Stereo Vision – A simple system

Dr. Gerhard Roth
Stereo

• Stereo
  • Ability to infer information on the 3-D structure and distance of a scene from two or more images taken from different viewpoints
  • Humans use only two eyes/images (try thumb trick)

• Two important problems in stereo
  • Correspondence and reconstruction

• Correspondence
  • What parts of left and right images are parts of same object?

• Reconstruction
  • Given correspondences in left and right images, and possibly information on stereo geometry, compute the 3D location and structure of the observed objects
Stereo

- Scene point
- Image plane
- Optical center
Stereo

Basic Principle: Triangulation

- Gives reconstruction as intersection of two rays
- Requires
  - Camera calibration
  - Point correspondence
Pinhole Camera Model

\[ x = -f \frac{X}{Z} \]

Image plane

\[ p = (x, y) \]

Virtual Image

\[ P = (X, Y, Z) \]
Simple Stereo System

• Left and right image planes are coplanar
  • Represented by $I_L$ and $I_R$

• So this means that all matching features are on the same horizontal line
  • So we can think of this as a 2D situation

Left

Right

scanline
Simple Stereo System (2D)

- Distance between centers of projection is called the baseline $T$
- Centers of projection of cameras $C_L$ and $C_R$
- Point $P$ in 3D space projects to $P_L$ and $P_R$
- $X_L$ and $X_R$ are co-ordinates of $P_L$ and $P_R$ with respect to principal points $C_L$ and $C_R$
- $Z$ is the difference between point $P$ and the baseline
  - $Z$ is called the depth
Simple Stereo System
Basic Stereo Derivations

Derive expression for $Z$ as a function of $x_1, x_2, f$ and $B$
Basic Stereo Derivations

Similar triangles $(P_L, P, P_R)$ and $(O_L, P, O_R)$

\[
\frac{T + x_l - x_r}{Z - f} = \frac{T}{Z}
\]

Define the disparity: $d = x_1 - x_2$

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

- \( D = ||x_1 - x_2|| \) measures the distance between corresponding points in two images
  - Normally disparity is stated as number of pixels
  - Clearly a particular simple stereo configuration has a maximum and minimum possible disparity

- Depth is inversely proportional to disparity

- If we compute the disparity for the entire images then we have a disparity map

- Display it as an image
  - Bright points have highest disparity (closest)
  - Dark points have lowest disparity (farthest)

- Disparity map is a 3D image
Disparity Map
Characteristics of Simple Stereo

- **FOV** is field of view of cameras
  - Overlap of the two cameras
- **Baseline** is a system parameter
  - It is also a tradeoff
- If B is the Baseline
  - Depth Error $\propto \frac{1}{B}$
- **PROS** of Longer baseline
  - better depth estimation
- **CONS**
  - smaller common FOV
  - Correspondence harder due to increased chance of occlusion
  - Occlusion means that a feature is visible in one image but not in another because something occludes it
Real-Time Stereo Systems

- There are a number of systems that can compute disparity maps.
- In practice, systems only work if there is texture in the regions that must be matched.
- Often such systems return sparse depth:
  - A few thousand images in regions where there is texture.
  - Do some interpolation when there is no texture.
- Point Grey research makes such a camera:
  - A successful Canadian company.
- Produces a variety of stereo cameras.
BumbleeBee
Example image from BumbleBee