Chapter 1

Scene statistics and 3D surface perception

Brian Potetz and Tai Sing Lee

1.1 Introduction

The inference of depth information from single images is typically performed by devising models of image formation based on the physics of light interaction and then inverting these models to solve for depth. Once inverted, these models are highly underconstrained, requiring many assumptions such as Lambertian surface reflectance, smoothness of surfaces, uniform albedo, or lack of cast shadows. Little is known about the relative merits of these assumptions in real scenes. A statistical understanding of the joint distribution of real images and their underlying 3D structure would allow us to replace these assumptions and simplifications with probabilistic priors based on real scenes. Furthermore, statistical studies may uncover entirely new sources of information that are not obvious from physical models. Real scenes are affected by many regularities in the environment, such as the natural geometry of objects, the arrangements of objects in space, natural distributions of light, and regularities in the position of the observer. Few current computer vision algorithms for 3D shape inference make use of these trends. Despite the potential usefulness of statistical models and the growing success of statistical methods in vision, few studies have been made into the statistical relationship between images and range (depth) images. Those studies that have examined this relationship in nature have uncovered meaningful and exploitable statistical trends in real scenes which may be useful for designing new algorithms in surface inference, and also for understanding how humans perceive depth in real scenes [32, 18, 46]. In this chapter, we will highlight some results we have obtained in our study on the statistical relationships between 3D scene structures and 2D images, and discuss their implications on understanding human 3D surface perception and its underlying computational principles.

1.2 Correlation between brightness and depth

To understand the statistical regularities in natural scenes that allow us to infer 3D structures from their 2D images, we carried out a study to investigate the correlational structures between depth and light in natural scenes. We collected a database of coregistered intensity and high-resolution range images (corresponding pixels of the two images correspond to the same point in space) of over 100 urban and rural scenes. Scans were collected using the Riegl LMS-Z360 laser range scanner. The Z360 collects coregistered range and color data using an integrated CCD sensor and a time-of-flight laser scanner with a rotating mirror. The scanner has a maximum range of 200 m, and a depth accuracy of 12 mm. However, for each scene in our database, multiple scans were averaged to obtain an accuracy under 6 mm. Raw range measurements are given in meters. All scanning is performed in spherical coordinates. Scans were taken of a variety of rural and urban scenes. All images were taken outdoors, under sunny conditions, while the scanner was level with ground. Typical spatial resolution was roughly 20 pixels per degree.

To begin to understand the statistical trends present between 3D shape and 2D appearance we start our statistical investigation by studying simple linear correlations within 3D scenes. We analyzed corresponding intensity and range patches, computing the correlation between a specific pixel (in either image or range patch) with other pixels in the image patch or the range patch, obtained with equation,

$$\rho = \mathrm{cor}[X, Y] = \frac{\mathrm{cov}[X, Y]}{\sqrt{\mathrm{var}[X]\mathrm{var}[Y]}} \tag{1.1}$$

The patch size is 25 x 25 pixels, slightly more than 1 degree visual angle in each dimension, and in calculating the covariance, both of the image patch and the range patch have subtracted their corresponding means across all patches.

One significant source of variance between images is the intensity of the light source illuminating the scene. Differences in lighting intensity result in changes to the contrast of each image patch, which is equivalent to applying a multiplicative constant. In order to compute statistics that are invariant to lighting intensity, previous studies of the statistics of natural images (without range data) study the logarithm of the light intensity values, rather than intensity itself [48, 8]. Zero-sum linear filters will then be insensitive to changes in image contrast. Likewise, we take the logarithm of range data as well. As explained by Huang *et. al.* [19], a large object and a small object of the same shape will appear identical to the eye when the large object is positioned appropriately far away and the small object is close. However, the raw range measurements of the large, distant object will differ from those of the small

object by a constant multiplicative factor. In the log range data, the two objects will differ by an additive constant. Therefore, a zero-sum linear filter will respond identically to the two objects.

Figure 1 shows three illustrative correlation plots. Figure 1a shows the correlation between intensity at center pixel $(13, 13)$ and all of the pixels of the intensity patch. Figure 1b shows the correlation between range at pixel $(13, 13)$ and the pixels of the range patch. We observe that neighboring range pixels are much more highly correlated with one another than neighboring luminance pixels. This suggests that the low frequency components of range data contain much more power than in luminance images, and that the spatial Fourier spectra for range images drops off more quickly than for luminance images, which are known to have roughly $\frac{1}{f}$ spatial Fourier amplitude spectra [37]. This finding is reasonable because factors that cause high-frequency variation in range images, such as occlusion contours or surface texture, tend to also cause variation in the luminance image. However, much of the high-frequency variation found in luminance images, such as shadow and surface markings, are not observed in range images. These correlations are related to the relative degree of smoothness that is characteristic of natural images versus natural range images. Specifically, natural range images are in a sense more smooth than natural images. Accurately modeling these statistical properties of natural images and range images is essential for robust computer vision algorithms and for perceptual inference in general. Smoothness properties in particular are ubiquitous in modern computer vision techniques for applications such as image denoising and inpainting [36], image-based rendering [52], shape from stereo [38], shape from shading [31], and others.

Figure 1c shows correlation between intensity at pixel $(13, 13)$ and the pixels of the range patch. There are two important effects here. The first is a general vertical tilt in the correlation plot, showing that luminance values are more negatively correlated with depth at pixels lower within the patch. This result is due to the fact that the scenes in our database were lit from above. Because of this, surfaces facing upwards were generally brighter than surfaces facing downwards, and conversely, brighter surfaces were more likely to be facing upwards than darker surfaces. Thus, when a given pixel is bright, the distance to that pixel is generally less than the distance to pixels slightly lower within the image. This explains the increasingly negative correlations between the intensity at pixel $(13, 13)$ and the depth at pixels lower within the range image patch.

What is more surprising in Figure 1c is the correlation between depth and intensity is significantly negative. Specifically, the correlation between the intensity and the depth at a given pixel is roughly -0.20. In other words, brighter pixels tend to be closer to the observer. Historically, physics-based approaches to shape from shading have generally concluded that shading cues offer only relative depth information. Our findings show there is also an absolute depth cue available from image intensity data that could help to

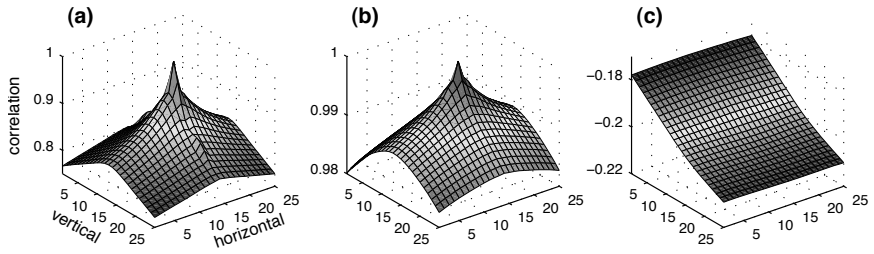


FIGURE 1.1: **A.** Correlation between intensity at central pixel $(13, 13)$ and all of the pixels of the intensity patch. Note that pixel $(1, 1)$ is regarded as the upper-left corner of the patch. **B.** Correlation between range at pixel $(13, 13)$ and the pixels of the range patch. **C.** Correlation between intensity at pixel $(13, 13)$ and the pixels of the range patch. For example, correlation between intensity at central pixel $(13, 13)$ and lower-right pixel $(25, 25)$ was -0.210.

more accurately infer depth from 2D images.

This empirical finding regarding natural 3D scenes may be related to an analogous psychophysical observation that, all other things being equal, brighter stimuli are perceived as being closer to the observer. This psychophysical phenomenon has been observed as far back as Leonardo da Vinci, who stated, "*among bodies equal in size and distance, that which shines the more brightly seems to the eye nearer.*" [26]. Hence, we referred to our empirical correlation as the *da Vinci correlation.* Artists sometimes make use of this cue to help create compelling illusions of depth [50, 39].

