Image statistics and surface perception

Edward H. Adelson
Massachusetts Institute of Technology, MA, USA 02139

ABSTRACT

Images have characteristic statistics that can be characterized in terms of the responses of wavelet or Gabor-like filters. There has been a great deal of interest in the fact that images have sparse (kurtotic) statistics in the wavelet domain, with implications for efficient image encoding in biological and artificial systems. If we set aside the issue of efficiency, we are still left with the problem of seeing. We have been studying the ways in which filter statistics can reveal useful information about surfaces, including albedo, shading, and gloss. We find that odd order statistics such as skewness are quite useful in extracting information about reflectance and gloss, and we also find evidence that humans make use of this information. It is straightforward to compute skewness with physiological mechanisms.

1. INTRODUCTION

It is a matter of everyday experience that we are able to visually judge surface properties such as color, lightness, roughness, and gloss, and can often estimate complex properties such as softness or slipperiness. Moreover we can usually tell what things are made of, e.g., wood, metal, or cloth. We are highly skilled at evaluating the appearance of materials, which is why there are large industries devoted to controlling that appearance. Major parts of our economy are devoted to the surface appearance of skin, hair, food, clothing, furniture, and so on. Presumably our finely honed perception evolved for practical purposes, such as choosing a mate, choosing food, or choosing where to place our next footstep.

Given the ubiquity of material judgments, and their importance to our lives, it is surprising that so little is known about them. Recently a few labs, including our own, have been making progress by combining psychophysical observations with ideas taken from machine vision, computer graphics, and computational neuroscience. This work links up with the problems studied in the field of "natural image statistics," in which one characterizes the distributions of various simple image properties such as filter outputs.

Researchers in human and machine vision often make the simplifying assumption that the world is made of Lambertian surfaces. But Lambertian shading looks quite artificial, and replicating the qualities of natural surfaces has attracted much attention in computer graphics (Glassner, 1995). Recently, the topic has been getting attention in computer vision, notably in the development of databases with the bidirectional reflectance distribution functions (BRDF's) of natural surfaces (Dana et al., 1999).

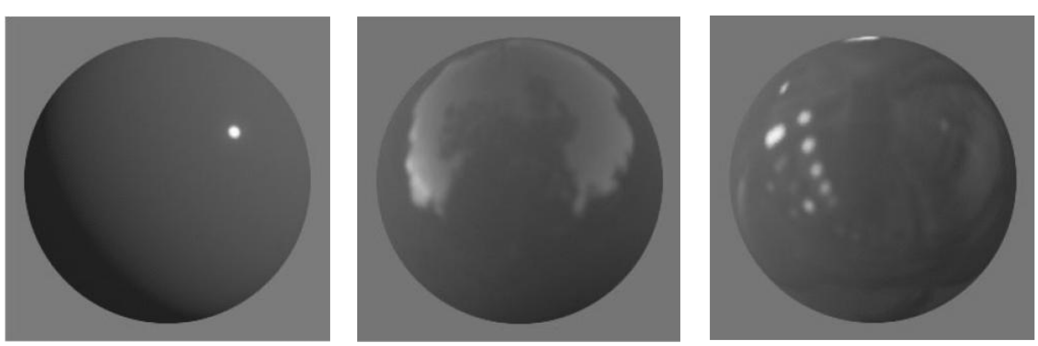

Figure 1: A glossy sphere under three different illuminations

Human Vision and Electronic Imaging XIII, edited by Bernice E. Rogowitz, Thrasyvoulos N. Pappas,
Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 6806, 680602, © 2008 SPIE-IS&T · 0277-786X/08/$18

We might use the term "rich surface perception" to refer to the perception of the complex surfaces of the real world, viewed in the complex illumination of everyday life. There is a richness in the information present in the image, a richness in the material qualities in the surface itself, and a richness in the perceptual qualities that are extracted by the visual system. Ultimately one would like to understand what is going on when we look at the skin on someone's forehead and notice that it is a bit pale, slightly moist, and showing signs of aging. This is too much to do all at once, but it is good to keep an ambitious endpoint in mind.

The appearance of an object depends on the following: its shape, its optical properties, the distribution of light falling on it (the illumination), and the point of view from which it is observed. (These are exactly the things that must be specified if an object is to be rendered in a computer graphics system). Point of view can be simply described, but the other factors can be quite complex, and their interaction to produce the image can be even more complex. Therefore it is wise to study some specific problems that bring in some richness, but not too much.

