Motion Estimation

Outline

• What is motion estimation
• Optical flow
• Gradient-based approach
• Aperture problem
• Matching-based approach
• Local matching approach
• Global matching approach

Motion Estimation

Objective:
• Studies how to recover the apparent motion of objects in a scene based on a sequence of images:

Applications:
• Video compression
• Camera stabilization (remove jitter)
• Autonomous vehicle (collision detection & avoidance)
• Human-machine interaction (understanding gestures & facial expressions)
• Surveillance & monitoring (tracking & analyzing behaviors)

Motion Estimation (Cont’d)

Input:
• A sequence of images captured by a single camera
Output:
• Optical flow – a 2D array of velocities

Optical Flow

Definition:
• A velocity field in the image that transforms one image into the next image in a sequence
• Not uniquely determined

Optical flow ≠ projection of 3D motion
• Optical flow may be zero when the 3D object is in motion
  • Rotation of a sphere with uniform color
  • Optical flow may not be zero when the 3D object remains still
  • Movement of light changes object appearance

Brightness Constancy Constraint

Almost all motion estimation approaches assume the brightness constancy:
• The brightness of the same object does not change when the object shows up at different locations in 2 frames
  \[ I(x, y, t) = I(x + \Delta x, y + \Delta y, t + \Delta t) = I(x + u\Delta t, y + v\Delta t, t + \Delta t) \]
Different Motion Estimation Techniques

- **Gradient-based approach:**
  - Solve a linear approximation of the brightness consistency equation
  - Need to calculate the spatial-temporal derivative
  - Limitation:
    - Cannot handle large motions

- **Matching-based approach:**
  - Solve the brightness consistency equation directly
  - Convert the problem to an optimization problem
  - Limitation:
    - Can only provide quantified velocities

Gradient-based Approach

- Assume small motions:
  - Consider only the first-order terms in the Taylor series expansion of brightness consistency

\[
\frac{\partial I}{\partial t} + \frac{\partial I}{\partial x} \Delta x + \frac{\partial I}{\partial y} \Delta y + \frac{\partial I}{\partial t} \Delta t + \cdots \Rightarrow \frac{\partial I}{\partial x} \Delta x + \frac{\partial I}{\partial y} \Delta y + \frac{\partial I}{\partial t} \Delta t = 0
\]

\[
\Rightarrow \frac{\partial I}{\partial x} u + \frac{\partial I}{\partial y} v = -\frac{\partial I}{\partial t}
\]

- The last equation is the optical flow constraint
- One linear equation with 2 unknowns
- Under-determined; need additional constraints

Spatial-temporal Derivative

- When viewed from a small aperture, the motion direction of a contour is ambiguous
  - The motion parallel to the line cannot be inferred
  - Smoothness constraint is often applied
  - The velocity field in the image vary smoothly almost everywhere

Matching-based Approach

- Directly compute the motion of pixel \((p, q)\) through template matching
  - Create a template \(T\) centered at \((p, q)\) in frame \(k\)
  - Search around \((p, q)\) in frame \(k+1\) for patch that best matches to \(T\)
  - Areas used for searching depends on maximum object velocity

Aperture Problem

Local Matching Approach

- Pseudo-code:
  - For each pixel \((p, q)\) in the image:
    - Use area around \((p, q)\) in current frame as template.
    - For each displacement \((u, v)\):
      - Calculate match cost between the template and the local patch centered at \((p+u, q+v)\) in the next frame.
      - Find the displacement with the lowest cost.
    - Assign it to pixel \((p, q)\) of the optical flow.

- Complexity:
  - \(O(M \times N \times U \times V)\)
Result of Local Approach
- Horizontal velocity map
  - Background pixels are moving toward the center
  - Indicating the camera is zooming out
- Vertical velocity map
  - Ping-pong ball moving downward
  - Bat & hand moving upward
- The results are noise since different pixels are matched independently

Global Matching Approach
- Global matching approach optimize both the cost term & a smoothness term
- 4D cost space:
  - The value in voxel $(p, q, u, v)$ indicates the cost of assigning displacement $(u, v)$ to pixel $(p, q)$
- Cast the motion estimation problem to an optimization problem
- Different global optimization techniques can be used:
  - Dynamic programming
  - Genetic algorithm

Result of Global Approach
- The result is much cleaner with smoothness constraint
  - Motion for textured background is nicely captured
  - Motion for textureless areas, such as the ping-pong table, is still incorrect

Another Motion Example