Capturing and Modelling 3D Data

Miles Hansard

Centre for Intelligent Sensing
Queen Mary University of London
My research objectives

• Reconstruct 3D scenes from images
• Model 3D scene-statistics
• Understand human binocular vision
• Applications:
  – View synthesis, augmented reality
  – Re-rendering for 3D displays
  – Active robot vision
Parallax-based 3D reconstruction

- Traditional method, using multiple views
- Cameras must be calibrated
- Main problems are mismatched/unmatched regions
- Can use the same images (textures) for rendering
Time-of-flight cameras

- Active depth-sensing devices (infrared)
- Optically similar to a camera (lens, CCD, etc)
- Estimates distance to scene along each ray
- Works by measuring phase shift of reflection
- Very low spatial resolution $(176 \times 144)$
- No colour!
Stereo vs TOF reconstruction

• Need the TOF/stereo transformation for rendering
• TOF 3D contains a lot of local errors
  – Sensor noise
  – Scattering surfaces
• Stereo 3D often has global errors
  – Overall distortion of the scene
  – Caused by lack of camera calibration
• The two reconstructions are complementary
Mixed camera-systems

- Each system provides two 3-D reconstructions
- One TOF camera + two high-resolution RGB cameras
- Several of these TOF+2RGB systems can be combined
Projective alignment: theory

• TOF/stereo viewpoints and methods are different
• How are the two reconstructions related in 3D?
• Not just rotation, translation and scale
• But *flat* surfaces are flat in *both* reconstructions
• Flatness-preserving transformations are projective
• Examples of projectively equivalent shapes in 2D:
Projective alignment: algorithm

- Align the uncalibrated stereo reconstruction to the TOF data, by 3D projective transformation
- Can be done by a linear method (SVD based)
- Now any TOF point can be projected into the images, so the model can be rendered
Reprojection results

- To associate a colour with each 3-D point:
  - Backproject the TOF pixels to XYZ in the scene
  - Reproject them into RGB views (using estimated cameras)

Left, TOF, and right images, colour-coded by depth
Reprojection results - detail
Wide-baseline example
Resolution mismatch
Problems at depth-boundaries
From TOF to dense depth
Full four-system configuration

Setup is designed for 360° capture of human figures
Multi-system alignment

- Each TOF+2RGB system has been calibrated
- We now align the four stereo reconstructions
- One system is chosen to be the reference-frame
- We use \((4-1)\) rigid + \(4 \times 2\) projective transformations
Meshing

• Complete figure + room reconstructions give rise to difficult meshing problems
• Use the TOF data to pre-segment the figure
• Background can then be meshed easily, using a local method
• The figure can be meshed using a global method (e.g. Poisson Reconstruction)
• Also allows one or more figures to be placed in an alternative background
• No completely satisfactory solution yet!
Rendering

• Mesh representation is rendered using standard graphics hardware (OpenGL shaders)
• An additional advantage of the alignment method is that multiple textures are available
• Blended, or switched according to relationship between the surface and viewpoint
• Models are rendered in real-time, using live TOF+RGB data.
Segmented figure

The cuboids represent one of the TOF+2RGB systems
Rendered figure & background

Note sharp boundary between figure and background
Top view of a three-system reconstruction
Reprojected figure-mesh

3D mesh, reprojected into one of the texture-images
Figure reconstructions
Collaborators

- Radu Horaud, Georgios Evangelidis, Michel Amat
  - INRIA Grenoble, France
- Seungkyu Lee, Ouk Choi
  - Samsung Advanced Institute of Technology, South Korea
Random scenes

• Highly **cluttered environments**, in which the depth-structure is not dominated by any particular object
• E.g. forests (important for evolution!)
• Very wide-angle laser range-scan:
Geometric model

- Left: Green rectangle must be empty for visibility
- Right: Both must be empty for binocular visibility
- Red line is the scene-boundary (empty in front)
Scene and observer models

- If scene has a **Poisson** distribution of intensity $\lambda$, then distance to visible object has **exponential** distribution $F$:
  \[
  \text{prob}(s|\lambda) = F(s, 2\varepsilon\lambda)
  \]
- This is not realistic in typical imaging conditions
- Peak of distribution along any ray would be at zero (the optical centre)
- Impose a **scene-boundary**, at random distance from the observer, according to Gaussian distribution $G$:
  \[
  \text{prob}(t|\theta, \mu, \sigma) = G \left( t, \frac{\mu}{\cos \theta}, \frac{\sigma}{\sin \theta} \right)
  \]
Binocular joint-distribution

- Total distance to first object along a ray is the scene-penetration \textit{plus} the distance to the scene-boundary

- Probability of a sum \( \rho = s + t \) is the \textit{convolution} of the densities \( F(s) \) and \( G(t) \)

- This is a re-parameterized \textit{ex-Gaussian} distribution \( H \):
  \[
  \text{prob}(\rho|\theta) = H \left( \rho, 2\epsilon\lambda, \frac{\mu}{\cos \theta}, \frac{\sigma}{\cos \theta} \right)
  \]

- Tend to see fewer distant objects, in clutter

- A point is binocularly visible if \textit{both} left and right rays are unobstructed:
  \[
  \text{prob}(\rho_L,\rho_R) = \text{prob}(\rho_L|\theta_L) \times \text{prob}(\rho_R|\theta_R)
  \]
Fits to forest data

- The parameters to be estimated are $2\varepsilon\lambda$, $\mu$ and $\sigma$

$$\text{prob}(\rho | \theta) = H \left( \rho, 2\varepsilon\lambda, \frac{\mu}{\cos \theta}, \frac{\sigma}{\cos \theta} \right)$$

- Each fit defines a one-parameter family, ranging from coarse/dense to fine/sparse

- Maximum Likelihood fits, by numerical minimization:
Monocular conditional-distribution

- Joint-distribution determines several other distributions
- But the joint-distribution is not *observable*; it is parameterized by *scene*-distances
- More useful to ask: given an image point in one view, where is the corresponding point in the other view?
- This is the *conditional* distribution along an *epipolar line*:

\[
\text{prob}(\theta_R | \theta_L) = \text{prob}(\rho_L, \rho_R) \times J_R(\theta_R)/S_R(\theta_L)
\]

- Jacobian $J_R(\theta_R)$ and normalizing constant $S_R(\theta_L)$ ensure that $\text{prob}(\theta_R | \theta_L)$ is a proper probability density
- Tend to see more of a scene, ‘per pixel’, in the distance
Prediction of re-projected forest data

- The image densities can be used as Bayesian priors for image-matching in cluttered scenes.
- These are *predictions*, not fits, given the estimated density:
Future work

- More on statistical scene-models
- Attempt to collect outdoor range scans
- Extend TOF work to Kinect devices
- Combination of cameras and inertial sensors
- Biological vision, geometric models of human stereopsis
- HTML5 data-visualization, including mobile devices