Shape Anchors for Data-driven Multi-view Reconstruction

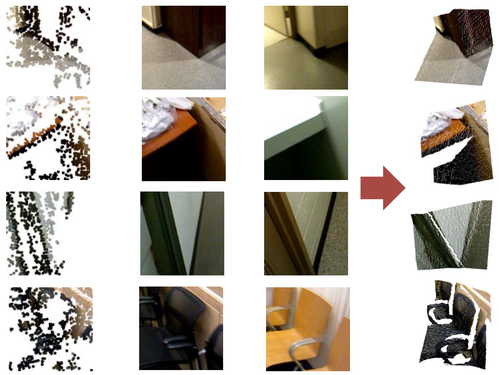
Andrew Owens, MIT
Jianxiong Xiao, Princeton*
Antonio Torralba, MIT
William Freeman, MIT

* Work done while at MIT

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.



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

[1] Y. Furukawa, J. Ponce. Accurate, Dense, and Robust Multiview Stereopsis. Trans. PAMI 2010.

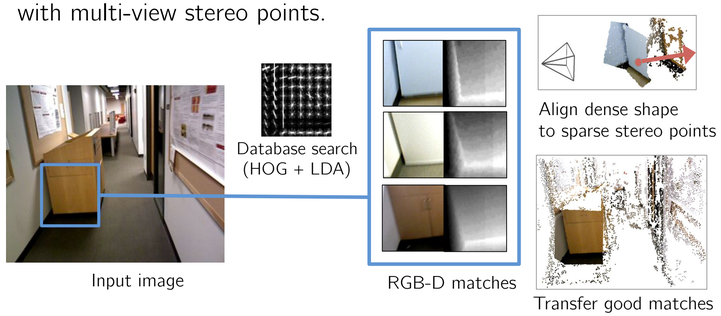
[2] J. Xiao, A. Owens and A. Torralba. SUN3D: A Database of Big Spaces Reconstructed using SfM and Object Labels. ICCV 2013.

[3] D. Fouhey, A. Gupta, and M. Hebert. Data-Driven 3D Primitives for Single Image Understanding. ICCV 2013.

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.



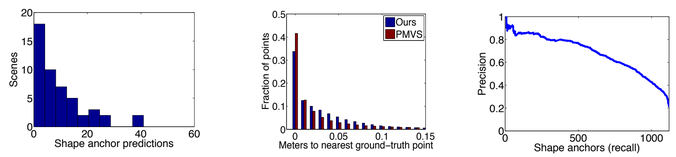
We train a random forest to distinguish correct vs. incorrect shape matches. Features include:

Multi-view evidence: Histograms of nearest-neighbor distances between multi-view stereo points and dense shape prediction.

Image evidence: Convolution score, relative window positions, 3D consistency of similar patches (similar to [3]).

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



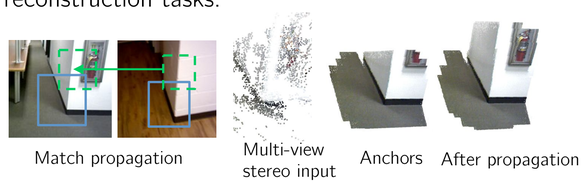
Acc: median distance to ground-truth point.

Comp: fraction of ground-truth points within 3cm of reconstruction.

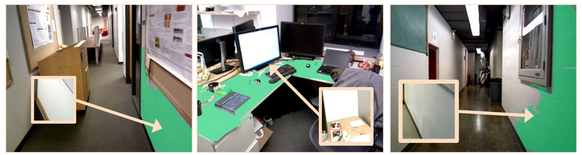
	Win. Acc.	Win. Comp.
Anchors	3.8cm	63.4%
PMVS	2.1cm	44.2%
	Full Acc.	**Full Comp.**
Transferred	1.8cm	21.6%
PMVS	1.6cm	27.4%
Combined	1.8cm	35.6%

Interpreting geometry

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



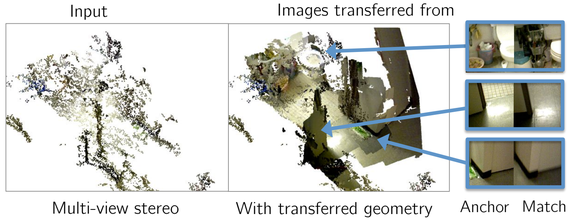
- We propagate shape anchor matches and expand the predicted geometry using contextual information.



- We use shape anchors to guide a plane-fitting/ segmentation algorithm.
- We use occlusion cues to remove erroneous shape anchors and to refine their geometry.

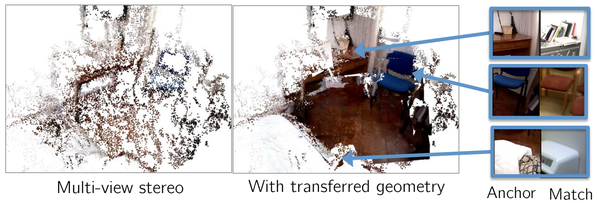
Transferring from very similar scenes:

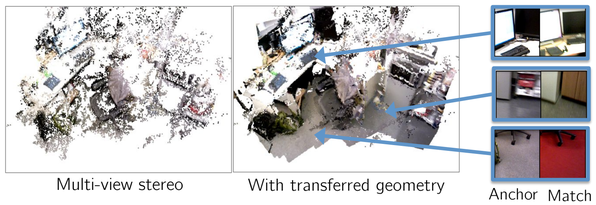
When the dataset contains similar scenes, the result is often dense. Below, our database contains **other apartment units in the same building**, with similar layouts and objects.

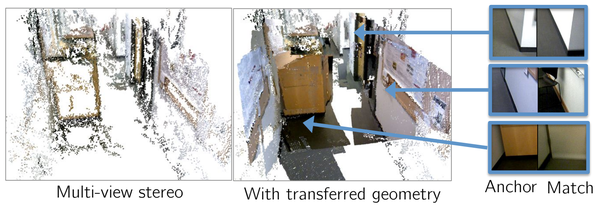


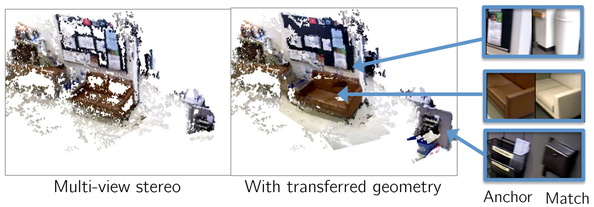
Results

Visualizations of geometry reconstructed using shape anchors; scenes chosen for large number of anchors.









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.