In the last century, psychophysicists validated da Vinci's observations in rigorous, controlled experiments [1, 2, 43, 3, 6, 41, 47, 23, 53]. In psychology literature, this effect is known as *relative brightness* [27]. Numerous possible explanations have been offered as to why such a perceptual bias exists. One common explanation is that light coming from distant objects has a greater tendency to be absorbed by the atmosphere [5]. However, in most conditions, as in outdoor sunlit scenes, the atmosphere tends to scatter light from the sun directly towards our eyes, making more distant objects appear brighter under hazy conditions [28]. Furthermore, our database was acquired under sunny, clear conditions, under distances insufficient to cause atmospheric effects (maximum distances were roughly 200m). Other explanations of a purely psychological explanation have also been advanced [43]. While these might be contributing factors for our perceptual bias, they cannot account for empirical observations of real scenes.

By examining which images exhibited the da Vinci correlation most strongly, we concluded that the major cause of the correlation was primarily due to shadow effects within the environment [32]. For example, one category of images where correlation between nearness and brightness was most strong was

images of trees and leafy foliage. Since the source of illumination comes from above, and outside of any tree, the outermost leaves of a tree or bush are typically the most illuminated. Deeper into the tree, the foliage is more likely to be shadowed by neighboring leaves, and so nearer pixels tend to be brighter. This same effect can cause a correlation between nearness and brightness in any scene with complex surface concavities and interiors. Because the light source is typically positioned outside of these concavities, the interiors of these concavities tend to be in shadow, and more dimly lit than the object's exterior. At the same time, these concavities will be further away from the viewer than the object's exterior. Piles of objects (such as figure 1.2) and folds in clothing and fabric are other good examples of this phenomenon.

To test our hypothesis, we divided the database into urban scenes (such as building facades, and statues) and rural scenes (trees and rocky terrain). The urban scenes contained primarily smooth, man-made surfaces with fewer concavities or crevices, and so we predicted these images to have reduced correlation between nearness and brightness. On the other hand, were the correlation found in the original dataset due to atmospheric effects, we would expect the correlation to exist equally well in both the rural and urban scenes. The average depth in the urban database (32 meters) was similar to that of the rural database (40 meters), so atmospheric effects should be similar in both datasets. We found that correlations calculated for the rural dataset increased to -0.32, while those for the urban dataset are considerably weaker, in the neighborhood of -0.06.

In Langer and Zucker [22], it was observed that for continuous Lambertian surfaces of constant albedo, lit by a hemisphere of diffuse lighting and viewed from above, a tendency for brighter pixels to be closer to the observer can be predicted from the equations for rendering the scene. Intuitively, the reason for this is that under diffuse lighting conditions, the brightest areas of a surface will be those that are the most exposed to the sky. When viewed from above, the peaks of the surface will be closer to the observer. Although these theoretical results have not been extended to more general environments, our results show that, in natural scenes, these tendencies remain, even when scenes are viewed from the side, under bright light from a single direction, and even when that lighting direction is oblique to the viewer. In spite of these differences, both phenomena seem related to the observation that concave areas are more likely to be in shadow. The fact that all of our images were taken under cloudless, sunny conditions and with oblique lighting from above suggests that this cue may be more important than at first realized.

It is interesting to note that the correlation between nearness and brightness in natural scenes depends on several complex properties of image formation. Complex 3D surfaces with crevices and concavities must be present, and cast shadows must be present to fill these concavities. Additionally, we expect that without diffuse lighting and lighting interreflections (light reflecting off of several surfaces before reaching the eye), the stark lighting of a single point light source would greatly diminish the effect [22]. Cast shadows, complex 3D

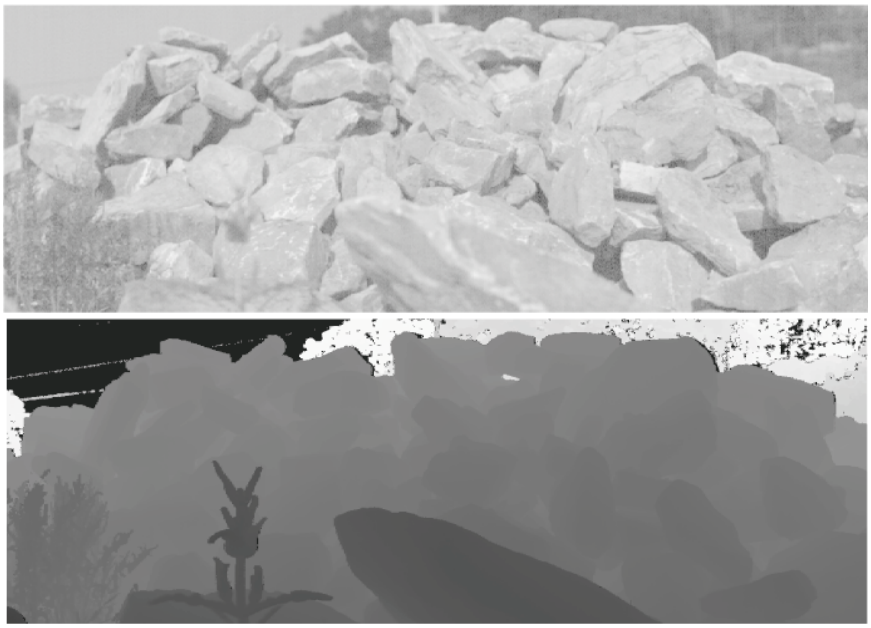


FIGURE 1.2: An example color image (top) and range image (bottom) from our database. For purposes of illustration, the range image is displayed by displaying depth as shades of gray. Notice that dark regions in the color image tend to lie in shadow, and that shadowed regions are more likely to lie slightly further from the observer than the brightly lit outer surfaces of the rock pile. This example image from our database had an especially strong correlation between closeness and brightness.

surfaces, diffuse lighting, and lighting interreflections are all image formation phenomena that are traditionally ignored by methods of depth inference that attempt to invert physical models of image formation. The mathematics required for these phenomena are too cumbersome to invert. However, taken together, these image formation behaviors result in the simplest possible relationship between shape and appearance: an absolute correlation between nearness and brightness. This finding illustrates the necessity of continued exploration of the statistics of natural 3D scenes.

1.3 Characterizing the Linear Statistics of Natural 3D Scenes

In the previous section, we explained the correlation between the intensity of a pixel and its nearness. In this section, we expand this analysis to include the correlation between the intensity of a pixel and nearness of other pixels in the image. The set of all such correlations forms the cross-correlation between depth and intensity. The cross-correlation is an important statistical tool: as we explain later, if the cross-correlation between a particular image and its range image were known completely, then given the image, we could use simple linear regression techniques to infer 3D shape perfectly. While perfect estimation of the cross-correlation from a single image is impossible, we demonstrate that this correlational structure of a single scene follows several robust statistical trends. These trends allow us to approximate the full cross-correlation of a scene using only three parameters, and these parameters can be measured even from very sparse shape and intensity information. Approximating the cross-correlation this way allows us to achieve a novel form of statistically-driven depth inference that can be used in conjunction with other depth cues, such as stereo.

Given an image $i(x, y)$ with range image $z(x, y)$, the cross-correlation for that particular scene is given by

$$(i \star z)(\Delta x, \Delta y) = \iint i(x, y)\, z(x+\Delta x, y+\Delta y)\, dx\, dy \tag{1.2}$$

It is helpful to consider the cross correlation between intensity and depth within the Fourier domain. If we use $I(u, v)$ and $Z(u, v)$ denote the Fourier transform of $i(x, y)$ and $z(x, y)$ respectively, then the Fourier transform of $i \star z$ is $Z(u, v)I^*(u, v)$. ZI^* is known as the *cross-spectrum* of i and z. Note that ZI^* has both real and imaginary parts. Also note that in this section, no logarithm or other transformation was applied to the intensity or range data (measured in meters). This allows us to evaluate ZI^* in the context of the Lambertian model assumptions, as we demonstrate later.

