Unfolding an Indoor Origami World
Extended Data Analysis
David F. Fouhey, Abhinav Gupta, Martial Hebert

We now present some additional analysis of the results. In particular, we answer (1) are these results by chance? and (2) how sensitive is the final solution to the potential trade-off parameters.

1 Statistical Analysis of the Results

To evaluate the stability of the method, we use bootstrapped confidence intervals[1], which characterize how a statistic (e.g., median, % Pixels $< 11.25°$) varies while making few assumptions about the data. Throughout the analysis, we compare the Origami World system with its informative point of comparison, the Manhattan-world version of 3D Primitives.

Variability of Results We show intervals (plus statistics) computed on both systems in Table 1. The results are quite stable, plus or minus about a degree and percent or so.

Variability of Performance Gain: The above does not does not fully analyze the relative performance of the systems as the results are paired (i.e., for each pixel, there's a prediction from each system)[2]. We therefore compute intervals on the *relative performance* per pixel. This properly measures the uncertainty in the question "how much gain do I expect". We report results in Table 2. In all cases, 0 gain (i.e., no difference) is never in the confidence interval. Note that the median is non-linear; thus, while the median prediction from Origami World is much better (1.3°) than the median prediction from 3D Primitives, the median *gain* considerably underestimates this (as most of the gains will be small).

Table 1: Bootstrapped confidence intervals for erorrs for each method.

	Summary Statistics (Lower Better)					
	Mean		Median		RMSE	
3D Primitives	36.0	(35.3,36.9)	20.5	(19.5,22.0)	49.4	(48.7,50.3)
Origami World	**35.1**	**(34.2,36.0)**	**19.2**	**(17.9,20.5)**	**48.7**	**(47.8,49.5)**

	% Good Pixels (Higher Better)					
	11.25°		22.5°		30°	
3D Primitives	35.9	(34.5,37.3)	52.0	(50.6,53.2)	57.8	(56.6,58.9)
Origami World	**37.6**	**(36.0,39.1)**	**53.3**	**(51.9,54.8)**	**58.9**	**(57.7,60.2)**

Table 2: Bootstrapped confidence intervals on gains of Origami World over the next-best method, 3D Primitives. Differences may not match Table 1 due to rounding.

	Summary Statistics						%Good Pixels					
	Mean		Median		RMSE		11.25°		22.5°		30°	
Gain	0.9	(0.2,1.7)	0.3	(0.1,0.5)	1.2	(0.4,1.6)	1.4	(0.5,2.3)	0.7	(0.2,1.7)	1.2	(0.3,2.2)

[1] We use 10,000 replicates and the bias-corrected accelerated method [1] to compute the intervals. Since errors within images are correlated, we use a block resampling scheme that samples entire images at once (i.e., each dataset replicate is created by sampling over images, and then including all the pixels within an image). Intuitively, our procedure is checking how the statistic varies as one samples datasets.

[2] Not accounting for the pairing makes the gain look weaker than it is intuitively because you're double counting intrinsic variability (e.g., due to image difficulty) when computing intervals for both method. The double counting is especially pronounced since 3DP is a cue in the Origami World system.

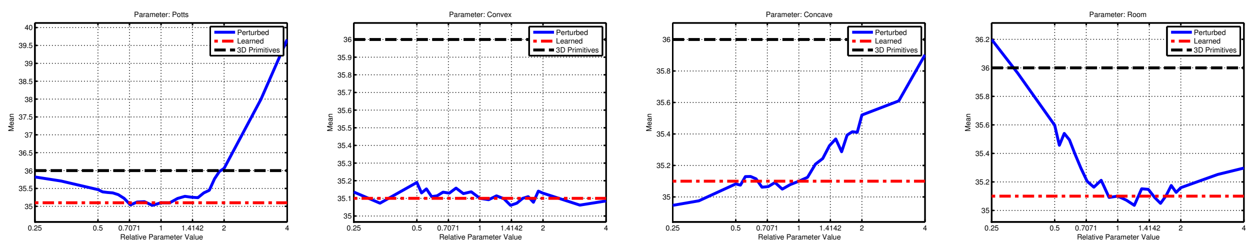
2 Stability Analysis

In learning, we hold the weight on the local evidence/3D Primitive potential fixed and vary the other parameters. We now show an assessment of the stability of the method when we perturb each learned trade-off parameters ($\{\lambda_k\}$, $\{\alpha_l\}$) of the method from the found values. We then plot performance on the validation set. If the paramaters deviate by a factor of 50% (i.e., 66% or 150% of their optimal value), the method still outperforms the baseline by a strong margin.

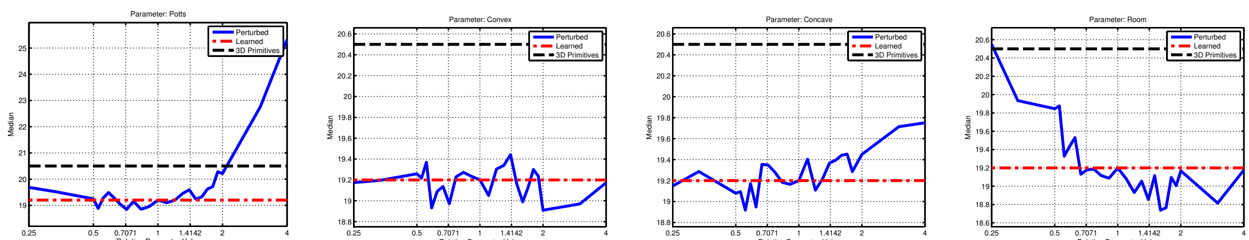
Of course, as the parameters are increasingly perturbed, performance drops. Of all the parameters, the method is most sensitive to oversmoothing (potts parameter). Nonetheless, the method still outperforms all the baselines with the smoothing (and other) potential set to nearly 2× the optimal value. For other parameters, this can be set to nearly 4× or 0.25× the optimal value.