We have looked at the problem of the perception of glossy spheres in both human and machine vision (Dror et al, 2001; Fleming et al., 2003). We restricted ourselves to spheres with uniform reflectance (e.g., with a single color and gloss over the surface). By using spheres, we set aside the problem of shape: all spheres project to circles in the image, and so all information about the surface is contained in the distribution of light within that circle. We did not restrict ourselves to point source illumination. This is important, as our most interesting findings depend on the use of rich real-world illumination.

Figure 1 shows three images of the same sphere under different illuminations. On the left is point source illumination; in the middle is outdoor (natural) illumination; on the right is indoor (man-made) illumination. The two images with real world illumination were synthesized in Radiance (Ward, 1994), using panoramic illumination data gathered by Debevec (Debevec, 1998). The point source illumination, which is conceptually simple, leads to a rather weak sense of the surface reflectance. The real world illuminations, whether outdoor or indoor, are quite complex, and lead to compelling percepts of gloss. It is interesting that the richness of the illumination seems to aid the perceptual task rather than hindering it.

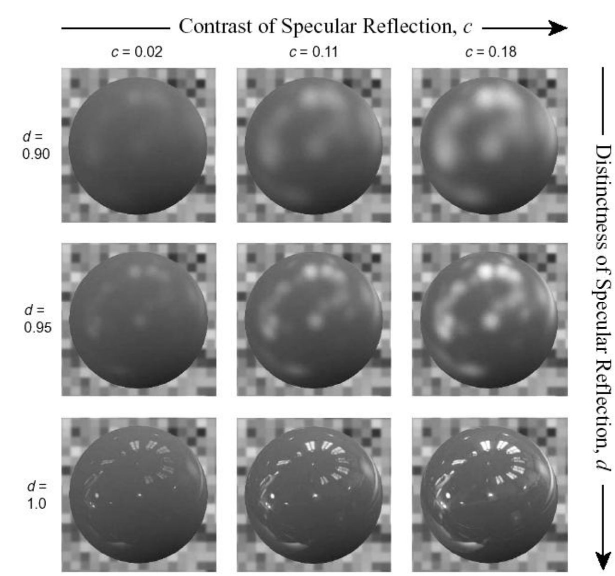

Figure 2: A set of spheres with varying gloss parameters.

When we speak of the ability to judge gloss, we are in effect speaking of gloss constancy: the ability to tell the gloss parameters of a surface regardless of its shape or the illumination conditions. In order to quantify this notion, we have utilized a technique of asymmetric gloss matching (asymmetric in the sense that two objects can look substantially different, yet be judged to be equally glossy). A given synthetic object is rendered with a variety of gloss parameters, and it is used to match another object.

The most popular parametric model for surface gloss in computer graphics is the Phong model, which has three parameters (in the achromatic case): the amount of diffuse reflectance (albedo), the amount of specular reflectance, and the width of the specular highlight. The width of the highly is implicitly attributed to the roughness of surface at a microscopic scale. We used a variant introduced by Ward (Ward, 1992), which corrects some of the physically unrealizable aspects of the Phong model. We placed virtual spheres in a variety of illuminations. Each illumination was, in effect, a pattern on a hollow sphere of infinite radius; this is known as an environment map or illumination map in computer graphics.

After rendering the spheres, we placed them against a background of random checks, which was intentionally out of focus to give the impression that the spheres were floating at a distance from the background. An array of such images is shown in Figure 2. The background checks included a range of color and luminance in order to provide consistent and articulated anchoring for the stimuli. The same background was used for all the stimuli.

Although the background did not correspond to the actual environment in which the rendering occurred, subjects were generally unaware of the inconsistency, and they found it easy to judge the reflectance properties of the sphere. We varied two parameters: the amount of specular reflectance, and the roughness of the surface (i.e., the blur of the specular highlights). We did this within the context of the perceptual space described by Pellacini et al. (2000). A test sphere and a match sphere were presented side by side. The subject adjusted the parameters of the match sphere to have the same apparent gloss as the test sphere. (All of the match sphere images were precomputed to allow real time interactive adjustment).