If the cross-spectrum is known for a given image, and is sufficiently bounded away from zero, then 3D shape could be estimated from a single image using linear regression: $Z = I(ZI^*/II^*)$. In this section, we demonstrate that given only three parameters, a close approximation to ZI^* can be constructed. Roughly speaking, those three parameters are the strength of the nearness/brightness correlation in the scene, the prevalence of flat shaded surfaces in the scene, and the dominant direction of illumination in the scene. This model can be used to improve depth inference in a variety of situations.

Figure 1.3a shows a log-log polar plot of $|real[ZI^*(r, \theta)]|$ from one image in our database. The general shape of this cross-spectrum appears to closely follow a power law. Specifically, we found that ZI^* can be reasonably modeled

a) |real[ZI^*]|

b) Example $B_K(\theta)$ vs degrees counter-clockwise from horizontal

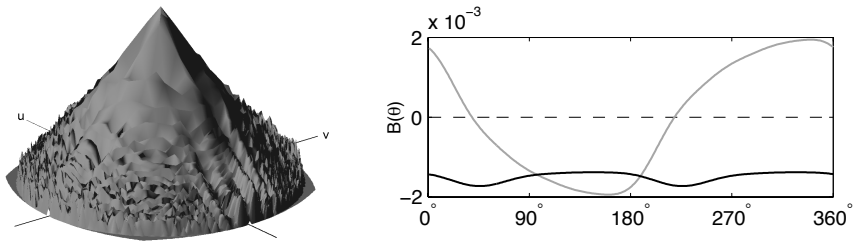


FIGURE 1.3: **a)** The log-log polar plot of $|real[ZI^*(r,\theta)]|$ for a scene from our database. **b)** $B(\theta)$ for the same scene. $real[B_K(\theta)]$ is drawn in black and $imag[B_K(\theta)]$ in grey. This plot is typical of most scenes in our database. As predicted by equation 1.5, imag$[B_K(\theta)]$ reaches its minima at the illumination direction (in this case, to the extreme left, almost 180°). Also typical is that real$[B_K(\theta)]$ is uniformly negative, most likely caused by cast shadows in object concavities [32].

by $B(\theta)/r^\alpha$, where r is spatial frequency in polar coordinates, and $B(\theta)$ is function that depends only on polar angle θ, with one curve for the real part and one for the imaginary part. We test this claim by dividing the Fourier plane into four 45° octants (vertical, forward diagonal, horizontal, and backward diagonal), and measuring the drop-off rate in each octant separately. For each octant, we average over the octant's included orientations and fit the result to a power-law. The resulting values of α (averaged over all 28 images) are listed in the table below:

orientation	II^*	real$[ZI^*]$	imag$[ZI^*]$	ZZ^*
horizontal	2.47 ±0.10	3.61 ±0.18	3.84 ±0.19	2.84 ±0.11
forward diagonal	2.61 ±0.11	3.67 ±0.17	3.95 ±0.17	2.92 ±0.11
vertical	2.76 ±0.11	3.62 ±0.15	3.61 ±0.24	2.89 ±0.11
backward diagonal	2.56 ±0.09	3.69 ±0.17	3.84 ±0.23	2.86 ±0.10
mean	2.60 ±0.10	3.65 ±0.14	3.87 ±0.16	2.88 ±0.10

For each octant, the correlation coefficient between the power-law fit and the actual spectrum ranged from 0.91 to 0.99, demonstrating that each octant is well-fit by a power-law (Note that averaging over orientation smoothes out some fine structures in each spectrum). Furthermore, α varies little across orientations, showing that our model fits ZI^* closely.

Note from the table that the image power spectra $I(u,v)I^*(u,v)$ also obey a power-law. The observation that the power spectrum of natural images obeys a power-law is one of the most robust and important statistic trends of natural images [37], and it stems from the scale invariance of natural images. Specifically, an image that has been scaled up, such as $i(\sigma x, \sigma y)$, has similar statistical properties as an unscaled image. This statistical property predicts

that $II^*(r,\theta) \approx 1/r^2$. The power-law structure of the power spectrum II^* has proven highly useful in image processing and computer vision, and has led to advances in image compression, image denoising, and several other applications. Similarly, the discovery that ZI^* also obeys a power spectrum may prove highly useful for the inference of 3D shape.

As mentioned earlier, knowing the full cross-covariance structure of an image/range image pair would allow us to reconstruct the range image using linear regression via the equation $Z = I(ZI^*/II^*)$. Thus, we are especially interested in estimating the regression kernel $K = ZI^*/II^*$. IK is a perfect reconstruction of the original range image (as long as $II^*(u,v) \neq 0$). The findings shown in the above table predict that K also obeys a power-law. Subtracting α_{II^*} from $\alpha_{\text{real}[ZI^*]}$ and $\alpha_{\text{imag}[ZI^*]}$, we find that $real[K]$ drops off at $1/r^{1.1}$ and $imag[K]$ drops off at $1/r^{1.2}$. Thus, we have that $K(r,\theta) \approx B_K(\theta)/r$.

Now that we know that K can be fit (roughly) by a $1/r$ power-law, we can offer some insight into why K tends to approximate this general form. Note that the $1/r$ drop-off of K cannot be predicted by scale invariance. If images and range images were jointly scale invariant, then II^* and ZI^* would both obey $1/r^2$ power laws, so that K would have roughly uniform magnitude. Thus, even though natural *images* appear to be statistically scale invariant, the finding that $K \approx B_K(\theta)/r$ disproves scale invariance for natural *scenes* (meaning images and range images taken together). In other words, while natural images retain similar statistical properties when scaled, and natural range images very nearly have this same property, the statistics of images and range images when taken together will not have this property. When images and range images are both scaled together, their joint statistics will vary according to a multiplicative constant.

The $1/r$ drop-off in the *imaginary* part of K can be explained by the linear Lambertian model of shading, with oblique lighting conditions. Recall that Lambertian shading predicts that pixel intensity is given by

$$i(x,y) \propto \vec{n}(x,y) \cdot \vec{L} \tag{1.3}$$

where $\vec{n}(x,y)$ is the unit surface normal at point (x,y) and $\vec{L}$ is the unit lighting direction. The linear Lambertian model is obtained by taking only the linear terms of the Taylor series of the Lambertian reflectance equation. Under this model, if constant albedo and illumination conditions are assumed, and lighting is from above, then $i(x,y) = a\,\partial z/\partial y$, where a is some constant. In the Fourier domain, $I(u,v) = a2\pi jvZ(u,v)$, where $j = \sqrt{-1}$. Thus, we have that

$$ZI^*(r,\theta) = -\frac{j}{a2\pi\, r\sin(\theta)} II^*(r,\theta) \tag{1.4}$$

$$K(r,\theta) = -j\,\frac{1}{r}\,\frac{1}{a2\pi\sin(\theta)} \tag{1.5}$$

Thus, Lambertian shading predicts that imag$[ZI^*]$ should obey a power-law, with $\alpha_{\text{imag}[ZI^*]}$ being one more than $\alpha_{\text{imag}[II^*]}$, which is consistent with the findings in the table above.

