Scaling Laws in Natural Scenes and the Inference of 3D Shape

Brian Potetz,[1, 2] Tai Sing Lee[1, 2]

[1]Computer Science Department, [2]Center for the Neural Basis of Cognition, Carnegie Mellon University

Abstract

- Using a database of natural images and coregistered laser scans, we explore the statistical relationship between natural images and their underlying 3D surfaces (range images), and how that relationship changes over scale. We begin by analyzing and evaluating a previous model of scaling in natural scenes used in a technique known as shape recipes[3]. We then advance a new model of scaling based on the statistics of natural scenes.
- We apply our new model by extending the shape recipe technique for enhancing low-resolution range data using a full-resolution color image. We then evaluate our method using natural scenes with ground-truth high-resolution range data. We demonstrate a two-fold improvement in performance over the previous method.
- We provide theoretical insight into the depth cues and statistical regularities that contribute to our model, and study their relative strengths. We show that, in natural scenes, shadow-based depth cues may contribute more to linear shape-from-shading than traditional Lambertian shading cues.

Introduction

Traditional physics-based methods for inferring depth from single images require many restrictive assumptions, while overlooking many useful statistical regularities in real scenes.

Little is known about the joint statistics of images and 3D surface shape in real scenes, despite the potential benefits of statistical models and the growing success of statistical methods in vision

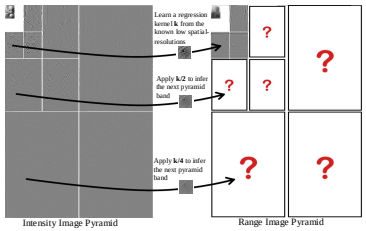
Using a database of natural images and coregistered laser scans, we explore the statistical correlations between images and range data in natural scenes, and we examine how that relationship changes over spatial scale.

One model for scaling in the range/intensity relationship has already been advanced implicitly in a technique known as shape recipes.

Shape recipes is a technique for enhancing low-resolution range data using a high-resolution intensity image, which is one immediate application for statistical models of scaling in the range/intensity relationship.

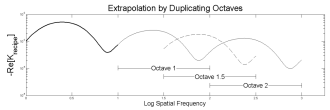
Shape recipes works by learning a relationship between 3D shape and monocular cues in the low spatial scales and applying that relationship to the high frequencies. One advantage of this approach is that hidden variables important to inference from monocular cues (such as illumination direction) may be implicitly learned from the low-resolution range and intensity images.

Analysis of Shape Recipes



Shape Recipe Algorithm for the super-resolution of range images given a full-resolution intensity image[3]:
1. Decompose the full-resolution intensity image and low-resolution range image into steerable pyramids.
2. For each orientation, learn a linear regression kernel from the low-resolution intensity band to the low-resolution range band.
3. Apply this kernel to the high resolution bands, (reducing the kernel amplitude by half for each octave)

Assumptions of Shape Recipes:
1. The learned regression kernel should change slowly across spatial scales.
2. The regression kernel amplitude is reduced by half for each octave.



Assumption 1: We show that this is mathematically self-contradictory. Shape recipes can be viewed as a single linear operation:

$$Z_{high}(r,\theta) = I(r,\theta)K_{recipe}(r,\theta)$$

K_{recipe} approximates the true kernel $K = Z_{high}I^*/II^*$

We show that K_{recipe} extrapolates the known low-spatial-frequencies of K by replicating a single low-frequency octave into the high frequencies, as shown below. The result is that intermediate octaves of K_{recipe} do not resemble the replicated octave. This means that shape recipes learn so many parameters that they cannot all simultaneously generalize to the higher spatial frequencies.

Assumption 2 is based on the linear lambertian shading equations, and can be tested by exploring the natural scene statistics:

$$i(x,y) \approx \frac{\partial}{\partial y} z(x,y) \qquad I(r,\theta) \approx 2\pi i r \cos(\theta) Z(r,\theta)$$

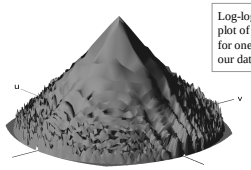
Statistics of Real Scenes

Using a laser-scanner with integrated color photosensor, we collected a database of single-material scenes. We studied the first order statistics of these scenes, including covariance between intensity and distance.

In the paper, we show that ZI^* (which is the Fourier transform of the cross-covariance matrix cov[z,i]) can be modeled by $ZI^*(r,\theta) \approx B(\theta)/r^{\alpha}$, where α was measured at 3.65 ± 0.14, and $B(\theta)$ is a parameter of the scene which depends on lighting direction and other scene properties.

II^* was fit by $A/r^{2.60}$, which means that K may be approximated as

$$K(r,\theta) \approx B(\theta)/r$$



Log-log polar plot of |real[ZI]| for one scene in our database.

The power-law drop-off of $1/r$ verifies the second assumption of shape recipes.

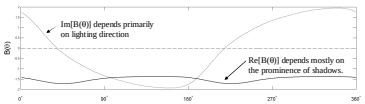
However, the linear lambertian assumption that this assumption was based on predicts that the real part of K should be zero.

Yet, in our database, we found that the **real part of K was often the stronger component.**

In a previous paper, we found that dark areas of an image were more likely to be further away[7]. We attributed this to cast shadows: object interiors and cavities are more likely to lie in shadow than object exteriors, and the interiors are further from the observer. In the dataset used in this paper, the correlation between brightness and distance was $\rho = 0.37$. This direct correlation contributes to the real part of K.

