Urban 3D Reconstruction

- Capture ground-level models
- Video-only as an alternative to LIDAR
- Must process hours of video efficiently:
  Partial video capture of Chapel Hill:
  2,500,000 frames @ 30fps = 23 hours

Plane-Sweeping Stereo with Multiple Sweeping Directions

- Simple, robust, multi-view approach
- Accounts for slanted surfaces
- Efficiently implemented on GPU
- Uses 3D points from SfM as a prior.
- Takes advantage of urban scenes, but performs no worse on general scenes.

Identifying Sweeping Directions

Urban scenes have 3 dominant directions:

**Ground normal:**
\[ V = \text{vertical (found from vanishing point)} \]
\[ M = \text{camera motion} \]
Ground: \[ G = \frac{(V \times M) \times M}{\| (V \times M) \times M \|} \]

**Façade normals:**
- Look for 2 façade normals which are vertical & orthogonal
- Determined by one rotation about vertical axis
  - Use 3D points from SfM
  - For each \( \Theta \) in \([0,90)\)
    - rotate points by \( \Theta \) about \( V \)
    - build histograms in x and y
    - compute entropy
  - record \( \Theta \) with minimum entropy

Incorporating Plane Priors

Use distribution of 3D points on planes as a prior for stereo matching:

\[
P(\Pi_m|C(x, y)) = \frac{P(C(x, y)|\Pi_m)P(\Pi_m)}{P(C(x, y))}
\]
\[
P(C(x, y)|\Pi_m) = e^{-\frac{C(x, y)}{\sigma}}
\]
Minimizing the negative log-likelihood yields:

\[
\hat{\Pi}(x, y) = \arg\min_{\Pi_m} \left\{ -\log e^{-\frac{C(x, y)}{\sigma}} P(\Pi_m) \right\}
\]
\[
= \arg\min_{\Pi_m} \left\{ C(x, y) - \sigma \log P(\Pi_m) \right\}
\]
Quick sweep: only test plane hypotheses with highest prior probability. 5.26Hz \(\rightarrow\) 20.0Hz

Results

- 11 images, 512x384 grayscale, 3 sweeping directions, 282 total plane hypotheses, processing rate 5.26 Hz
- 1474 frame reconstruction, 11 images in stereo, 512x384 grayscale, 48 plane hypotheses (quick sweep), processing rate 20.0 Hz

Acknowledgements:
DARPA UrbanScape and DTO VACE programs, as well as D. Nistér, A. Akbarzadeh and the rest of the UNC & UKY UrbanScape team.