Equation 1.4 predicts that only the imaginary part of ZI^* should obey a power-law, and the real part of ZI^* should be zero. Yet, in our database, the real part of ZI^* was typically stronger than the imaginary part. The real part of ZI^* is the Fourier transform of the even-symmetric part of the cross-correlation function, and it includes the direct correlation cov$[i, z]$, corresponding to the da Vinci correlation between intensity pixel and range pixel discussed earlier. The form of real$[ZI^*]$ is related to the rate at which the da Vinci correlation drops off over space. One explanation for the $1/r^3$ drop-off rate of real$[ZI^*]$ is the observation that deeper crevices and concavities should be more shadowed and therefore darker than shallow concavities. Conversely, for two surface concavities with equal depths, the one with the narrower aperture should be darkest. If images and range images were jointly scale invariant, then the correlation between an image and a range image that were both convolved by an aperture filter f would be independent of the spatial scale of f:

$$\text{cor}[f(\sigma x, \sigma y) * i, f(\sigma x, \sigma y) * z] = \text{const} \tag{1.6}$$

However, this is not the case. Real scenes violate joint scale invariance because crevices of smaller aperture yield higher da Vinci correlations. When II^*, ZI^* and ZZ^* all obey the power laws shown in the table above, it can be shown that

$$\text{cor}[f(\sigma x, \sigma y) * i, f(\sigma x, \sigma y) * z] = \text{const} * \sigma^{\frac{\alpha_{II}+\alpha_{ZZ}}{2\alpha_{ZI}}} \approx \text{const} * \sigma^{0.75} \tag{1.7}$$

As σ increases, the filter aperture decreases, and the da Vinci correlation increases. Thus, the $1/r$ drop-off rate of K can be explained by the relationship between aperture size and the strength of the da Vinci correlation.

Figure 1.4 shows examples of B_K in urban and rural scenes. These plots illustrate that the real part of B_K is strongest (most negative) for rural scenes with abundant concavities and shadows. These figures also illustrate how the imaginary part of K follows Equation 1.5, and imag$[B_K(\theta)]$ closely follows a sinusoid with phase determined by the dominant illumination angle. Thus, B_K (and therefore also K and ZI^*) can be well approximated using only three parameters: the strength of the da Vinci correlation (which is related to the extent of complex 3D surfaces and shadowing present in the scene), the angle of the dominant lighting direction, and the strength of the Lambertian relationship in the scene (i.e. the coefficient $1/a$ in equation 1.5, which is related to the prominence of smooth Lambertian surfaces in the scene). In the following section, we show how we can use this approximation to improve depth inference.

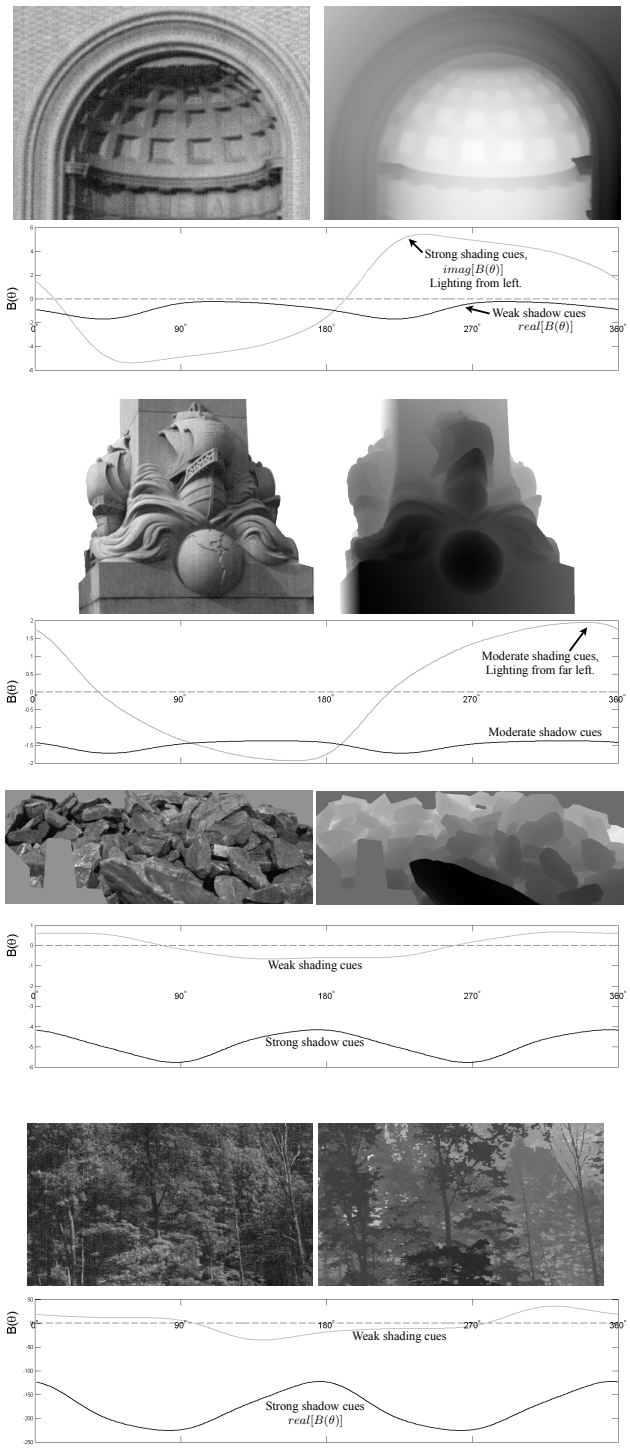


FIGURE 1.4: Natural and urban scenes and their $B_K(\theta)$. Images with surface concavities and cast shadows have significantly negative $real[B(\theta)]$ (black line), and images with prominent flat shaded surfaces have strong $imag[B(\theta)]$ (grey line).

1.4 Implications Towards Depth Inference

Armed with a better understanding of the statistics of real scenes, we are better prepared to develop successful depth inference algorithms. One example is range image super-resolution. Often, we may have a high-resolution color image of a scene, but only a low spatial resolution range image (range images record the 3D distance between the scene and the camera for each pixel). This often happens if our range image was acquired by applying a stereo depth inference algorithm. Stereo algorithms rely on smoothness constraints, either explicitly or implicitly, and so the high-frequency components of the resulting range image are not reliable [4, 38]. Laser range scanners are another common source of low-resolution range data. Laser range scanners typically acquire each pixel sequentially, taking up to several minutes for a high-resolution scan. These slow scan times can be impractical in real situations, so in many cases only sparse range data is available. In other situations, inexpensive scanners are used that can capture only sparse depth values.

It should be possible to improve our estimate of the high spatial frequencies of the range image by using monocular cues from the high-resolution intensity (or color) image. One recent study suggested an approach to this problem known as *shape recipes* [9, 45]. The basic principle of shape recipes is that a relationship between shape and appearance could be *learned* from the low resolution image pair, and then *extrapolated* and applied to the high resolution intensity image to infer the high spatial frequencies of the range image. One advantage of this approach is that hidden variables important to inference from monocular cues, such as illumination direction and material reflectance properties, might be implicitly learned from the low-resolution range and intensity images.

From our statistical study, we now know that fine details in $K = ZI^*/II^*$ do not generalize across scales, as was assumed by shape recipes. However, the coarse structure of K roughly follows a $1/r$ power-law. We can exploit this statistical trend directly. We can simply estimate $B_K(\theta)$ using the low-resolution range image, use the $1/r$ power-law to extrapolate $K \approx B_K(\theta)/r$ into the higher spatial frequencies, and then use this estimate of K to reconstruct the high frequency range data. Specifically, from the low-resolution range and intensity image, we compute low resolution spectra of ZI^* and II^*. From the highest frequency octave of the low-resolution images, we estimate $B_{II}(\theta)$ and $B_{ZI}(\theta)$. Any standard interpolation method will work to estimate these functions. We chose a $cos^3(\theta + \pi\phi/4)$ basis function based on steerable filters [49]. We now can estimate the high spatial frequencies of the range image, z. Define

