Motion and optical flow

Thurs Sept 17

Many slides adapted from S. Seitz, R. Szeliski, M. Pollefeys, S. Lazebnik

Last time

- Texture is a useful property that is often indicative of materials, appearance cues
- **Texture representations** attempt to summarize repeating patterns of local structure
- **Filter banks** useful to measure redundant variety of structures in local neighborhood
  - Feature spaces can be multi-dimensional
- Neighborhood statistics can be exploited to “sample” or **synthesize** new texture regions
  - Example-based technique

Today

- Optical flow: estimating motion in video
- Background subtraction

Video

- A video is a sequence of frames captured over time
- Now our image data is a function of space \((x, y)\) and time \((t)\)

Uses of motion

- Estimating 3D structure
- Segmenting objects based on motion cues
- Learning dynamical models
- Recognizing events and activities
- Improving video quality (motion stabilization)

Motion field

- The motion field is the projection of the 3D scene motion into the image
Motion parallax

http://psych.hanover.edu/KRANTZ/MotionParallax/MotionParallax.html

Motion field + camera motion

Length of flow vectors inversely proportional to depth $Z$ of 3D point

Figure from Michael Black, Ph.D. Thesis

Motion field + camera motion

Motion estimation techniques

- **Direct methods**
  - Directly recover image motion at each pixel from spatio-temporal image brightness variations
  - Dense motion fields, but sensitive to appearance variations
  - Suitable for video and when image motion is small

- **Feature-based methods**
  - Extract visual features (corners, textured areas) and track them over multiple frames
  - Sparse motion fields, but more robust tracking
  - Suitable when image motion is large (10s of pixels)

Optical flow

- Definition: optical flow is the *apparent* motion of brightness patterns in the image
- Ideally, optical flow would be the same as the motion field
- Have to be careful: apparent motion can be caused by lighting changes without any actual motion

Apparent motion $\neq$ motion field

Figure from Horn book
Problem definition: optical flow

How to estimate pixel motion from image H to image I?
- Solve pixel correspondence problem
  - given a pixel in H, look for nearby pixels of the same color in I

Key assumptions
- color constancy: a point in H looks the same in I
  - For grayscale images, this is brightness constancy
- small motion: points do not move very far

This is called the optical flow problem

Brightness constancy

Optical flow constraints

Optical flow equation

The aperture problem
The aperture problem

- How to get more equations for a pixel?
- **Spatial coherence constraint**: pretend the pixel's neighbors have the same \((u,v)\)

Solving the aperture problem

Prob: we have more equations than unknowns

\[
A \cdot d = b
\]

Solution: solve least squares problem

\[
\begin{bmatrix}
\sum I_x I_x & \sum I_x I_y \\
\sum I_x I_y & \sum I_y I_y
\end{bmatrix}
\begin{bmatrix}
u \\
v
\end{bmatrix} = -
\begin{bmatrix}
\sum I_x I_t \\
\sum I_y I_t
\end{bmatrix}
\]

Conditions for solvability

When is this solvable?

- \(Ra\) should be invertible
- \(Ra\) should not be too small
- eigenvalues \(\lambda_1\) and \(\lambda_2\) of \(Ra\) should not be too small
- \(Ra\) should be well-conditioned
- \(\lambda_1/\lambda_2\) should not be too large (\(\lambda_1\) = larger eigenvalue)

The barber pole illusion

http://en.wikipedia.org/wiki/Barberpole_illusion
Edge
- gradients very large or very small
- large $\lambda_1$, small $\lambda_2$

Low-texture region
- gradients have small magnitude
- small $\lambda_1$, small $\lambda_2$

High-texture region
- gradients are different, large magnitudes
- large $\lambda_1$, large $\lambda_2$

(Example applications with optical flow)

Today
- Optical flow: estimating motion in video
- Background subtraction

Video as an “Image Stack”

Can look at video data as a spatio-temporal volume
- If camera is stationary, each line through time corresponds to a single ray in space

Alyosha Efros, CMU
Background subtraction

- Simple techniques can do ok with static camera
- ...But hard to do perfectly

- Widely used:
  - Traffic monitoring (counting vehicles, detecting & tracking vehicles, pedestrians),
  - Human action recognition (run, walk, jump, squat),
  - Human-computer interaction
  - Object tracking
Frame differences vs. background subtraction

- Toyama et al. 1999

Mean Filter

- In this case the background is the mean of the previous n frames:
  \[
  B(x, y, t) = \frac{1}{n} \sum_{i=0}^{n-1} I(x, y, t - i)
  \]
  \[
  |I(x, y, t) - \frac{1}{n} \sum_{i=0}^{n-1} I(x, y, t - i)| > Th
  \]
- For n = 10:

  Estimated Background

  Foreground Mask

Median Filter

- Assuming that the background is more likely to appear in a scene, we can use the median of the previous n frames as the background model:
  \[
  B(x, y, t) = \text{median}\{I(x, y, t - i)\}
  \]
  \[
  |I(x, y, t) - \text{median}\{I(x, y, t - i)\}| > Th \text{ where } i \in \{0, \ldots, n-1\}.
  \]
- For n = 10:

  Estimated Background

  Foreground Mask

Average/Median Image

Background Subtraction

Alyosha Efros, CMU
Pros and cons

**Advantages:**
- Extremely easy to implement and use!
- All pretty fast.
- Corresponding background models need not be constant, they change over time.

**Disadvantages:**
- Accuracy of frame differencing depends on object speed and frame rate
- Median background model: relatively high memory requirements.
- Setting global threshold Th...

*When will this basic approach fail?*

Background mixture models

Idea: model each background pixel with a mixture of Gaussians; update its parameters over time.

So far: features and filters

Transforming images, gradients, textures, edges, flow