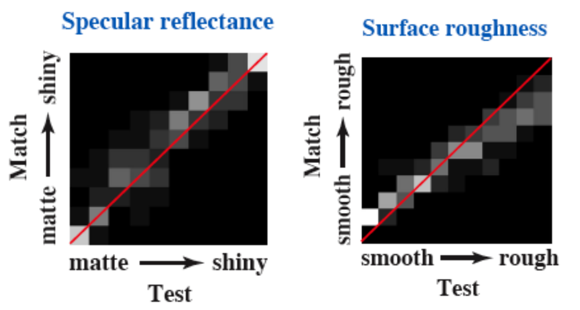

Figure 3: Subjective matches of reflectance parameters between spheres in different illuminations.

Figure 3 shows the results for one subject. The test sphere and the match sphere were rendered under different illumination maps, and the subject had to adjust the two spheres to have the same gloss. This subject was able to perform the task fairly well for both the amount of specular reflection (left) and the roughness (right).

There is an inherent ambiguity in this task, because a smooth sphere in a blurry environment looks the same as a rough sphere in a sharp environment (Ramamoorthi & Hanrahan, 2001). Under the right assumptions, this is a complete ambiguity: in both cases the environment map has effectively been convolved with a blur function, and it is impossible to tell whether the blur was already in the environment before it was reflected off the sphere surface. Not surprisingly, if subjects are given a sphere rendered in an environment that is blurry, they will incorrectly attribute some of the blur to the surface, and this leads to a systematic bias in their matches. However, it is interesting that most real-world illuminations have enough information to prevent this bias.

What aspects of the visual environment are important to supporting gloss constancy? Figure 4 shows a sphere rendered three environments. The first is an indoor environment. The second is a collection of point sources. The third is 1/f noise. It is interesting that the real-world environment gives a much better sense of the surface than the two synthetic environments. Most interesting is the fact that 1/f noise does poorly even though it has the same spectrum as natural scenes. (Note: since 1/f noise is normally defined on a plane, not a sphere, we synthesized environments using spherical harmonic spectra rather than Fourier spectra.).

To quantify the quality of gloss perception, we had subjects make matches between various environments and a single standard real-world environment, and measured the accuracy. Figure 6 shows the results, expressed as RMS error, for 8 real-world environments (lighter bars) and 5 synthetic environments (darker bars), for the judgments of the roughness parameter. The errors were much greater for the synthetic scenes than for the real-world scenes, with one exception: the extended rectangular source.

The error values in such experiments can have two sources, which may be illustrated by the binned scatter plots of Figure 5. The lighter squares have more counts in their bins. Figure 5(a) shows the pooled data over all subjects for the

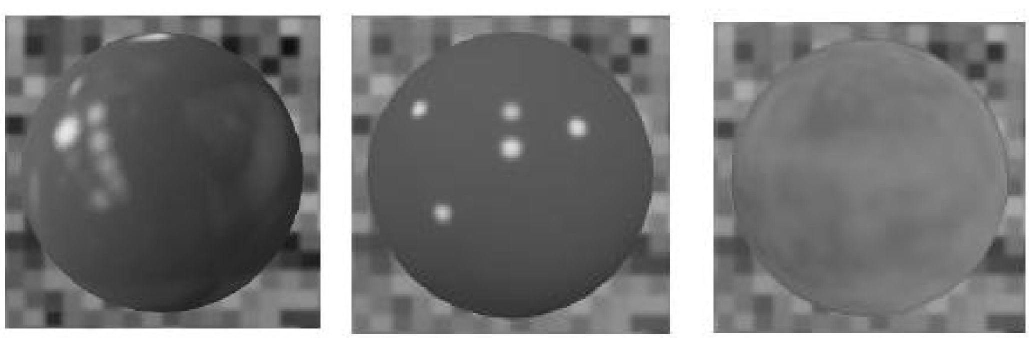

Figure 4 : A shiny sphere rendered in three environments: an indoor scene, a set of point sources, and a surround of 1/f noise.

estimates of reflectance Uffizi environment, while Figure 5(b) shows it for 1/f noise. The Uffizi case shows a systematic

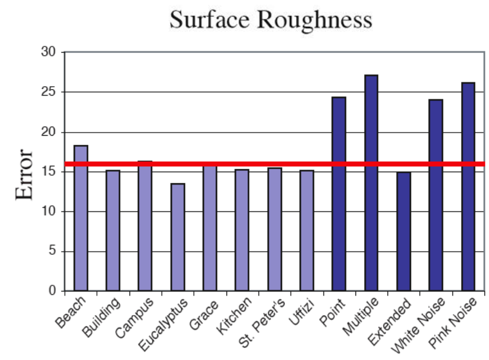

Figure 6: Errors in gloss matches for a variety of real-world and synthetic illuminations.