$$K_{powerlaw}(r,\theta) = (B_{ZI}(\theta)/B_{II}(\theta))/r \tag{1.8}$$

$$Z_{powerlaw} = F_{low}(r)\, Z \;+\; (1-F_{low}(r))\, IK_{powerlaw} \tag{1.9}$$

a) Original Intensity Image b) Low-Res. Range Data c) Power-law Technique

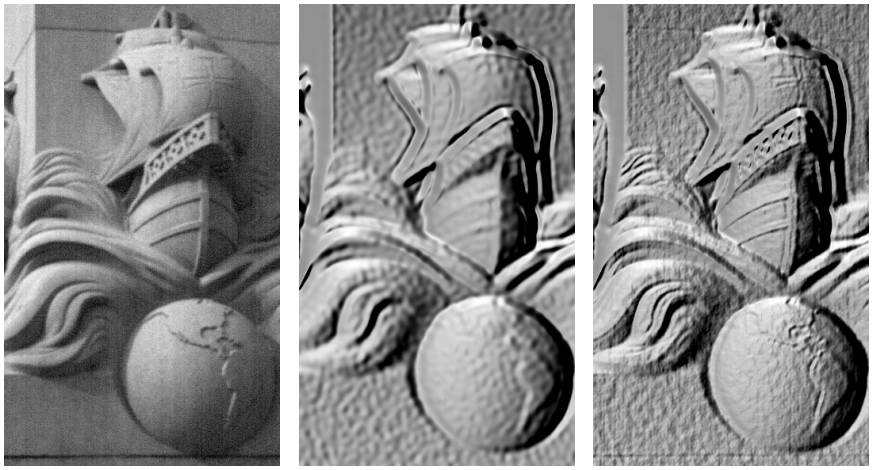


FIGURE 1.5: (a) A example intensity image from our database. (b) A computer-generated Lambertian rendering of the corresponding laser-acquired low-resolution range image. This figure shows the low-resolution range image which, for purposes of illustration, has been artificially rendered as an image. Note the over-smoothed edges and lack of fine spatial details that result from the down-sampling. (c) Power-law method of inferring high-resolution 3D shape from a low-resolution range image and a high-resolution color image. High spatial-frequency details of the 3D shape have been inferred from the intensity image (left). Notice that some high-resolution details, such as the cross in the sail, are not present at all in the low-resolution range image, but were inferred from the full-resolution intensity image.

where F_{low} is the low-pass filter that filters out the high spatial frequencies of z where depth information is either unreliable or missing.

Because our model is derived from scene statistics and avoids some of the mistaken assumptions in the original shape recipe model, our extension provides a two-fold improvement over Freeman and Torralba's original approach [45], while using far fewer parameters. Figure 1.5 shows an example of the output of the algorithm.

This power-law based approach can be viewed as a statistically informed generalization of a popular shape-from-shading algorithm known as *linear shape from shading* [30], which remains popular due to its high efficiency. Linear shape from shading attempts to reconstruct 3D shape from a single image using equation 1.5 alone, ignoring shadow cues and the da Vinci correlation. As mentioned previously, the da Vinci correlation is a product of cast shadows, complex 3D surfaces, diffuse lighting, and lighting interreflections. All four of these image formation phenomena are exceptionally cumbersome to invert in a deterministic image formation model, and subsequently they have been ignored by most previous depth inference algorithms. However,

taken together, these phenomena produce a very simple statistical relationship that can be exploited using highly efficient linear algorithms such as equation 1.9. It was not until the statistics of natural range and intensity images were studied empirically that the strength of these statistical cues was made clear.

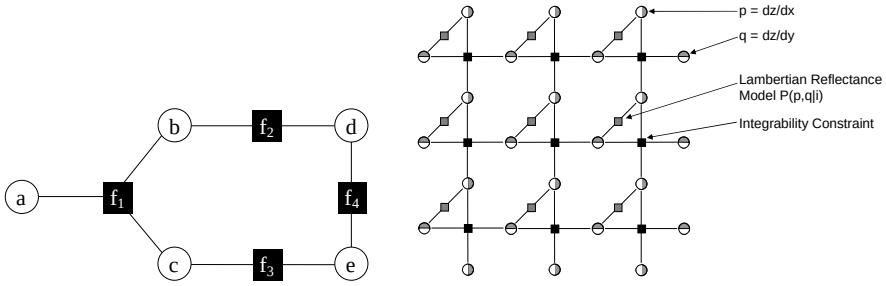
The power-law algorithm described here presents a new opportunity to test the usefulness of the da Vinci shadow cues, by comparing the power-law algorithm results to the linear shape from shading technique [30]. When our algorithm was made to use only shading cues (by setting the real part of $K_{powerlaw}(r, \theta)$ to zero), the effectiveness of the algorithm was reduced to 27% of it's original performance. When only shadow cues were used (by setting the imaginary part of $K_{powerlaw}(r, \theta)$ to zero), the algorithm retained 72% of it's original effectiveness [33]. Thus, in natural scenes, linear shadow cues proved to be significantly more powerful than linear shading cues. These results show that shadow cues are far more useful than was previously expected. This is an important empirical observation, as shape from shading has received vastly more attention in computer vision research than shape from shadow. This finding highlights the importance of shadow cues, and also the benefits of statistical studies of natural scenes.

As expected given the analysis of the da Vinci correlation above, the relative performance of shadow and shading cues depends strongly on the category of the images considered. Shadow cues were responsible for 96% of algorithm performance in foliage scenes, 76% in scenes of rocky terrain, and 35% in urban scenes.

1.5 Statistical Inference for Depth Inference

This approach described above shows the limits of what is possible using only second-order linear statistics. The study of these simple models is important, because it helps us to understand the statistical relationships that exist between shape and appearance. However, simple linear systems capture only a fraction of what is achievable using a complete statistical inference framework. The problem of inferring 3D shape from image cues is both highly complex and highly underconstrained: for any given 2D image, there are countless plausible 3D interpretations of that scene. Our goal is to find solutions that are especially likely. Powerful statistical methods will be necessary to achieve these goals. In this section, we discuss the use of modern statistical inference techniques for inferring 3D shape from images.

In recent years, there has been a great deal of progress made in computer vision using graphical models of large joint probability distributions [44, 57, 7, 10, 40, 42]. Graphical models offer a powerful framework to incor-



(a) Factor Graph (b) Graph for Shape from shading

FIGURE 1.6: **a)** An example factor graph. This graph represents the factorization of a joint probability distribution over five random variables: $P(a, b, c, d, e) \propto f_1(a, b, c) f_2(b, d) f_3(c, e) f_4(d, e)$. **b)** A factor graph to solve the classical Lambertian shape-from-shading problem using linear constraint nodes. The representation of 3D shape is twice overcomplete, including p and q slope values at each pixel. The linear constraint nodes are shown as black squares, and enforce the consistency (integrability) of the solution. The grey squares represent factor nodes encoding the reflectance function.

porate rich statistical cues from natural scenes and can be applied directly to the problem of depth inference. Bayesian inference of shape (depth) Z from images I involves estimating properties of the posterior distribution $P(Z|I)$. The dimensionality of the posterior distribution $P(Z|I)$, however, is far too great to model directly. An important observation relevant to vision is that the interdependency of variables tend to be relatively local. This allows the factorization of the joint distribution into a product of "potential functions," each of lower dimensionality than the original distribution (as shown in Figure 1.6). In other words,

$$P(I, Z) \propto \prod_a \phi_a(\vec{x}_a) \tag{1.10}$$

where $\vec{x}_a$ is some subset of variables in I and Z. Such a factorization defines an example of a graphical model known as a "factor graph": a bipartite graph with a set of variable nodes (one for each random variable in the multivariate distribution) and a set of factor nodes (one for each potential function). Each factor node is connected to each variable referenced by its corresponding potential function (see figure 1.6 for an example, or reference [11] for a review of factor graphs). Factor graphs that satisfy certain constraints can be expressed as Bayes networks, or for other constraints, as Markov Random Fields (MRF). Thus, these approaches are intimately connected and are equivalent in terms of neural plausibility.

Exact inference on factor graphs is possible only for a small subclass of problems. In most cases approximate methods must be used. There are a

variety of existing approaches to approximating the mode of the posterior distribution (MAP, or maximum *a posteriori*) or the its mean (MMSE, or minimum mean-squared error), such as Markov chain Monte Carlo (MCMC) sampling, graph cuts, and belief propagation. In this section, we explore the use of the belief propagation algorithm. Belief propagation is advantageous in that it imposes fewer restrictions on the potential functions than graph cuts [20] and is faster than MCMC. Belief propagation is also interesting in that it is highly neurally plausible [25, 35], and has been advanced as a possible model for statistical inference in the brain [29].

