Stereo Vision

ECE 285
Abhay Bhopat
13th Mar 2013
Reference


• Y Boykov, O Veksler, R Zabih, Fast approximate energy minimization via graph cuts, IEEE Trans PAMI, 23(11), 1222-1239 (2001)


http://vision.middlebury.edu/stereo/
What is the goal of stereo vision?

- The recovery of the 3D structure of a scene using two or more images of the 3D scene, each acquired from a different viewpoint in space.
- The term **binocular vision** is used when two cameras are employed.
Stereo setup

Cameras in arbitrary position and orientation:

More convenient setup:
Stereo terminology

- **Fixation point**: the point of intersection of the optical axes.

- **Baseline**: the distance between the centers of projection.

- **Conjugate pair**: any point in the scene that is visible in both cameras will be projected to a pair of image points in the two images (corresponding points).
Stereo terminology (cont’d)

- **Disparity**: the distance between corresponding points when the two images are superimposed.

- **Disparity map**: the disparities of all points form the disparity map.
Triangulation

- Determines the position of a point in space by finding the intersection of the two lines passing through the center of projection and the projection of the point in each image.
The two problems of stereo

(1) The *correspondence* problem.

(2) The *reconstruction* problem.
The correspondence problem

- Finding pairs of matched points such that each point in the pair is the projection of the same 3D point.

left camera  right camera
The correspondence problem (cont’d)

• Triangulation depends **crucially** on the solution of the correspondence problem!

• Ambiguous correspondences may lead to several different consistent interpretations of the scene.
The reconstruction problem

• Given the corresponding points, we compute the disparity map.

• The disparity map can be converted to a 3D map of the scene assuming that the stereo geometry is known.
Recovering depth (i.e., reconstruction)

- Consider recovering the position of $P$ from its projections $p_l$ and $p_r$

$P_l$: $P$ in left camera coordinates

$P_r$: $P$ in right camera coordinates

Right camera

$x_r = f \frac{X_r}{Z_r}$

Left camera

$x_l = f \frac{X_l}{Z_l}$
Correspondence problem

- Some points in each image will have no corresponding points in the other image.
  - i.e., cameras might have different field of view.
- A stereo system must be able to determine the image parts that should not be matched.
Correspondence problem (cont’d)

• Two main approaches:

  **Intensity-based**: attempt to establish a correspondence by matching image intensities.

  **Feature-based**: attempt to establish a correspondence by matching a sparse sets of image features.
Intensity-based Methods

- Match image sub-windows between the two images (e.g., using *correlation*).
Correlation-based Methods (cont’d)

• How to choose the window size $W$?
  – Too small a window may not capture enough image structure, and may be too noise sensitive (i.e., less distinctive, many false matches).
  – Too large a window makes matching less sensitive to noise (desired) but also harder to match.
Feature-based Methods

• Look for a feature in an image that matches a feature in the other.
  – Edge points
  – Line segments
  – Corners

• A set of geometric *features* are used for matching.
Intensity-based vs feature-based approaches

• Intensity-based methods
  – Provide a dense disparity map.
  – Need textured images to work well.
  – Sensitive to illumination changes.

• Feature-based methods:
  – Faster than correlation-based methods.
  – Provide sparse disparity maps.
  – Relatively insensitive to illumination changes.
Structured lighting

• Feature-based methods cannot be used when objects have smooth surfaces or surfaces of uniform intensity.
• Patterns of light can be projected onto the surface of objects, creating “interesting” points even in regions which would be otherwise smooth.
What is binocular stereo vision?

• A way of getting depth (3-D) information about a scene from two 2-D views (images) of the scene
What is binocular stereo vision?

• A way of getting depth (3-D) information about a scene from two 2-D views (images) of the scene

• Used by humans and animals
What is binocular stereo vision?