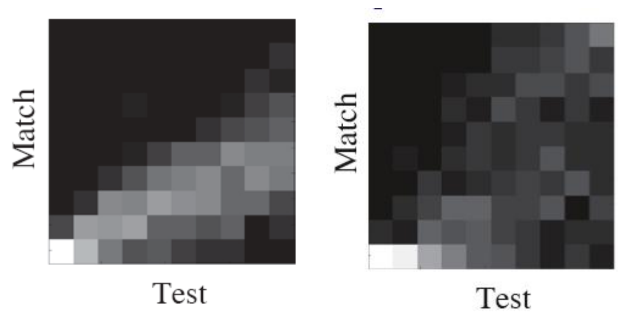

Figure 5: Binned scatterplots of the matches made with a real world environment and a synthetic one.

bias, but a reasonably consistent set of judgments. The 1/f noise case shows a disorganized pattern with great variability, indicating that the subjects had difficulty making the judgment. These two aspects of the errors are both informative. The systematic bias helps us determine what makes a reflected pattern look more or less glossy. The variability of the judgments can tell us whether subjects have a strong and consistent impression of the gloss parameters.

2. CUES FOR GLOSS

What cues are humans using in judging gloss? We can reject a number of hypotheses already. It is not just the spectrum of the reflected pattern, because 1/f noise has the same spectrum as natural images, and its spectrum is modified by the reflection in the same way as images are. We have also examined the matches for various illuminations, and have found that we can reject a variety of simple features such as the luminance of the brightest point, or the sharpness of the sharpest gradient. We can reject a variety of features (e.g., the steepest gradient, or the power spectrum) by considering the case of spheres with reversed contrast, as shown in Figure 7. This sphere does not look shiny at all; the dark spots seem to be marks or stains on the surface. Note that this image is physically realizable, since it is what would be seen if the shiny sphere were rendered in an inverted environment. However, the environment in which we live usually has a few bright spots (e.g., light sources), rather than a few black spots against an otherwise luminous background.

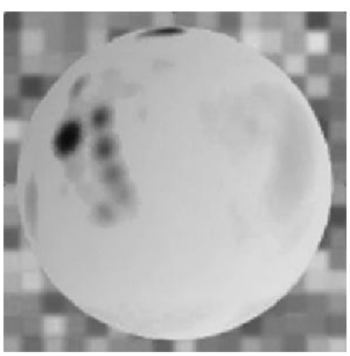

Figure 7: A glossy sphere shown in inverted contrast.

The critical point is that the appearance of glossy surfaces depends on the appearance of the surrounding world. Gloss perception is only possible because this world has certain visual characteristics. The characteristics are only statistical, but apparently the statistics are reliable enough to allow humans to estimate gloss under a variety of situations. Humans are implicitly aware of these stable aspects of the world, at least as they are expressed in the patterns seen reflected in glossy surfaces.

The statistical nature of the problem suggests that there may be useful analogies to the problem of texture perception. We can think of the three squares in Figure 8 as being three samples from a random process that makes patches of random noise with certain parameters. Each sample is completely different at the pixel level, but there is a visual quality that they all share. Likewise, we can think of the three sphere images below as three samples from a random process that makes images of mirrored spheres. Again, each image is completely different at the pixel level, and again, there is some visual quality that they all share. For both the textures and the spheres, these qualities are statistical. There is no template that can be used to match a texture in the same way that there can be for a letter "A".

When Julesz began studying texture perception, he used arrays of random checks rather like the ones in Figure 8. He proposed (Julesz, 1962) that humans responded to first and second order pixel statistics, and that two textures that were matched up to second order statistics would be matched in textural appearance. It turned out that this didn't work, and Julesz proposed an alternative approach based "textons," (Julesz, 1981) but these were never formalized. Nonetheless the notion of texture perception as the estimation of statistical parameters seems to be correct.