Belief propagation has been applied successfully to a wide variety of computer vision problems [10, 40, 42, 31], and has shown impressive empirical results on a number of other problems [21, 12]. Initially, the reasons behind the success of belief propagation were only understood for those cases where the underlying graphical model did not contain loops. The many empirical successes on graphical models that did contain loops were largely unexplained. However, recent discoveries have provided a solid theoretical justification for "loopy" belief propagation by showing that when belief propagation converges, it computes a minima of a measure used in statistical physics known as the Bethe free energy [54]. The Bethe free energy is based on a principled approximation of the KL-divergence between a graphical model and a set of marginals, and has been instrumental in studying the behaviors of large systems of interacting particles, such as spin glasses. The connection to Bethe free energy had the additional benefit that it inspired the development of algorithms that minimize the Bethe free energy directly, resulting in variants of belief propagation that guarantee convergence [56, 15], improve performance [54, 55], or in some cases, guarantee that belief propagation computes the *globally* optimal MAP point of a distribution [51].

Belief propagation estimates the marginals $b_i(x_i) = \sum_{X \setminus x_i} P(\vec{X})$ by iteratively computing *messages* along each edge of the graph according to the equations:

$$m_{i \to f}^{t+1}(x_i) = \prod_{g \in \mathcal{N}(i) \setminus f} m_{g \to i}^{t}(x_i) \tag{1.11}$$

$$m_{f \to i}^{t+1}(x_i) = \sum_{\vec{x}_{\mathcal{N}(f) \setminus i}} \left(\phi_f \left(\vec{x}_{\mathcal{N}(f)} \right) \prod_{j \in \mathcal{N}(f) \setminus i} m_{j \to f}^{t}(x_j) \right) \tag{1.12}$$

$$b_i(x_i) \propto \prod_{g \in \mathcal{N}(i)} m_{g \to i}^{t}(x_i) \tag{1.13}$$

where f and g are factor nodes, i and j are variable nodes, and $\mathcal{N}(i)$ is the set of neighbors of node i [14]. Here, $b_i(x_i)$ is the estimated marginal of variable i. Note that the expected value of $\vec{X}$, or equivalently, the minimum mean-squared error (MMSE) point estimate, can be computed by finding the mean of each marginal. If the most likely value of $\vec{X}$ is desired, also known

as the maximum *a posteriori* (MAP) point estimate, then the integrals of equation 1.12 are replaced by suprema. This is known as max-product belief propagation.

For many computer vision problems, belief propagation is prohibitively slow. The computation of Equation 1.12 has a complexity of $\mathcal{O}(M^N)$, where M is the number of possible labels for each variable, and N is the number of neighbors of factor node f. In many computer vision problems, variables are continuous or have many labels. In these cases, applications of belief propagation have nearly always been restricted to pairwise connected Markov Random Fields, where each potential function depends on only two variable nodes (i.e. $N = 2$) [10, 40]. However, pairwise connected models are often insufficient to capture the full complexity of the joint distribution, and thus would severely limit the expressive power of factor graphs. Developing efficient methods for computing non-pairwise belief propagation messages over continuous random variables is therefore crucial for solving the complex problems with rich, higher-order statistical distributions encountered in computer vision.

In the case that the potential function ϕ can be expressed in terms of a weighted sum of its inputs, we have developed a set of techniques to speed up the computation of messages considerably. For example, suppose the random variables a, b, c, and d are all variable nodes in our factor graph, and we want to constrain them such that $a + b = c + d$. We would add a factor node f connected to all four variables with potential function

$$\phi_f(a, b, c, d) = \delta(a + b - c - d) \tag{1.14}$$

To compute $m^{t+1}_{f\to A}$ we use equation 1.12:

$$m^{t+1}_{f\to A}(a) = \sum_{b,c,d} \delta(a + b - c - d)\; m^t_{B\to f}(b)\; m^t_{C\to f}(c)\; m^t_{D\to f}(d) \tag{1.15}$$

$$= \sum_{b,c} m^t_{B\to f}(b)\; m^t_{C\to f}(c)\; m^t_{D\to f}(a + b - c) \tag{1.16}$$

$$= \sum_{x,y} m^t_{B\to f}(x - a)\; m^t_{C\to f}(x - y)\; m^t_{D\to f}(y) \tag{1.17}$$

$$= \sum_{x} m^t_{B\to f}(x - a) \left(\sum_{y} m^t_{C\to f}(x - y)\; m^t_{D\to f}(y) \right) \tag{1.18}$$

where $x = a + b$ and $y = a + b - c$. Notice that in equation 1.18, the second summand (in parenthesis) does not depend on a. This summand can be computed in advance by summing over y for each value of x. Thus, computing $m^{t+1}_{f\to A}(a)$ using equation 1.18 is $\mathcal{O}(M^2)$, which is far superior to a straightforward computation of equation 1.15, which is $\mathcal{O}(M^4)$. In [34], we show how this same approach can be used to compute messages in time

$\mathcal{O}(M^2)$ for all potential functions of the form

$$\phi(\vec{x}) = g\left(\sum_i g_i(x_i)\right) \tag{1.19}$$

This reduces a problem from exponential time to linear time with respect to the number of variables connected to a factor node. Potentials of this form are very common in graphical models, in part because they offer advantages in training graphical models from real data [13, 60, 16, 36].

This approach reduces a problem from exponential time to linear time with respect to the number of variables connected to a factor node. With this efficient algorithm, we were able to apply belief propagation towards the classical computer vision problem of shape from shading, using the factor graph shown in Figure 1.6 (see [31] for details). Previously, the general problem of shape from shading was solved using gradient descent based techniques. In complex, highly nonlinear problems like shape from shading, these approaches often become stuck inside local, suboptimal minima. Belief propagation helps to avoid difficulties with local minima in part because it operates over whole probability distributions. While gradient descent approaches maintain only a single 3D shape at a time, iteratively refining that shape over time, belief propagation seeks to optimize the single-variate marginals $b_i(x_i)$ for each variable in the factor graph.

Solving shape from shading using belief propagation performs significantly better than previous state of the art techniques (see figure 1.7). Note without the efficient techniques described here, belief propagation would be intractable for this problem, requiring over 100,000 times longer to compute each iteration. In addition to improved performance, solving shape from shading using belief propagation allows us to relax many of the restrictions typically assumed by shape from shading algorithms in order to make the problem tractable. The classical definition of the shape from shading problem specifies that lighting must originate from a single point source, that surfaces should be entirely matte, or Lambertian in reflectance, and that no markings or colorations can be present on any surface. The flexibility of the belief propagation approach allows us to start relaxing these constraints, making shape from shading viable in more realistic scenarios.

