review all material covered so far

- Image formation
  - Perspective, orthographic projection properties, equations, effects
  - Pinhole cameras
  - Thin lens
  - Field of view, depth of field
- Color
  - Color opponent theory
  - Spectral power distribution
  - Color mixing
  - Color matching
  - Color space
  - Human perception
- Binary image analysis
  - Histograms and thresholding
  - Connected components
  - Morphological operators
  - Region properties and invariance
  - Region-based segmentation
  - Chamfer distance
- Filters
  - Application of filters to
    - Convolution properties
    - Noise models
    - Mean, median, Gaussian, derivative filters
    - Separability
- Edges, pyramids, sampling
  - Image gradients
  - Effects of noise
  - Derivative of Gaussian, Laplacian filters
  - Canny edge detection
  - Corner detection
  - Sampling and aliasing
  - Pyramids – construction and applications
- Texture
  - Analysis vs. synthesis
  - Representations
  - Grouping
    - Gestalt principles
    - Clustering: agglomerative k-means, mean shift, graph-based
    - Graph and affinity matrices
- Fitting
  - Hough transform
  - Generalized Hough transform
  - Least squares
  - Incremental fitting, r-estimates
  - Robust fitting: RANSAC, M-estimators
  - Deformable contours, energy functions
- Stereo vision

Outline

- Brief review of deformable contours
- Fundamentals of stereo vision
- Epipolar geometry

Last time: deformable contours

a.k.a. active contours, snakes

initial
intermediate
final
Snake energy function

The total energy of the current snake defined as
\[ E_{\text{total}} = E_{\text{int}} + E_{\text{ex}} \]

Internal energy encourages smoothness or any particular shape

External energy encourages curve onto image structures (e.g. image edges)

Internal energy incorporates prior knowledge about object boundary, which allows a boundary to be extracted even if some image data is missing

We will want to iteratively minimize this energy for a good fit between the deformable contour and the target shape in the image

Discrete energy terms

• An external energy term for a (discrete) snake based on image edge

\[ E_{\text{ex}} = - \sum_{i=0}^{n-1} |G_{x}(x_i, y_i)|^2 + |G_{y}(x_i, y_i)|^2 \]

Discrete energy terms

• If the curve is represented by \( n \) points

\[ v_i = (x_i, y_i) \quad i = 0 \ldots n-1 \]

\[ \frac{d}{ds} v_i = \frac{v_{i+1} - v_i}{2} \quad \frac{d^2}{ds^2} (v_{i+1} - v_i) - (v_i - v_{i-1}) = v_{i+1} - 2v_i + v_{i-1} \]

\[ E_{\text{int}} = \sum_{i=0}^{n-1} \alpha |v_{i+1} - v_i|^2 + \beta |v_{i+1} - 2v_i + v_{i-1}|^2 \]

Energy minimization

• Many algorithms proposed to fit deformable contours
  – Greedy search
  – Gradient descent
  – Dynamic programming (for 2d snakes)

Problems with snakes

• Depends on number and spacing of control points
• Snake may oversmooth the boundary
• Not trivial to prevent curve self intersecting
• Cannot follow topological changes of objects

Problems with snakes

• May be sensitive to initialization, get stuck in local minimum

• Accuracy (and computation time) depends on the convergence criteria used in the energy minimization technique
Problems with snakes

- External energy: snake does not really "see" object boundaries in the image unless it gets very close to it.

Depth unavailable in single views

What cues can indicate 3d shape?

Focus/Defocus

Texture

Shading
Motion

Figures from L. Zhang http://www.brainconnection.com/teasers/?main=illusion/motion-shape

Estimating scene shape

• Shape from X: Shading, Texture, Focus, Motion…

• Stereo:
  – shape from motion between two views
  – infer 3d shape of scene from two (multiple) images from different viewpoints

Accommodation and focus

From Bruce and Green, Visual Perception, Physiology, Psychology and Ecology

The lens modifies the image focus by adjusting its focal length.

Accommodation and focus

From Bruce and Green, Visual Perception, Physiology, Psychology and Ecology

Fixation, convergence

From Palmer, "Vision Science", MIT Press

Human stereopsis: disparity

Figure 5-2.3  Convergence as a function of distance. The angle of convergence changes rapidly with distances up to a meter or two but very little after that.