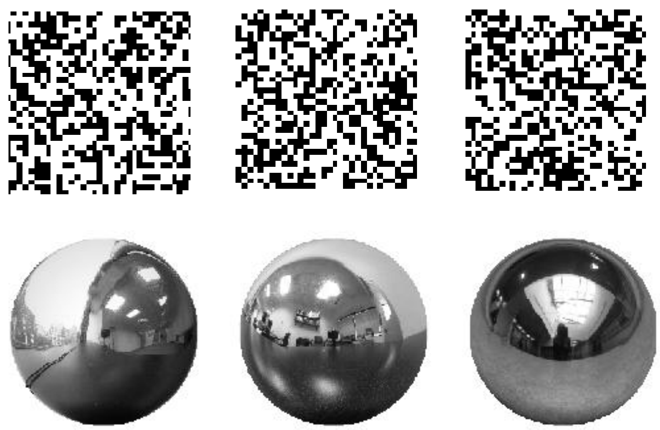

Figure 8: Top: three samples of a random texture. Bottom: three samples of chrome spheres.

There have been a number of important psychophysical models of texture perception based on banks of Gabor-like linear filters. Many of these models can be thought of as measuring the local energy in multiple bands and performing various non-linear interactions between the outputs (e.g., Malik & Perona 1990; Bergen & Landy 1991).

An important advance was described by Heeger and Bergen (1995). This paper was nominally about texture synthesis. It showed how to make a new sample of texture that resembles a given sample -- for instance, given some wood grain, make some more wood grain. Heeger and Bergen developed a texture metric (i.e., a function that describes the perceptual distance between two textures), and a method of modifying a given texture so that it is more similar to another texture. By iterating this procedure, they could start with a noise texture and force it to move to a desired position in texture space. Insofar as the metric was correct, the new texture should look similar to the original. When previous texture models were used in this way, they failed badly. Heeger and Bergen did much better by matching the full histograms of wavelet subbands (rather than just the variances), as well as the full histograms of pixel luminances. These two sets of constraints were far more successful than those in the traditional texture models. Since then, other papers on texture analysis and synthesis have continued the theme (Zhu et al., 1997; Portilla & Simoncelli (2000)).

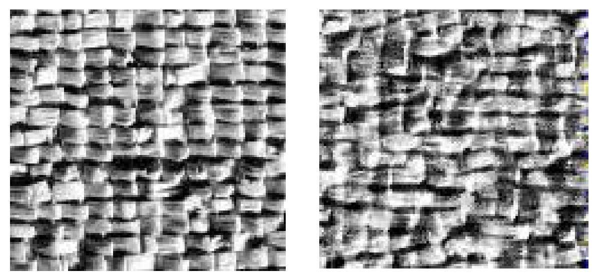

Figure 9: Left: A sample of raffia texture from the Brodatz database. Right: A synthetic sample of texture generated by Portilla and Simoncelli's algorithm, with matching textural parameters.

In broad terms, the texture analysis/synthesis process of these papers can be described as follows: Run a lot of filters over the image, do some non-linear things to them, and measure a lot of statistics on the outputs. Also measure statistics on the pixel values themselves. Put all these numbers (the "features") into a big feature vector. Synthesize another texture that has a similar vector and you will hopefully have a similar looking texture. When you choose the right textural features (the ones that matter to humans), you get textures that look the same to humans. Figure 9 shows an example of an original texture, and a synthesized texture, using the method of Portilla and Simoncelli (2000), which involves joint statistics.

The problem of gloss perception under real world illumination has a similar statistical character to the problem of texture perception, it is interesting to perform some simple "textural" modifications of some glossy images, to see what happens to the appearance. Figure 10(a) shows an original photo of a chrome-plated sphere. The histogram of pixel values within the sphere is shown below. It covers a broad range of intensities as would be expected since it is merely reflecting a distorted image of the environment.