1.6 Concluding Remarks and Future Directions

The findings described here underline the importance of studying the statistics of natural scenes; specifically, to study not only the statistics of images alone, but images together with their underlying scene properties. Just as the statistics of natural images has proven invaluable for understanding efficient

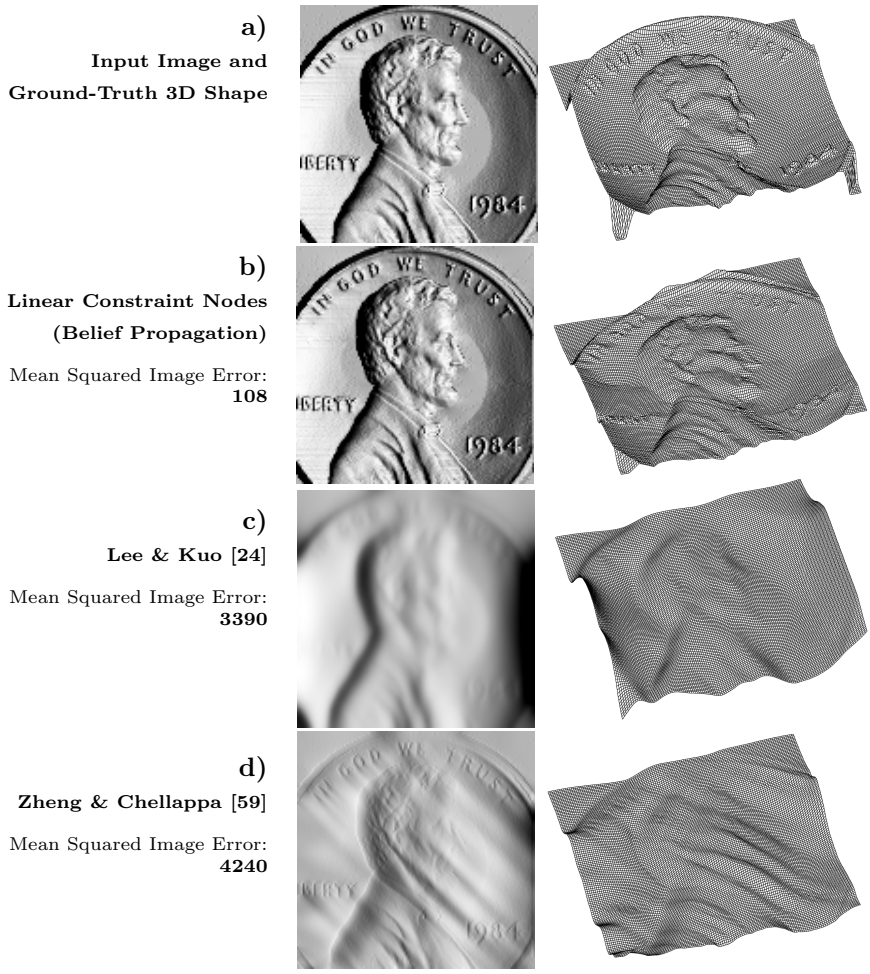


FIGURE 1.7: Comparison between our results of inferring shape from shading using loopy belief propagation (row b) with previous approaches (rows c and d). Each row contains a 3D wire mesh plot of the surface (right) and a rendering (left) of that surface under a light source at location (1,0,1). **(a)** The original surface. The rendering in this column serves as the input to the SFS algorithms in the next three columns. **(b)** The surface recovered using our linear constraint node approach. **(c)** The surface recovered using the method described by Lee and Kuo [24]. This algorithm performed best of the six SFS algorithms reviewed in the recent survey paper [58]. **(d)** The surface recovered using the method described by Zheng and Chellappa [59]. Our approach (row b) offers a significant improvement over previous leading methods. It is especially important that re-rendering that recovered surface very closely resembles the original input image. This means that the Lambertian constraint at each pixel was satisfied, and that any error between the original and recovered surface is primarily the fault of the simplistic model of prior probability of natural 3D shapes used here.

image coding, transmission, and representation, the joint statistics of natural scenes stands to greatly advance our understanding of perceptual inference. The discovery of the da Vinci correlation described here illustrates this point. This absolute correlation between nearness and brightness observed in natural 3D scenes is among the simplest statistical relationships possible. However, it stems from the most complex phenomena of image formation; phenomena that have historically been ignored by computer vision approaches to depth inference for the sake of mathematical tractability. It is difficult to anticipate this statistical trend by only studying the physics of light and image formation. Also, because the da Vinci correlation depends on intrinsic scene properties such as the roughness or complexity of a 3D scene, physical models of image formation are unable to estimate the strength of this cue, or its prevalence in real scenes. By taking explicit measurements using laser range finders, we have demonstrated that this cue is very strong in natural scenes, even under oblique, non-diffuse lighting conditions. Further, we have shown that for linear depth inference algorithms, shadow cues such as the da Vinci correlation are 2.7 times as informative as shading cues in a diverse collection of natural scenes. This result is especially significant, because depth cues from shading has received far more attention than shadow cues. We believe that continued investigation into natural scene statistics will continue to uncover important new insights into visual perception that are unavailable to approaches based on physical models alone.

Another conclusion we wish to draw is the benefit of statistical methods of inference for visual perception. The problem of shape from shading described above was first studied in the 1920s in order to reconstruct the 3D shapes of lunar terrains [17]. Since that time, approaches to shape from shading were primarily deterministic, and typically involved iteratively refining a single shape hypothesis until convergence was reached. By developing and applying efficient statistical inference techniques that consider *distributions* over 3D shapes, we were able to advance the state of shape from shading considerably. The efficient belief propagation techniques we have developed have similar applications in a variety of perceptual inference tasks. These and other statistical inference techniques promise to significantly advance the state of the art in computer vision and to improve our understanding of perceptual inference in general.

In addition to improved performance, the approach to shape from shading described above offers a new degree of flexibility that should allow shading to be exploited in more general and realistic scenarios. Previous approaches to shape from shading typically relied heavily on the exact nature of the Lambertian reflectance equations, and so could only be applied to surfaces with specific (i.e. matte) reflectance qualities with no surface markings. Also, specific lighting conditions were assumed. The approach described above applies directly to a statistical model of the relationship between shape and shading, and so it does not depend on the exact nature of the Lambertian equation or specific lighting arrangements. Also, the efficient higher-order belief prop-

agation techniques described here make it possible to exploit stronger, non-pairwise models of the prior probability of 3D shapes. Because the problem of depth inference is so highly underconstrained, and natural images admit large numbers of plausible 3D interpretations, it is crucial to utilize an accurate model of the prior probability of 3D surface. Knowing what 3D shapes commonly occur in nature, and what shapes are *a priori* unlikely or odd is a very important constraint for depth inference. Finally, the factor graph representation of the shape from shading problem (see figure 1.6) can be generalized naturally to exploit other depth cues, such as occlusion contours, texture, perspective, or the da Vinci correlation and shadow cues. The state of the art approaches to the inference of depth from binocular stereo pairs typically employ belief propagation over a markov random field. These approaches can be combined with our shape from shading framework in a fairly straightforward way, allowing both shading and stereo cues to be simultaneously utilized in statistically optimal way. Statistical approaches to depth inference make it possible to work towards a more unified and robust depth inference framework, which is likely to become a major area of future vision research.

References

[1] M. Ashley. Concerning the significance of light in visual estimates of depth. *Psychological Review*, 5(6):595–615, 1898.

[2] H. Carr. *An Introduction to Space Perception*. Longmans, Green and Co, New York, 1935.

[3] J. Coules. Effect of photometric brightness on judgments of distance. *Journal of Experimental Psychology*, 50:19–25, 1955.

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

[5] J. E. Cutting and P. M. Vishton. Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth. In William Epstein and Sheena J Rogers, editors, *Perception of space and motion*, Handbook of perception and cognition, pages 69–117. Academic Press, San Diego, CA, USA, 1995.

[6] M. Farne. Brightness as an indicator to distance: Relative brightness per se or contrast with the background? *Perception*, 6:287–293, 1977.

[7] L. Fei-Fei and P. Perona. A bayesian hierarchical model for learning natural scene categories. In *CVPR*, 2005.

[8] D. J. Field. What is the goal of sensory coding? *Neural Computing*, 6:559–601, 1994.

[9] W. T. Freeman and A. Torralba. Shape recipes: Scene representations that refer to the image. In *Advances in Neural Information Processing Systems 15 (NIPS)*, 2003.

[10] William T. Freeman, Egon Pasztor, and Owen T. Carmichael. Learning low-level vision. *Int. J. Comp. Vis.*, 40(1):25–47, 2000.

[11] B.J. Frey. *Graphical models for machine learning and digital communication*. MIT Press, 1998.

[12] Brendan J. Frey and Delbert Dueck. Clustering by passing messages between data points. *Science*, January 2007.

[13] Jerome H. Friedman, Werner Stuetzle, and Anne Schroeder. Projection pursuit density estimation. *Journal of the American Statistical Association*, 79:599–608, 1984.