Disparity occurs when eyes verge on one object; others appear at different visual angles
Human stereopsis: disparity

Disparity: \( d = r - l = D - F \).

Random dot stereograms

- Julesz 1960: Do we identify local brightness patterns before fusion (monocular process) or after (binocular)?

- To test: pair of synthetic images obtained by randomly spraying black dots on white objects

Random dot stereograms

When viewed monocularly, they appear random; when viewed stereoscopically, see 3d structure.

- Conclusion: human binocular fusion not directly associated with the physical retinas; must involve the central nervous system

- Imaginary "cyclopean retina" that combines the left and right image stimuli as a single unit
Generating a random dot stereogram

http://www.wellesley.edu/CS/LED/Fall00/Pozdolski/Brault.paper20.html

Autostereograms

Images from magiceye.com

Exploit disparity as depth cue using single image
(Single image random dot stereogram, Single image stereogram)

Images from magiceye.com

Autostereograms

Images from magiceye.com

Stereo photography and stereo viewers

Take two pictures of the same subject from two slightly different viewpoints and display so that each eye sees only one of the images.

Invented by Sir Charles Wheatstone, 1838

Image courtesy of fisher-price.com

http://www.johnsonshawnmuseum.org
Stereo

- Main issues
  - Geometry: what information is available, how do the camera views relate?
  - Correspondences: what feature in view 1 corresponds to feature in view 2?
  - Triangulation, reconstruction: inference in presence of noise

Multi-view geometry

- Relate
  - 3-D points
  - Camera centers
  - Camera orientation
  - Camera intrinsics
Camera parameters

Extrinsic:
- Camera frame ↔ Reference frame
- Image coordinates relative to camera ↔ Pixel coordinates

Intrinsic:
- Image coordinates relative to camera

- Extrinsic params: rotation matrix and translation vector
- Intrinsic params: focal length, pixel sizes (mm), image center point, radial distortion parameters

Geometry for a simple stereo system

- First, assuming parallel optical axes, known camera parameters:

- Parameters in this case:
  - Camera centers (Ol, Or)
  - Focal length (f)
  - Baseline (T)

Geometry for a simple stereo system

- First, assuming parallel optical axes, known camera parameters:

Similar triangles (pl, P, pr) and (Ol, P, Or)

\[ T + x_l - x_r = T \]

\[ Z - f \]

\[ Z = f \frac{T}{d} \]

where \( d = x_r - x_l \)

Stereo constraints

- Given p in left image, where can corresponding point p' be?
Stereo constraints

• Given p in left image, where can corresponding point p' be?

Epipolar geometry

Now the optical axes are not necessarily parallel.

• Baseline: line joining the camera centers
• Epipole: point of intersection of baseline with the image plane
• Epipolar plane: plane containing baseline
• Epipolar line: intersection of epipolar plane with the image plane

• All epipolar lines intersect at the epipole
• An epipolar plane intersects the left and right image planes in epipolar lines

Epipolar constraint

• Potential matches for p have to lie on the corresponding epipolar line l'.
• Potential matches for p' have to lie on the corresponding epipolar line l.
Example: converging cameras

As position of 3d point varies, epipolar lines "rotate" about the baseline

Example: forward motion

Epipole has same coordinates in both images.
Points move along lines radiating from $e$: "Focus of expansion"

Reconstruction by triangulation

- Assuming intrinsic and extrinsic parameters are known, compute 3d location of point $P$ from projections $p$ and $p'$:
  - Intersect rays $R = Op$ and $R' = O'p'$.

But, in practice, parameters and image locations only approximately known...

Triangulation with non-intersecting rays (1)

Construct line segment perpendicular to $R$ and $R'$ that intersects both rays
Midpoint of this segment is closest point to the two rays, use as $P$ estimate of scene point
Triangulation with non-intersecting rays (2)

Estimate scene point Q as the point that minimizes summed squared distance between p and q, and p' and q' (non-linear least squares, iterative, not closed form)

Next

- 3d reconstruction
- Building stereo algorithms
  - correspondences