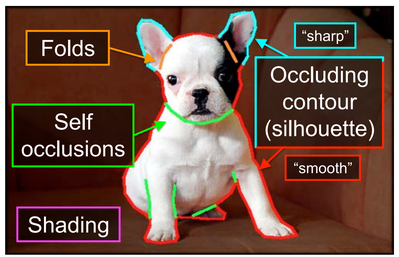
Boundary Cues for 3D Object Shape Recovery

Kevin Karsch[1], Zicheng Liao[1], Jason Rock[1], Jonathan T. Barron[2], Derek Hoiem[1]
[1]University of Illinois at Urbana-Champaign [2]University of California Berkeley

We evaluate how a variety of single-image "boundary" cues affect shape reconstruction for *real pictures*.

Image from PASCAL



Annotated boundary cues

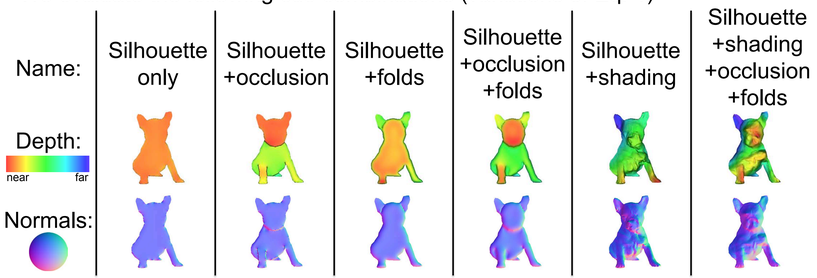
Shape is computed through a novel SFS optimization (extending [1,2]):

$$\underset{Z,R,L}{\text{minimize}}\ \delta_{sfc} f_{sfc}(Z) + \delta_{selfocc} f_{selfocc}(Z) + \delta_{fold} f_{fold}(Z) + \delta_{reg} f_{reg}(Z) + \delta_{sfs} f_{sfs}(g(R) + h(L))$$
$$\text{subject to } c_{sfs}(Z, R, L) = 0 \quad (1)$$

(Z: shape, R: reflectance, L: illumination)

Our experiments aim to find:
- How accurate are reconstructions?
- Are the shapes recognizable?
- Which cues are "most important"?
- Directions for improving SFS

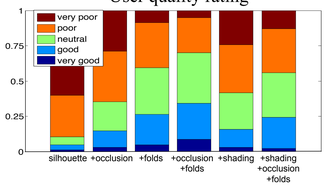
We evaluate the following cue combinations (variations of Eq. 1)



[1] J Malik and D Maydan. "Recovering three-dimensional shape from a single image of curved objects." *TPAMI*, 1989.
[2] JT Barron and J Malik. "Color constancy, intrinsic images, and shape estimation." *ECCV*, 2012.

Reconstruction Quality

User quality rating



Percentage of times that one method rates higher than another

- **Subjects consistently prefer reconstructions using self occlusion and folds to other cue combinations**
- Reconstructions using *both* self occlusions and folds rated higher than using one or none of these cues by a ratio of about 5:1
- No method achieves good/very good ratings consistently

Surface normal (N) and depth (Z) error on MIT Intrinsic Dataset

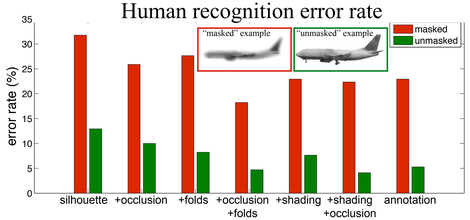
	N-MSE°	Z-MAE°	N-MSE†	Z-MAE†
sih	0.573	25.533	0.521	25.637
+selfocc	0.565	25.198	0.498	25.342
+folds	0.496	25.562	0.501	25.400
+occ+folds	0.487	25.161	0.482	24.983
+shading	0.874	38.968	0.310	25.793
+shading+occ+folds	0.574	27.379	0.350	24.492

° Optimization weights set uniformly (=1)
† Optimization weights trained from held out set

- **Quantitative errors roughly consistent with preferences found in user study**
- Uniform and trained weights produce perceptually similar results yet significant differences in error
- Training improves performance on synthetic data; no training data available yet for real images (e.g. PASCAL)

Shape Recognition

Human recognition error rate



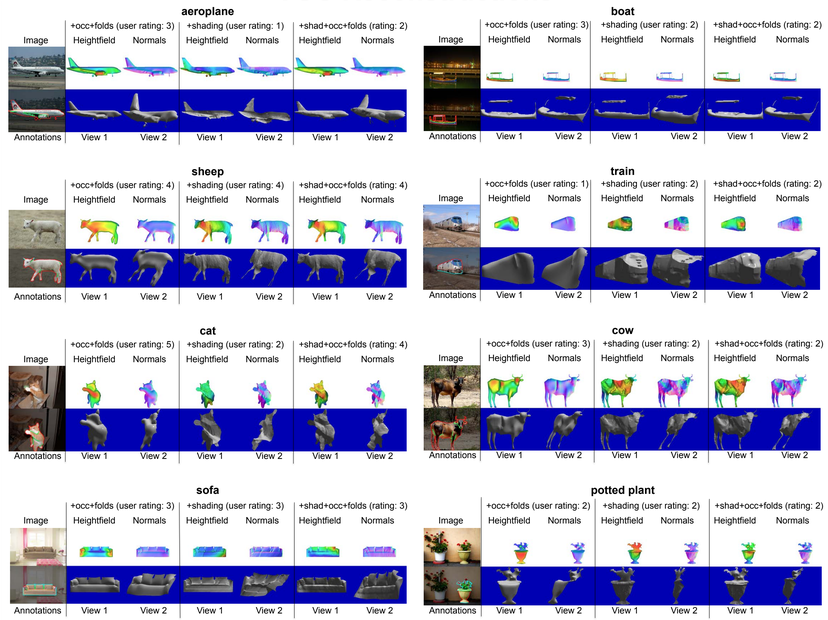
- **Both shading and boundary cues aid human recognition of object classes**, but silhouette alone is already a very revealing cue
- Viewing only annotation lines is nearly as helpful (gestalt effect)
- Masking object boundary makes classification much more difficult (e.g. "masked" results)

Automatic recognition rate on PASCAL VOC with different features

	RGBD kernel	PHOW
rgb	55.29	-
+occ+folds+rgb	**70.00**	-
+shading+occ+folds+rgb	62.35	-
sih	47.06	45.29
+selfocc	65.88	**54.12**
+folds	51.76	47.65
+occ+folds	65.88	51.76
+shading	48.24	41.76
+shading+occ+folds	52.94	42.94

- **Approximate depth is very useful for building classification features**
- Boundary cues appear to drive classification; high frequency shading features seem to confuse classifiers
- RGB features (results above double lines) greatly improve accuracy

VOC Reconstructions



Conclusions

- Boundary cues provide coarse shape detail, complementing fine-scale detail provided by shading cues
- Automatically detecting boundary cues like self occlusions and folds can greatly improve single-image shape reconstruction
- Features extracted from shape reconstructions improve existing object descriptors/classifiers (e.g. RGB+*approximate* depth)