In Figure 10(b), the same photograph has been modified by selecting the circular region of the sphere and compressing

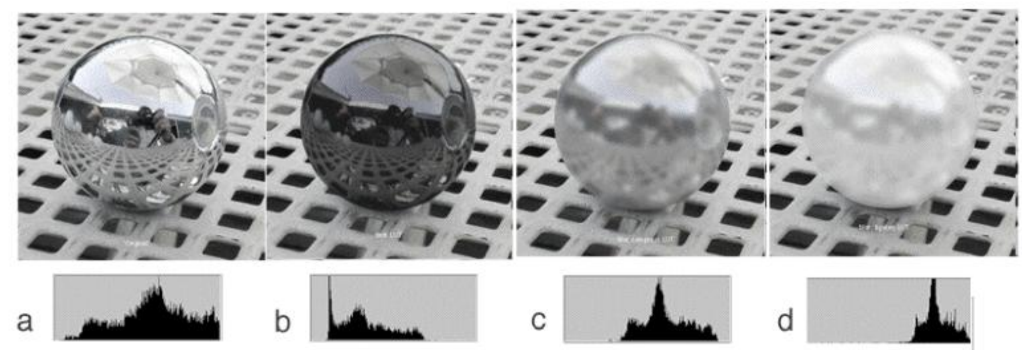

Figure 10: Four images created by modifying an original image. (a) The original image of a chrome-plated sphere. The histogram of pixel intensities (in the sphere image) is shown below. (b) the image after compressing the grayscale range of the sphere image. (c) The result of blurring the image, reducing the range, and adding a constant term. (d) The result of blur, more compression and a larger additive term.

the intensity range, so that everything becomes darker. The compressed histogram is shown below. The resulting image looks like a black shiny sphere, such as a billiard ball made out of black plastic. It should not surprise us that it is possible to make a chrome plated sphere look like a black plastic sphere since in both cases the dominant reflection is specular, with a different magnitude.

In Figure 10(c) the sphere region's histogram has been shifted upward and a Gaussian blur has been applied to that region. The sphere now takes on the appearance of brushed or sandblasted metal. Note that for an actual roughened sphere the blurring would take place as a convolution in the spherical domain rather than being uniformly applied in the image plane, but this physically incorrect blur still gives a good impression.

Figure 10(d) shows what happens if we take the image of Figure 10(c) and compress the histogram further in the direction of white. Now the sphere begins to take on a pearly appearance. This result makes sense if we consider that a pearl consists of many thin shells of slightly inhomogeneous transparent material, where Fresnel reflection and scattering can occur across the multiple layers.

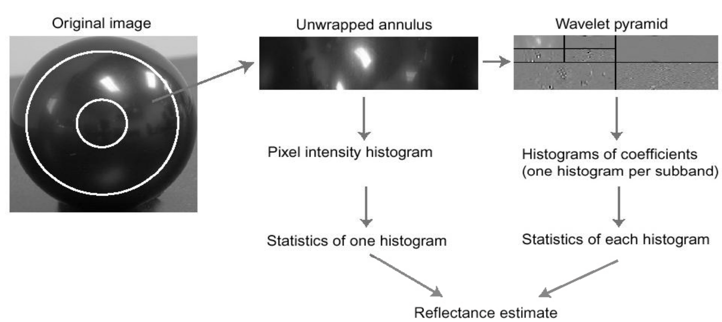

Figure 11: Flow diagram for extracting statistical features from an image of a sphere.

The experiments of Figure 10 suggest at least two domains that might useful for characterizing gloss. The first, in the image domain, is the shape of the luminance histogram. The second would be related to spatial frequency, where blur and sharpness give us cues about the material being viewed. Recall that Heeger and Bergen found it necessary to constrain their textures both in terms of subband statistics and in terms of luminance statistics.

Encouraged by these observations, we built a machine vision system for characterizing the optical qualities of surfaces viewed in real world settings (Dror et al., 2001). Figure 11 shows a flow chart of how we applied these basic ideas for analyzing and classifying the surfaces of spheres. For simplicity we took an annulus from our image and unwrapped it into rectangular coordinates. We then built a wavelet pyramid from the image. We derived statistics from the histograms of the wavelet subbands and also from the raw pixel luminance values. These measurements were used as features that were handed to a pattern classification system. We trained the system by showing it multiple examples of a finite set of spheres, viewed in multiple real-world environments. We did this with real photos of real spheres, and also with synthetic images of spheres rendered using panoramic data from real environments. For each image in the sample, we measured a set of feature vectors and attached the known label (e.g., "gray shiny"). The question was whether the system could learn to recognize a gray shiny sphere regardless of which environment it was in. That is, we tested for gloss constancy under variation in illumination conditions.

