Structure From Motion
The promise, the results, the hand-wringing

Last several days we’ve set up:

- The relationship between camera motion, and image pairs --- either in the discrete motion or the continuous motion case.
- When we’ve talked about applications, they’ve been to solve for the R, T
- Today we’ll think about 3d reconstructions.

Direct Motion Estimation Using BCCE

\[
(u, v) = (Tx, Ty) = \frac{u_w}{Z} + u_{\text{ext}} + \alpha_x + \beta_x + \gamma_y + \delta_y
\]

\[
v = \frac{v_w}{Z} + \alpha_y + \beta_y - \gamma_x - \delta_x
\]

\[1_u + 1_v + 1_z = 0\]

We talked about ambiguity at the end of the last class.

Ambiguity is minimized if camera is moving towards a point in the field of view, when there are sharp depth discontinuities in the scene, and when the camera field of view is large.

Lets think about 3d models?
Structure from motion results

Input frames (3 of 5)

Texture-mapped model
Shaded model

Output 3-D model

(Images courtesy of Luc Van Gool)

Why is a surface better than a point cloud? Should we solve first for the cloud then the surface?

Alternating minimization.
Assume that you know Z.
Solve for U, V, W, alpha, Beta, gamma.
Assume you know U, V, W, alpha, Beta, gamma, solve for 1/Z

Convergence?
Speed?

Once common initial guess: assume Z is infinite.

Structure from Motion

- Not so good. Image derivatives are very noisy. When solving for Z, there is only one equation per pixel.
- Solutions. Have a model of the depth.
- Locally planar? (Where are the planes?)
- Another alternating minimization.
  - Assume you know which pixels are on the same plane.
  - Solve for the best motion/planes (Model Fitting)
  - For each pixel, calculate the error it would have had if it was on another plane. Reassign the pixels to the depth plane that best fits its data. (Reassignment)
  - Loop up to "Assume".
- A version of "Expectation Maximization".

Crap.

On some objects, segmentation algorithm does not converge, or converges to a solution that does not correspond to depth discontinuities. The boundaries (of the badly segmented regions) still indicate likely region boundaries.
• Exploit consistency of the world.

• Another alternating minimization.
  - Assume you know which pixels are on the same plane.
  - Solve for the best motion/planes (Model Fitting)
  - For each pixel on boundaries between two regions, calculate the error it would have had if it was on the neighboring plane. Reassign the pixels to the depth plane that best fits its data. (Reassignment)
  - Loop up to “Assume”.

• A (more bastardized) version of “Expectation Maximization”.

In contrast to global reassignments, region growing maintains locality information. Small patches whose boundaries respect scene discontinuities is all that is really necessary for motion computation.

### Segmentation III.I

In contrast to global reassignments, region growing maintains locality information. Small patches whose boundaries respect scene discontinuities is all that is really necessary for motion computation.

### Piecewise Planar Reconstructions

• Assume smoothness, and solve for “mesh” representation. Represent the depth of whole scene by the depth of a set of control points. The depth of a pixel between control points is a linear combination of the depth at each control point.

\[
\begin{align*}
u &= -\frac{U + xW}{Z} + \alpha \gamma - \beta (x^2 + 1) + \gamma y \\
u &= \frac{1}{Z}(-U + xW) + \alpha \gamma - \beta (x^2 + 1) + \gamma y \\
u &= (k_1 \lambda_1 + k_2 \lambda_2 + k_3 \lambda_3)(-U + xW) + \alpha \gamma - \beta (x^2 + 1) + \gamma y \\
\lambda_2 &= 1/Z
\end{align*}
\]

### Differential Video Linking

Solve for the best motion in each frame. “Chaining” them together: Requires solution for translation magnitude. When are there errors?
Optimal Equiform transformation defines scale factor and rigid motion relating the two viewpoints, so that information from both motion fields can be combined.

Special Effects in The Matrix

- Recently used to create *Bullet Time* special-effect in *The Matrix* (www.whatisthematrix.com)
- Linear array of cameras replaces moving camera.
- Green screen makes segmentation easy.