Layered Image Representation

Chuck Dyer


Motivation

+ Standard flow assumes optical flow is smooth
+ Bad things happen at occlusion boundaries
+ Instead, decompose image sequence into a set of overlapping layers
+ Each layer is smooth in its own motion
Problem Definition

Figure 1: This figure shows the decomposition of an image sequence consisting of a hand moving in front of a checkerboard background. The conventional method of representing motion is by a dense motion field with motion discontinuities at object boundary. The layered representation describes these objects with smooth motions, and discontinuities in opacity. The apparent motion discontinuities result when the layers are composited according to the occlusion relationship between objects.

Example Input Video

a. Frame 0 of MPEG flower garden sequence. b. Frame 15 of MPEG flower garden sequence. c. Frame 30 of MPEG flower garden sequence.
a. Optic flow

Figure 9: a) Velocity estimates appear as samples in space. b) Standard optic flow algorithms impose some smoothing on the velocity estimates in order to deal with noise. c) Regularization algorithms model the velocity field as a set of smooth regions while allowing for sharp breaks at model boundaries. d) Shows the representation that we wish to attain. The velocity samples are explained by two affine models (straight lines) and discontinuities are explained by the occlusion of one object by the other.
Figure 2: This figure shows the technique used in motion segmentation. Affine motion models are determined by regression on the dense motion fields and the regions are assigned to minimize the error between the motion expected by the models and the estimated dense motion.
Motion Vectors vs Motions Hypothesis

- There are a number of different motion hypotheses available
  - In theory, each of these hypothesis corresponds to a distinct motion in the video
- Each pixel is assigned to the motion hypothesis that most closely approximates its motion vector
- This segments the frame into distinct regions, one for each motion hypothesis
Motion Hypothesis Generation

- For each region want motion hypothesis that best represents all pixel motions in that region
- Least squares fit to find best affine motion parameter in a region
- First iteration initialized with small blocks

Motion Hypothesis Refinement

- K-Means used to cluster motion hypotheses
- $K$ unknown
- Empty clusters removed
- Large clusters split to maintain minimum $k$ value
Region Segmentation

- For each pixel compare hypotheses to dense motion vectors
- Find closest hypothesis
- Group all pixels represented by a motion hypothesis into a region
- Pixels with large error unassigned
- Hypotheses without membership removed

Region Adjustment

- Region Splitter
  - Assumes areas with same motion are connected
  - Disconnects areas within a region are split into separate regions
  - Increases number of hypotheses for k-means
- Region Filter
  - Small regions give poor motion estimates
  - Remove all regions with area below threshold
  - Disconnected objects with same motion will be merged at next segmentation step
Algorithm Summary

• Dense motion estimation, region segmentation, and motion estimation performed for all pairs of consecutive frames
• For first pair, segmentation initialized to blocks and k-Means initialized to lattice in 6D affine space
• Subsequent frame pairs initialized with final segmentation and motion hypotheses from previous frame pair

Layer Synthesis

• Motion estimates relate each frame only to the previous frame
  • Frames are projected onto first video frame
  • Cumulative projection kept in 3x3 transformation matrix
• Layers are not necessarily ordered similarly between frames
  • Assume largest layer is background
• Median taken of all values projected to each pixel in final image
Affine Motion Segmentation

Video Mosaic of Each Layer

- Flower Bed regions in all images aligned
Motion Compensation

- Aligned regions

Tree  Flower Bed  House

a) Frame 1 aligned  b) Frame 15 as reference  c) Frame 30 aligned
3 Major Layers
Application: Video Synthesis

- Layered decomposition captures spatial coherence of object motion and temporal coherence of object shape and texture in a few semantically-meaningful layers
- Synthesize new sequences from the layers
c. Synthesize frame 30