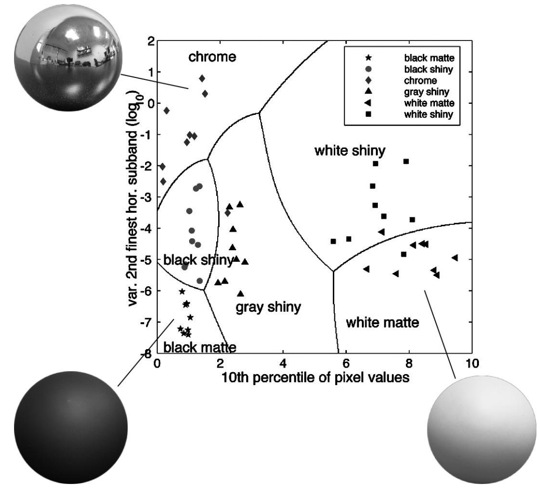

Figure 12: Classification of sphere images based on two features.

The features were chosen experimentally. Figure 12 shows the results for two particular features. The classification boundaries were done with a support vector machine, but the main point of interest here is the feature space itself. (Note that for better performance we used as many as 5 dimensions, but it is easier to visualize just two).

On the x-axis is the 10th percentile of the pixel luminance values within the sphere. On the y-axis is the variance of the second finest horizontal subband. The y-axis is basically a measure of spectral power in a high frequency band, and in some sense it is a stand-in for overall sharpness. There is nothing magic about this particular subband. The x-axis is more interesting. The 10th percentile of the luminance histogram tells us how dark the dark pixels tend to be. We had no preconceived idea that this would be a useful thing to measure, but it turns out to be quite diagnostic. For instance, it turns out that a chrome-plated sphere has a lot of very black pixels, and its 10th percentile score is about the same as that for a black matte sphere. The reason is that the world contains black objects, or regions that are in shadow and so appear black, and these are reflected in the mirrored surface. If the surface was not mirror-like, or if there was a diffuse component to the reflectance, then the blackness would be diluted, raising the floor of the histogram.

The two-component feature space of this diagram suggests a heuristic: for something to look chrome-like, it should have a lot of sharpness and blackness. It turns out that this heuristic is known and used by artists.

We have found it useful to look at surfaces with complex "mesostructure" (medium scale bumps and dips) like the two images of the sculpture shown in Figure 13. These are both renderings from the same 3D surface of a statue of St. Matthew. Indeed, the two images are based on a single rendering, the only difference being that the one on the left was put through an accelerating non-linearity (i.e., a lookup table), while the one on the right was put through a compressive non-linearity. This simple manipulation alters the skewness of the luminance histogram and greatly alters the appearance of the surfaces.

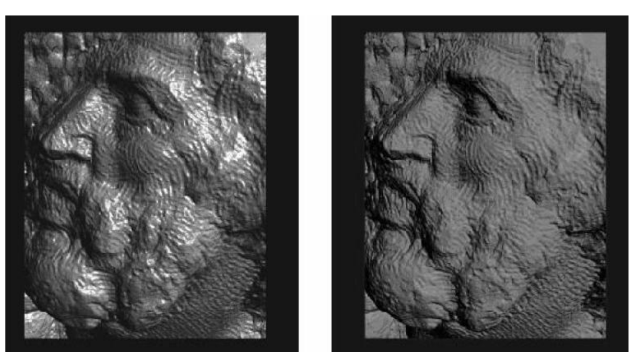

Figure 13: Two renderings of a sculpture with different skewness (courtesy Digital Michaelangelo Project).

For surfaces like these, skewness can be quite informative; dark glossy surfaces tend to have high skewness and light matte surfaces can have low or negative skewness. It is often useful to measure such quantities locally. It turns out that local skewness can be computed quite readily with filters such as those found early in the visual system.

Skewness, which characterizes the asymmetry of a distribution, is classically defined as the normalized third moment. However, there are other definitions of skewness, and we have no reason to think that this one is privileged. The point is to look at the balance between the positive tails and negative tails. It is the outliers that get greatest weight here, because the outliers are especially informative. Very bright points tend to be specular highlights; very dark points tend to be shadows, and so positive and negative outliers have quite different meaning.

