Shape Anchors for Data-driven Multi-view Reconstruction

Motivation

Task: Build dense 3D reconstructions from videos.

We use shape anchors to combine single- and multi-view cues. These are image patches whose geometry is “obvious” – they are so distinctive that we can recognize their dense 3D shapes using a database search.

Finding shape anchors

Idea: Predict dense geometry based on a single image; keep only the shape interpretations that have good photo-consistency evidence.

For each patch, we search for the best matches in an RGB-D video database [2] and transfer the 3D shape of the matches that agree with multi-view stereo points.

We train a random forest to distinguish correct vs. incorrect shape predictions.

We use shape anchors in conjunction with multi-view stereo [1] to estimate dense geometry. Our inputs are videos of real-world scenes with handheld camera motion.

Transferred geometry

- Shape estimates are coarse: less accurate than stereo but dense
- Classifier is trained to find matches with ≤10cm mean error
- Number of shape anchors varies with scene content

Interpreting geometry

We use the geometry provided by shape anchors for reconstruction tasks.

- We propagate shape anchor matches and expand the predicted geometry using contextual information.
- We use occlusion cues to remove erroneous shape anchors and to refine their geometry.

Transferring from very similar scenes:

When the dataset contains similar scenes, the result is often dense. Below, our database contains other apartment units in the same building, with similar layouts and objects.

Common failure cases:

- In many scenes, few shape anchors are found, so the results are sparse.
- Transfers can be too coarse for fine-scale geometry, or there are mismatches.