• A way of getting depth (3-D) information about a scene from two 2-D views (images) of the scene

• Used by humans and animals

• Computational stereo vision
  – Programming machines to do stereo vision
  – Studied extensively in the past 25 years
  – Difficult; still being researched
Fundamentals of Stereo Vision

• The goal of stereo analysis:
  – The inverse process: From 2-D image coordinates to 3-D scene coordinates
  – Requires images from at least two views
Fundamentals of Stereo Vision

- 3-D reconstruction
Fundamentals of Stereo Vision

- 3-D reconstruction
Fundamentals of Stereo Vision

- 3-D reconstruction
Fundamentals of Stereo Vision

- 3-D reconstruction
Fundamentals of Stereo Vision

- 3-D reconstruction
Fundamentals of Stereo Vision

- 3-D reconstruction
Fundamentals of Stereo Vision

- 3-D reconstruction
Fundamentals of Stereo Vision

- 3-D reconstruction
Recovering 3D from Images

• How can we automatically compute 3D geometry from images?
  – What cues in the image provide 3D information?
Visual Cues for 3D

• Shading
Visual Cues for 3D

- Shading
- Texture
Visual Cues for 3D

• Shading

• Texture

• Focus
Visual Cues for 3D

- Shading
- Texture
- Focus
- Motion
Visual Cues for 3D

- Shading
- Texture
- Focus
- Motion

- Others:
  - Highlights
  - Shadows
  - Silhouettes
  - Inter-reflections
  - Symmetry…
Depth from Convergence

Human performance: up to 6-8 feet

\[ d = \frac{c}{2 \tan\left(\frac{a}{2}\right)} \]
Depth from binocular disparity

P: converging point

C: object nearer projects to the outside of the P, disparity = +

F: object farther projects to the inside of the P, disparity = -

Sign and magnitude of disparity
Vergence

- Field of view decreases with increase in baseline and vergence
- Accuracy increases with baseline and vergence
Stereo Example

input image (1 of 2)  depth map  3D rendering
[Szeliski & Kang '95]
Stereo Example

left image  right image  depth map
Stereo Example
Ordering Constraint

- If an object $a$ is left on an object $b$ in the left image then object $a$ will also appear to the left of object $b$ in the right image.

Ordering constraint... ...and its failure
Correspondences

Match intensities sequentially between two scanlines.
Correspondences

Left scanline

Right scanline

Left occlusion

Right occlusion

Match

Match

Match
Search Over Correspondences

Three cases:

- Sequential – cost of match
- Left occluded – cost of no match
- Right occluded – cost of no match
Standard 3-move Dynamic Programming for Stereo

Dynamic programming yields the optimal path through grid. This is the best set of matches that satisfy the ordering constraint.
Graph cut

• One of the best algorithms
  – Produces high correct match scores
  – Computationally expensive
Stereo As a Pixel-Labeling Problem

- Let $P$ be a set of pixels, $L$ be a label set. The goal is find a labeling $f$ which minimize some energy. For stereo, the labels are disparities.
- The classic form of energy function is:

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

Data term \quad Smoothness term
Energy Function:

- The energy function $D_p(f_p)$ measures how appropriate a label is for the pixel $p$ given the observed data. In stereo, this term corresponds to the match cost or likelihood.
- The energy term $V_{p,q}(f_p, f_q)$ encodes the prior or smoothness constraint.

In stereo, the so-called Potts model is used:

$$V_{p,q}(f_p, f_q) = \begin{cases} 
0 & f_p = f_q \\
\rho_I(\Delta I) & \text{otherwise} 
\end{cases}$$
Stereo with Converging Cameras

• Stereo with Parallel Axes
  – Short baseline
    • large common FOV
    • large depth error
  – Long baseline
    • small depth error
    • small common FOV
    • More occlusion problems

• Two optical axes intersect at the Fixation Point
  – converging angle $\theta$
  – The common FOV Increases
Stereo with Converging Cameras

- Stereo with Parallel Axes
  - Short baseline
    - large common FOV
    - large depth error
  - Long baseline
    - small depth error
    - small common FOV
    - More occlusion problems

Two optical axes intersect at the Fixation Point
- converging angle $\theta$
- The common FOV Increases
Stereo with Converging Cameras

- Two optical axes intersect at the Fixation Point
  - converging angle $\theta$
  - The common FOV Increases

- Disparity properties
  - Disparity uses angle instead of distance
  - Zero disparity at fixation point
    - and the Zero-disparity horopter
  - Disparity increases with the distance of objects from the fixation points
    - $>0$ : outside of the horopter
    - $<0$ : inside the horopter

- Depth Accuracy vs. Depth
  - Depth Error $\propto$ Depth$^2$
  - Nearer the point, better the depth estimation
Stereo with Converging Cameras

- Two optical axes intersect at the Fixation Point
  - converging angle $\theta$
  - The common FOV Increases