Potts	Convex	Concave	Room Unary

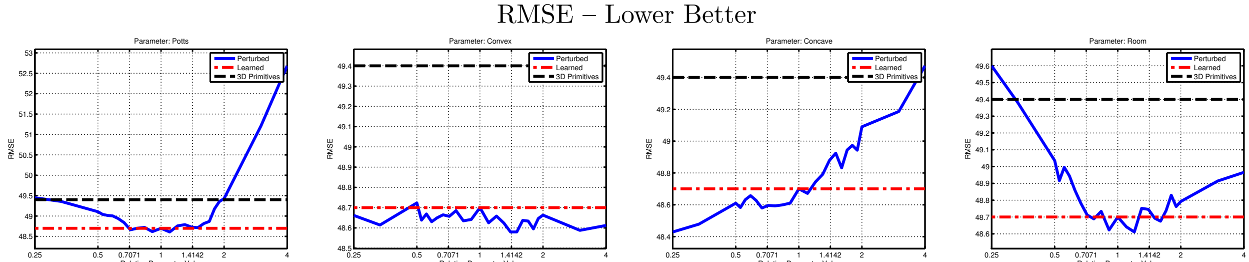
Mean Error – Lower Better



Median Error – Lower Better

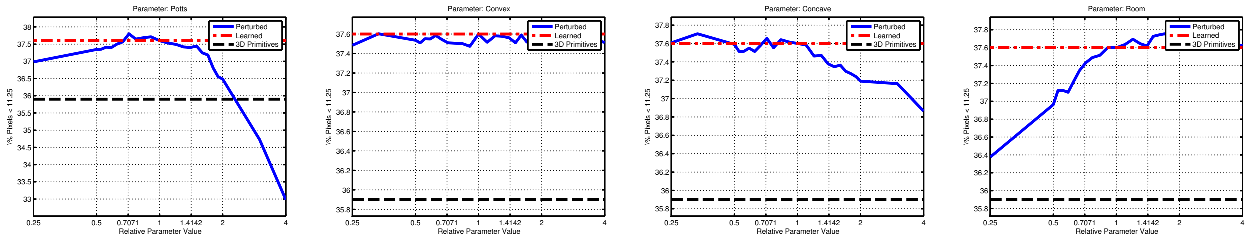


RMSE – Lower Better

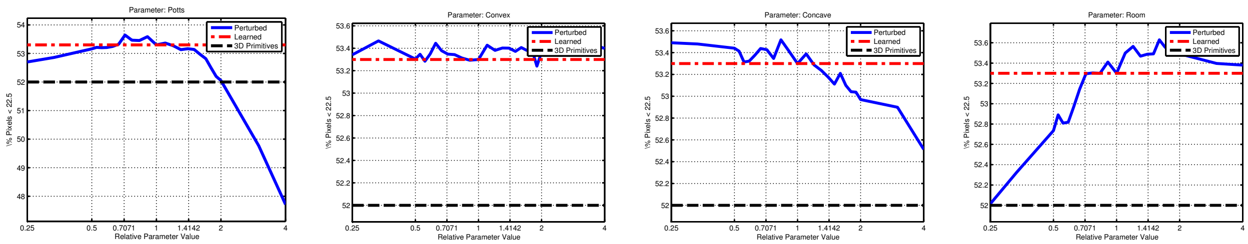


Potts	Convex	Concave	Room Unary

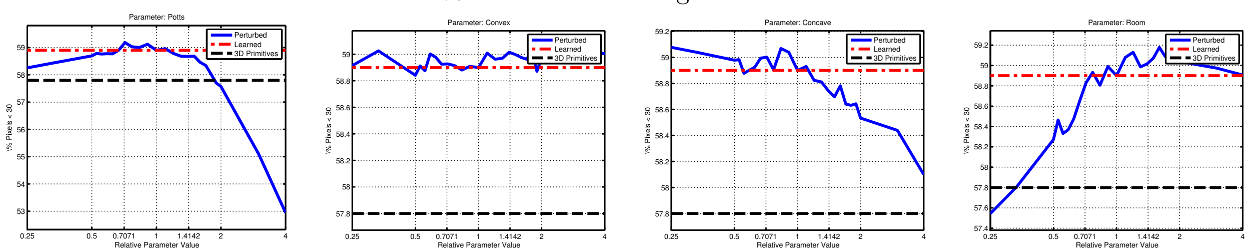
% Pixels $< 11.25^\circ$ – Higher Better



% Pixels $< 22.5^\circ$ – Higher Better



% Pixels $< 30^\circ$ – Higher Better



References

[1] B. Efron. Better bootstrap confidence intervals. *Journal of the American Statistical Association*, 82(397):171–185, 1987.