We find evidence for mechanisms in human vision that respond selectively to positive or negative skewness. For example, we find a skewness aftereffect. After adaptation to a pattern of random dots, a surface subsequently viewed will have an altered appearance. After adapting to a positively skewed pattern, the surface will look lighter and less glossy than after adapting to a negatively skewed pattern (Motoyoshi et al 2007).

References:

Beck, J. & Prazdny, S. (1981). Highlights and the perception of glossiness. Perception & Psychophysics, 30(4), 407-410.

Bergen, J. R., and Landy, M. S. (1991) "Computational modeling of visual texture segregation," in Computational Models of Vision Processing (M. S. Landy and J. A. Movshon, eds.), pp. 253--271, Cambridge MA: MIT Press.

Berzhanskaya, J., Swaminathan, G., Beck, J., & Mingolla, E. (2002). Highlights and surface gloss perception [Abstract]. Journal of Vision, 2(7), 93a, http://journalofvision.org/2/7/93/, DOI 10.1167/2.7.93

Blake, A. & Bülthoff, H. (1990). Does the brain know the physics of specular reflection? Nature, 343, 165-168.

K.J. Dana, B. van Ginneken, S.K. Nayar and J. J. Koenderink, (1999) Reflectance and Texture of Real World Surfaces, ACM Transactions on Graphics, Volume 18, No. 1, pp. 1-34

Debevec, P. E. (1998). Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography. Proceedings of SIGGRAPH, 1998, 189-198.

Dror, R. O., Adelson, E. H., and Willsky, A. S. (2001). Surface reflectance estimation and natural illumination statistics. In Proceedings of IEEE Workshop on Statistical and Computational Theories of Vision, Vancouver, Canada.

Field, D. J., (1987) Relations between the statistics of natural images and the response properties of cortical cells. J. Opt. Soc. Am. A, 4:2379.

Fleming, R. W., Dror, R. O., & Adelson, E. H. (2003). Real-world illumination and the perception of surface reflectance properties. Journal of Vision, 3(5), 347-368,

Glassner, A. S., (1995) Principles of Digital Image Synthesis, Morgan Kaufmann, San Francisco.

Hartung, B., and Kersten, D. (2002). Distinguishing shiny from matte [Abstract]. Journal of Vision, 2(7), 551a

Heeger, D. J., and Bergen, J. R. (1995). Pyramid-based texture analysis/synthesis. Proceedings of SIGGRAPH. 1995, 229—238.

Julesz, B., (1962) "Visual pattern discrimination," IRE transactions on Information Theory, vol. 8, pp. 84—92.

Julesz, B., (1981) Textons, the elements of texture perception, and their interactions. Nature, 290:91—97.

Malik, J., and Perona, P (1990). Preattentive texture discrimination with early vision mechanisms. Journal of Optical Society of America, 7(2):923-932

Motoyoshi, I., Nishida, S., Sharan, S., Adelson, E. H., (2007) Image statistics and the perception of surface qualities. Nature, 447:206-209.

Nishida, S., and Shinya, M. (1998). Use of image-based information in judgments of surface-reflectance properties. Journal of the Optical Society of America A, 15, 2951–2965

Pellacini, F., Ferwerda, J. A., and Greenberg, D. P. (2000). Toward a psychophysically-based light reflection model for image synthesis. Computer Graphics, 34(2), 55-64.

Portilla, J., and Simoncelli, E. P., (2000) A parametric texture model based on joint statistics of complex wavelet coefficients. Int. J. Computer Vision, pages 49-71,

Ramamoorthi, R., and Hanrahan, P. (2001) A signal-processing framework for inverse rendering. Computer Graphics (SIGGRAPH).

Rudermanm D. L., (1994). The statistics of natural images. Network: Comput. Neural Syst., 5:517-48.

Ward, G. J. (1992). Measuring and modeling anisotropic reflection. Computer Graphics, 26(2), 265—72.

Ward, G. J. (1994). The RADIANCE lighting simulation and rendering system. Computer Graphics, 28(2), 459-472.

Zhu, S. C., Wu, Z. N., and Mumford, D. (1997) Minimax entropy principle and its application to texture modeling. Neural Computation, 9(8):1627- 1660.