[14] Tom Heskes. On the uniqueness of loopy belief propagation fixed points. *Neural Comp.*, 16(11):2379–2413, 2004.

[15] Tom Heskes, Kees Albers, and Bert Kappen. Approximate inference and constrained optimization. In *UAI*, pages 313–320, 2003.

[16] Geoffrey Hinton. Products of experts. In *International Conference on Artificial Neural Networks*, volume 1, pages 1–6, 1999.

[17] Berthold K. P. Horn. Obtaining shape from shading information. pages 123–171, 1989.

[18] C. Q. Howe and D. Purves. Range image statistics can explain the anomalous perception of length. *Proc. Nat. Acad. Sci.*, 99:13184–13188, 2002.

[19] Jinggang Huang, Ann B. Lee, and David Mumford. Statistics of range images. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1324–1331, 2000.

[20] Vladimir Kolmogorov and Ramin Zabih. Computing visual correspondence with occlusions using graph cuts. In *International Conference on Computer Vision (ICCV)*, volume 2, pages 508–515. IEEE, 2001.

[21] Frank R. Kschischang and Brendan J. Frey. Iterative decoding of compound codes by probability propagation in graphical models. *IEEE Journal of Selected Areas in Communications*, 16(2):219–230, 1998.

[22] M. S. Langer and S. W. Zucker. Shape from Shading on a Cloudy Day. *Journal of the Optical Society of America - Part A: Optics, Image Science, and Vision*, 11(2):467–478, February 1994.

[23] M.S. Langer and H.H. Blthoff. Perception of shape from shading on a cloudy day. Technical Report 73, Tbingen, Germany, oct 1999.

[24] K.M. Lee and C.C.J. Kuo. Shape from shading with a linear triangular element surface model. *IEEE Trans. Pattern Anal. Mach. Intell.*, 15(8):815–822, 1993.

[25] Tai Sing Lee and David Mumford. Hierarchical bayesian inference in the visual cortex. *J. Opt. Soc. Amer. A*, 20:1434–1448, 2003.

[26] E. MacCurdy, editor. *The Notebooks of Leonardo da Vinci, Volume II*. Reynal & Hitchcock, New York, 1938.

[27] D. G. Myers. *Psychology*. Worth Publishers, New York, 1995.

[28] S.K. Nayar and S.G. Narasimhan. Vision in bad weather. In *ICCV*, volume 2, pages 820–827, 1999.

[29] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufman, San Francisco, CA, 1988.

[30] A. P. Pentland. Linear Shape From Shading. *International Journal of Computer Vision*, 4(2):153–162, March 1990.

[31] Brian Potetz. Efficient belief propagation for vision using linear constraint nodes. In *CVPR 2007: Proceedings of the 2007 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE Computer Society, Minneapolis, MN, USA, 2007.

[32] Brian Potetz and Tai Sing Lee. Statistical correlations between two-dimensional images and three-dimensional structures in natural scenes. *J. Opt. Soc. Amer. A*, 20(7):1292–1303, 2003.

[33] Brian Potetz and Tai Sing Lee. Scaling laws in natural scenes and the inference of 3d shape. In Y. Weiss, B. Schölkopf, and J. Platt, editors, *Advances in Neural Information Processing Systems 18*, pages 1089–1096. MIT Press, Cambridge, MA, 2006.

[34] Brian Potetz and Tai Sing Lee. Efficient belief propagation for higher order cliques using linear constraint nodes. *Computer Vision and Image Understanding*, 112(1):39–54, Oct 2008.

[35] R.P.N. Rao. Bayesian computation in recurrent neural circuits. *Neural Computation*, 16(1), 2004.

[36] Stefan Roth and Michael J. Black. Fields of experts: A framework for learning image priors. In *CVPR*, pages 860–867, 2005.

[37] D. L. Ruderman and W. Bialek. Statistics of natural images: scaling in the woods. *Physical Review Letters*, 73:814–817, 1994.

[38] Daniel Scharstein and Richard Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int. J. Comput. Vision*, 47(1-3):7–42, 2002.

[39] T. M. Sheffield, D. Meyer, B. Payne J. Lees, E. L. Harvey, M. J. Zeitlin, , and G. Kahle. Geovolume visualization interpretation: A lexicon of basic techniques. *The Leading Edge*, 19:518–525, 2000.

[40] Jian Sun, Nan-Ning Zheng, and Heung-Yeung Shum. Stereo matching using belief propagation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(7):787–800, 2003.

[41] R.T. Surdick, E.T. Davis, R.A. King, and L.F. Hodges. The perception of distance in simulated visual displays: A comparison of the effectiveness and accuracy of multiple depth cues across viewing distances. *Presence: Teleoperators and Virtual Environments*, 6:513–531, 1997.

[42] Kam Lun Tang, Chi Keung Tang, and Tien Tsin Wong. Dense photometric stereo using tensorial belief propagation. In *CVPR*, pages 132–139, 2005.

[43] I. L. Taylor and F. C. Sumner. Actual brightness and distance of individual colors when their apparent distance is held constant. *The Journal of Psychology*, 19:79–85, 1945.

[44] A. Torralba, K.P. Murphy, W.T. Freeman, and M.A. Rubin. Context-based vision system for place and object recognition. In *ICCV*, 2003.

[45] Antonio Torralba and William T. Freeman. Properties and applications of shape recipes. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 383–390, 2003.

[46] Antonio Torralba and Aude Oliva. Depth estimation from image structure. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(9):1226–1238, 2002.

[47] Christopher W. Tyler. Diffuse illumination as a default assumption for shape from shading in the absence of shadows. *The Journal of imaging science and technology*, 42(4):319–325, 1998.

[48] J. H. van Hateren and A. van der Schaaf. Independent component filters of natural images compared with simple cells in primary visual cortex. *Proceedings of the Royal Society of London B*, 265:359–366, 1998.

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

[50] C. Wallschlaeger and C. Busic-Snyder. *Basic Visual Concepts and Principles for Artists, Architects, and Designers.* McGraw Hill, Boston, 1992.

[51] Yair Weiss and William T. Freeman. What makes a good model of natural images? In *CVPR 2007: Proceedings of the 2007 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE Computer Society, Minneapolis, MN, USA, 2007.

[52] O. J. Woodford, I. D. Reid, P. H. S. Torr, and A. W. Fitzgibbon. Fields of experts for image-based rendering. In *Proceedings of the 17th British Machine Vision Conference, Edinburgh*, volume 3, pages 1109–1108, 2006.

[53] M. Wright and T. Ledgeway. Interaction between Luminance Gratings and Disparity Gratings. *Spatial Vision*, 17(1–2):51–74, 2004.

[54] Jonathan S. Yedidia, William T. Freeman, and Yair Weiss. Generalized belief propagation. In *NIPS*, pages 689–695, 2000.

[55] Jonathan S. Yedidia, William T. Freeman, and Yair Weiss. Understanding belief propagation and its generalizations. In *Exploring artificial intelligence in the new millennium*, pages 239–269. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2003.

[56] Alan L. Yuille. CCCP algorithms to minimize the Bethe and Kikuchi free energies: Convergent alternatives to belief propagation. *Neural Computation*, 14(7):1691–1722, 2002.

[57] S. C. Zhu Z. W. Tu. Image segmentation by data-driven markov chain monte carlo. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24:657–673, 2002.

[58] Ruo Zhang, Ping-Sing Tsai, James Edwin Cryer, and Mubarak Shah. Shape from shading: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.*, 21(8):690–706, 1999.

[59] Qinfen Zheng and Rama Chellappa. Estimation of illuminant direction, albedo, and shape from shading. *IEEE Trans. Pattern Anal. Mach. Intell.*, 13(7):680–702, 1991.

[60] Song Chun Zhu, Ying Nian Wu, and David Mumford. Frame : Filters, random fields and maximum entropy — towards a unified theory for texture modeling. *Int'l Journal of Computer Vision*, 27(2):1–20, 1998.