The $1/r$ drop-off rate of the real part of K is a new finding. We believe that this happens because concavities with smaller apertures but equal depths tend to be darker.

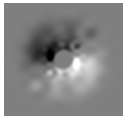
Below is a plot of $B(\theta)$ for one scene. The real part of $B(\theta)$ (in black) is uniformly negative, corresponding to the correlation between distance and darkness. The imaginary part, as predicted by the linear lambertian model, reaches its minima at the illumination direction (at the extreme left, almost 180^0, for this scene).

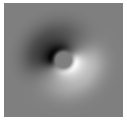


Extending Shape Recipes

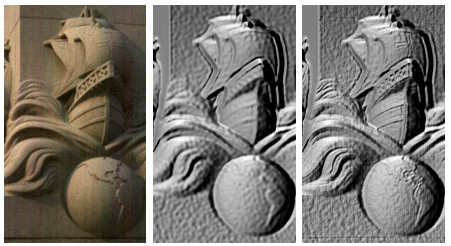
Our model of $K \approx B(\theta)/r$ suggests that we can extrapolate K by measuring $B(\theta)$ in the low spatial-frequencies. This gives us $K_{powerlaw}$, which we can use to estimate Z by:

$$Z_{powerlaw} = IK_{powerlaw} + Z_{low}$$

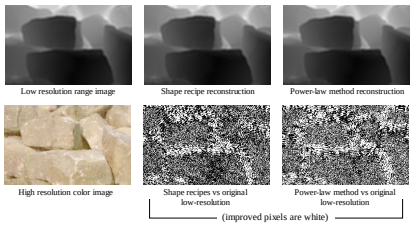

Example Imag[K_{recipe}]


Example Imag[$K_{powerlaw}$]

Inference Results in Real Scenes


High Resolution Color Image | Low Resolution Range Image (Lambertian rendering) | High Resolution Power-law Reconstruction

For each of the 28 single-materials scenes in our dataset, we down-sampled our high-resolution range images to create low-resolution range images, and used shape recipes and our power-law method to attempt to reconstruct the original high-resolution range images. Shape recipes improved 21 of the 28 images, for an average of 1.3% less mean squared error than the down-sampled low-resolution range image. The power-law method improved 26 of the 28 images, for a 2.2% reduction of mean squared error. We estimate that most of the remaining error (95%) is either due to nonlinearities in the intensity/range relationship or noise in the acquisition of our images.



Shadow cues more powerful than shading cues?

As noted previously, the imaginary part of ZI, which is expected to come from shading, was often smaller than the real part of ZI, which is expected to come from shadow cues.

What are the relative contributions of these cues?

We re-ran our algorithm with either the real or imaginary part of $K_{powerlaw}$ set to zero.
- The algorithm is 27% as effective with Imag[K] alone (no real component)
- The algorithm is 72% as effective with Real[K] alone (no imag component)

Thus, in our database, using only linear cues, shadow-based cues appear to be more powerful than shading-based cues.

Subdividing our database according to environment supports this theory. Shadow cues may be expected to be most powerful in scenes of foliage, where regions deeper into foliage are more heavily shadowed. Urban scenes contain fewer shadowed crevises and concavites, and many continuous, smooth surfaces where shading is most effective.

Environment:	% improvement from shadow-based cues alone
foliage (8 scenes)	96%
rocky terrain (8 scenes)	76%
urban scenes (building fascades and statues, 12 scenes)	35%

Linear shape from shading has received some attention since its introduction[6] as a fast, reasonably accurate, and biologically plausible method of inferring depth from shading. However, linear shape inference methods have so far utilized only Lambertian shading cues. Linear shape-from-shading algorithms could be adapted to exploit the correlations in the real part of ZI, which we believe are related to cast shadows.

References

[1] J. E. Cryer, P. S. Tsai and M. Shah, "Integration of shape from shading and stereo," Pattern Recognition, 28(7):1033--1043, 1995.

[2] W. T. Freeman, E. H. Adelson, "The design and use of steerable filters," IEEE Transactions on Pattern Analysis and Machine Intelligence, **13**, 891--906 1991.

[3] W. T. Freeman and A. Torralba, "Shape Recipes: Scene representations that refer to the image," Advances in Neural Information Processing Systems 15 (NIPS), MIT Press, 2003.

[4] C. Q. Howe and D. Purves, "Range image statistics can explain the anomalous perception of length," Proc. Nat. Acad. Sci. U.S.A. **99** 13184--13188 2002.

[5] M. S. Langer and S. W. Zucker, "Shape-from-shading on a cloudy day," J. Opt. Soc. Am. A **11**, 467--478 (1994).

[6] A. Pentland, "Shape information from shading: a theory about human perception," *Spat Vis*. Vol. 4, pp. 165--82, 1989.

[7] B. Potetz, T. S. Lee, "Statistical correlations between two-dimensional images and three-dimensional structures in natural scenes," J. Opt. Soc. Amer. A, **20**, 1292--1303 2003.

[8] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," IJCV 47(1/2/3):7--42, April-June 2002.

[9] A. Torralba, A. Oliva, "Depth estimation from image structure," IEEE Transactions on Pattern Analysis and Machine Intelligence. 24(9): 1226--1238 2002.

[10] A. Torralba and W. T. Freeman, "Properties and applications of shape recipes," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003.

[11] C. W. Tyler, "Diffuse illumination as a default assumption for shape-from-shading in the absence of shadows," J. Imaging Sci. Technol. **42**, 319--325 1998.