- Disparity properties
  - Disparity uses angle instead of distance
  - Zero disparity at fixation point
    - and the Zero-disparity horopter
  - Disparity increases with the distance of objects from the fixation points
    - $>0$ : outside of the horopter
    - $<0$ : inside the horopter

- Depth Accuracy vs. Depth
  - Depth Error $\propto$ Depth$^2$
  - Nearer the point, better the depth estimation
Stereo with Converging Cameras

- Two optical axes intersect at the Fixation Point
  - converging angle $\theta$
  - The common FOV Increases

- Disparity properties
  - Disparity uses angle instead of distance
  - Zero disparity at fixation point
    - and the Zero-disparity horopter
  - Disparity increases with the distance of objects from the fixation points
    - $>0$: outside of the horopter
    - $<0$: inside the horopter

- Depth Accuracy vs. Depth
  - Depth Error $\propto$ Depth$^2$
  - Nearer the point, better the depth estimation
Stereo with Converging Cameras

- Two optical axes intersect at the Fixation Point
  - converging angle $\theta$
  - The common FOV Increases

- Disparity properties
  - Disparity uses angle instead of distance
  - Zero disparity at fixation point
    - and the Zero-disparity horopter
  - Disparity increases with the distance of objects from the fixation points
    - $>0$: outside of the horopter
    - $<0$: inside the horopter

- Depth Accuracy vs. Depth
  - Depth Error $\propto$ Depth$^2$
  - Nearer the point, better the depth estimation
Stereo with Converging Cameras

- Two optical axes intersect at the Fixation Point
  - converging angle $\theta$
  - The common FOV Increases

- Disparity properties
  - Disparity uses angle instead of distance
  - Zero disparity at fixation point
    - and the Zero-disparity horopter
  - Disparity increases with the distance of objects from the fixation points
    - $>0$: outside of the horopter
    - $<0$: inside the horopter

- Depth Accuracy vs. Depth
  - Depth Error $\propto \text{Depth}^2$
  - Nearer the point, better the depth estimation
For Each point \((x_l, y_l)\) in the left image, define a window centered at the point.
Correlation Approach

- ... search its corresponding point within a search region in the right image
Correlation Approach

- The disparity \((dx, dy)\) is the displacement when the correlation is maximum.
Correlation Approach

• Elements to be matched
  – Image window of fixed size centered at each pixel in the left image

• Similarity criterion
  – A measure of similarity between windows in the two images
  – The corresponding element is given by window that maximizes the similarity criterion within a search region

• Search regions
  – Theoretically, search region can be reduced to a 1-D segment, along the epipolar line, and within the disparity range.
  – In practice, search a slightly larger region due to errors in calibration
Correlation Approach

• PROS
  – Easy to implement
  – Produces dense disparity map
  – Maybe slow

• CONS
  – Needs textured images to work well
  – Inadequate for matching image pairs from very different viewpoints due to illumination changes
  – Window may cover points with quite different disparities
Feature-based Approach

• Features
  – Edge points
  – Lines (length, orientation, average contrast)
  – Corners

• Matching algorithm
  – Extract features in the stereo pair
  – Define similarity measure
  – Search correspondences using similarity measure and the epipolar geometry
Feature-based Approach

• For each feature in the left image...
Feature-based Approach

- Search in the right image... the disparity \((dx, dy)\) is the displacement when the similarity measure is maximum
Feature-based Approach

• PROS
  – Relatively insensitive to illumination changes
  – Good for man-made scenes with strong lines but weak texture or textureless surfaces
  – Work well on the occluding boundaries (edges)
  – Could be faster than the correlation approach

• CONS
  – Only sparse depth map
  – Feature extraction may be tricky
    • Lines (Edges) might be partially extracted in one
Reconstruction by Triangulation

• Assumption and Problem
  – Under the assumption that both intrinsic and extrinsic parameters are known
  – Compute the 3-D location from their projections, \( p_l \) and \( p_r \)

• Solution
  – **Triangulation**: Two rays are known and the intersection can be computed
  – **Problem**: Two rays will not actually intersect in space due to errors in calibration and correspondences, and pixelization
  – **Solution**: find a point in space with minimum distance from both rays
Questions

• Explain Graph cut method.
• Explain scenario when ordering constraint fails.
• Explain Correlation and Feature based approach.
• Write short note on Binocular stereo vision.
• What is correspondence problem in stereo vision?
Thank you!