Introduction to Vision and Robotics: Computer Vision

Active vision

Bastiaan Boom

All slides in this lecture are due to Bob Fisher
Active Vision

Key points:
• Acting to obtain information
• Eye movements
• Depth from motion parallax
• Extracting motion information from a spatio-temporal pattern
• Obtaining structure from motion
The importance of activity

• Computer vision is often approached as a problem of passive extraction of information from single images
• But most natural vision systems try to actively sample the visual scene
• Can solve some problems with active system that are hard with passive system
Segmenting  (Kruger, 1998)
Improving resolution

Making best use of fixed number of receptors e.g. 100x100 pixels  (Ballard, 1991)
Eye movements

• Increase the effective resolution by *saccade* movements of high resolution area (fovea)
• Creates impression that see complete detailed scene, but this is illusory
Eye movement patterns indicate attention and task

1. Describe room.
2. What was happening before?
3. People’s ages.
Eye movements

• Increase the effective resolution by *saccade* movements of high resolution area (fovea)
• Creates impression that see detailed scene, but this is illusory
• Task dependent, indicates attention
Eye movements and localisation

• Knowing where the eye/camera is pointing tells us the direction of objects of interest (requires proprioception to know relative angles)

• Can also extract depth through motion parallax
Demo
Motion Parallax

Closer objects change retinal position more when we move

Geometrically equivalent to binocular stereopsis
Motion perception

• Like parallax, much important information about the world comes from sensing visual motion
• E.g. breaking camouflage, sensing self motion, seeing what is happening…
• ‘Active vision’ sometimes taken to mean vision based on sequences of images
• Aim is to extract the flow-field:
Explanation
Optical Flow

Let's look at these constraints more closely:
- brightness constancy:  Q: what's the equation?
  \[ H(x, y) = I(x+u, y+v) \]
- small motion: (u and v are less than 1 pixel)
suppose we take the Taylor series expansion of I:
  \[ I(x+u, y+v) = I(x, y) + \frac{\partial I}{\partial x} u + \frac{\partial I}{\partial y} v + \text{higher order terms} \]
  \[ \approx I(x, y) + \frac{\partial I}{\partial x} u + \frac{\partial I}{\partial y} v \]
Optical Flow 2

Combining these two equations:

\[ 0 = I(x + u, y + v) - H(x, y) \]

\[ \approx I(x, y) + I_x u + I_y v - H(x, y) \]

\[ \approx (I(x, y) - H(x, y)) + I_x u + I_y v \]

\[ \approx I_t + I_x u + I_y v \]

\[ \approx I_t + \nabla I \cdot [u \ v] \]

shorthand: \( I_x = \frac{\partial I}{\partial x} \)

The x-component of the gradient vector.
Structure from motion

• Motion field contains information about the 3-d structure of objects (e.g. strong depth effect)

• If rigid body, and can track points, can geometrically recover structure of scene and movement of camera - active field in Computer Vision.
Structure from motion 2

What is this?
Structure from motion 3

3D structure emerges from pattern of motion
Structure from motion

Source images

Reconstructed from tracked feature points:
Attention from motion

- Can use flow-field to determine where to redirect the eyes – moving stimuli are *salient*
- Mechanism to determine new eye position:
  - Calculate the flow field
  - Enhance changes to detect new stimuli
  - Smooth to offset noise
  - Implement ‘winner-take-all’ connection to choose most salient movement, and inhibit return to same location
- Note that then have to solve problem of mapping visual target onto correct motion of camera