Computer Vision for Recovering Information About Scene Geometry

Dan Huttenlocher
Computer Science Dept.
Geometry and Computer Vision

- Recovering “full” scene geometry from multiple images
  - depth or disparity maps from stereopsis
  - 3D surfaces or point sets from motion analysis

*Depth map from Boykov, Veksler and Zabih, 1998*
Partial Scene Geometry is Useful

- Two examples
  - background-plus-objects scene representation
    - emerging integrated approach to recovering scene geometry from multiple views
    - combines motion analysis and epipolar geometry from stereopsis
  - flexible models of objects
    - long history of graph-based models (late 60’s) for articulated objects composed of rigid subparts
    - recent algorithmic advances and applications to recognizing both static and dynamic configurations
Background-Plus-Objects

- Most scenes have a background that appears to be a “plane at infinity”
  - changes in viewpoint (locally) do not reveal 3D geometric structure of this background

- Objects are then parts of scene where 3D geometry is apparent from images
  - closer than the “plane at infinity”
    - causes motion parallax between views
  - moving independently in the scene
    - motion that does not obey parallax
Illustrating Background-Plus-Objects

- **Top**
  - sequence with moving camera and car

- **Bottom**
  - background aligned (left)
  - parallax motion aligned (right)

*From Anandan and Irani, 1997*
Background-Plus-Objects Modeling

- Camera centers $C_t$ and $C_s$, image point $P_i$
  - projection onto “virtual” reference plane $\pi$
  - re-projection onto image plane

*From Irani, Anandan and Weinshall, 1998*
Background-Plus-Objects Modeling

- Virtual reference plane maps to each image plane via projective transformation
  - modeled using projective transformation between two images
    - affine approximation often reasonable

- Remaining image motion (parallax) in static scene due to height above reference plane
  - meets epipolar constraint
  - don’t actually need to find epipoles
    - important because not always reliably computable
Applications of this Geometric Model

- Replacing objects in video
  - e.g. advertisements in live broadcast of sporting events (e.g., Sarnoff)

- Detection of “interesting” objects in video
  - removing independently moving objects
  - hypothesizing objects to follow

Tracking results from Bell, Felzenszwalb and Huttenlocher, 1999
Representing Flexible Objects

- Modeling articulated objects such as people, some animals, many manmade items
  - “stick figure” models (e.g., Bregler, Adelson)
  - “cardboard cutout” models (e.g., Black, Yacoob)
- Models composed of parts and connections between parts
  - commonly use graph-based representation with parts as nodes and connections as edges
  - connection cost reflects deformation of model parts with respect to an “ideal configuration”
Illustrating Flexible “Cardboard” Models

- Relative geometry captured by connections
  - simple appearance model of each part

From Felzenszwalb and Huttenlocher, 1999
Finding Good Configurations of Parts

- For \( n \) parts and \( m \) possible locations of each part, \( O(m^n) \) configurations
  - cost measures how well each part matches image and how much model is deformed

- Seek low cost configuration
  - can be set up in a principled manner as a MAP estimation problem (Bayesian framework)

- In general find local rather than global minimum due to high computational cost
  - fine for tracking – “nearby” configuration
New Result: Efficient Global Solution

- Many flexible objects are tree-structured
  - no cycles in the graph
    - “hand not connected to foot”

- An $O(m^2n)$ method for such objects
  - modification of standard Viterbi-style dynamic programming – but still not practical, $m^2$ is huge

- Second level of dynamic programming yields $O(mn)$ method that runs in a few seconds
  - based on a novel generalization of distance transforms from point sets to functions
Flexible Model Matching Examples

- Simple part models can be mismatched due to color change
  - primarily measuring deformation geometry
Flexible Model Matching Examples

- Currently no model of occluded parts
  - find best match that accounts for all parts (allowing overlap)
The Geometry You Need?

Variety of techniques in computer vision for extracting geometric information

- “full” 3D recovery only one possibility
  - recent improvements in stereo algorithm accuracy
- partial 3D recovery using background-plus-objects representation
  - objects with “depth” or moving “independently”
    - potentially more stable than methods that explicitly recover epipolar geometry
- geometric model-based matching for identifying people and their activities