Proceedings of the. SPIE Vol. 4299, pp. 1-12, Human Vision and Electronic Imaging VI, B. E. Rogowitz; T. N. Pappas; Eds. (2001)

On Seeing Stuff: The Perception of Materials by Humans and Machines

Edward H. Adelson

Massachusetts Institute of Technology, NE20-444H, Cambridge MA 02139

ABSTRACT

The perception of objects is a well-developed field, but the perception of materials has been studied rather little. This is surprising given how important materials are for humans, and how important they must become for intelligent robots. We may learn something by looking at other fields in which material appearance is recognized as important. Classical artists were highly skilled at generating convincing materials. The simulation of material appearance is a topic of great importance in 3-D computer graphics. Some fields, such as mineralogy, use the concept of a habit, which is a combination of shape and texture, and which may be used for characterizing certain objects or materials. We have recently taken steps toward material recognition by machines, using techniques derived from the domain of texture analysis.

Keywords: Perception, materials, reflectance, BRDF.

INTRODUCTION: THINGS AND STUFF

Ask someone what vision is for and you may get an answer about recognizing objects. Few people will tell you that vision is about recognizing materials. Yet materials are just as important as objects are. Our world involves steel and glass, paper and plastic, food and drink, leather and lace, ice and snow, not to mention blood sweat and tears. Nonetheless, if you peruse the scientific literature in human and machine vision, you will also find a great deal of attention paid to the problem of recognizing objects, and very little to the problem of recognizing materials. Why should this be?

Perhaps it is related to the general preference we have for talking about things rather than stuff. Linguists distinguish count nouns, such as chairs, from mass nouns, such as snow. Chairs are objects that can be counted, while snow is a material of unspecified extent. Our world contains both things and stuff, but things tend to get the attention. This prejudice is certainly true in perception. Even at the low-level stages of vision, Julesz speaks of his textons as the quarks of vision, and Marr built his primal sketch on lists of individuated edges and blobs. There seems to be a desire to put vision on a firm foundation by emulating particle physics. However, classical physics is built on stuff (e.g. mass, heat, entropy), and Adelson and Bergen[1] have argued that early vision can be thought of as the extraction of stuffish properties in an image. As we get to high-level vision, physical materials become important, and they deserve more attention than they have received.

THE IMPORTANCE OF MATERIALS

To appreciate the ubiquitous importance of materials in everyday life, one can survey the advertisements in a magazine. Here is an ad for skin cream, or for silk scarves, or for milk, or for perfume, or for house paint, or for eye shadow, or for laundry detergent. Consider the detergent: it is a material (detergent) that removes another material (dirt) from a third material (fabric). It improves the fabric s looks, and in addition, it may influence a fabric s mechanical properties, such as softness and flexibility, and even electrical properties such as static cling. Similarly, shampoo may give hair a silky shiny look, and extra body, i.e., it may modify the optical and the mechanical properties.

Looking in a recent issue of a consumer magazine I found articles about iced tea, lipstick, paint, and glue. These are just a few of the materials that pervade our lives. We often care about a great many details of a material s properties. For example, here is how the magazine described one brand of lipstick: Ultima II Lipsexxxy Lipcolor (sheen/frost).   Felt creamy while applying but became powdery as it "dried." Easy to apply, blot-resistant, matte, opaque, not slippery. Felt light on lips. Slightly easier to launder than others. In this short description, numerous optical, mechanical, and chemical properties are discussed, and they are all important to the purchaser. Another magazine has a review of salami, and it describes Marco

Polo brand: Dull appearance, slightly slimy mouth feel; good tasting though quite salty. Again, optical, mechanical, and chemical descriptors are used.

Do we simply care about materials because we are made of flesh, and have to eat food, wear clothes, and excrete body fluids? How would a robot feel? If it were a legged robot that walked over terrain, it would need to figure out what it was walking on. The walking technique will differ for grass, mud, gravel, or concrete. Suppose the robot is lucky enough to be walking on a concrete sidewalk, and it sees a dark patch ahead. Is it a patch of asphalt? A puddle of water? A patch of clear ice? The answer may be critical. Another example: if a robot colony is set up on Mars to build a factory, it must mine the right materials and process them correctly.

Closer to home, suppose we have a domestic robot that will clean up the kitchen, and suppose that this robot encounters a patch of white stuff on the floor. It might be spilled milk, a crumpled napkin, a pile of flour, or some cream cheese. In each case the robot must use the appropriate cleaning technique. A sponge will be useful for the milk but not for the crumpled napkin. A vacuum cleaner will deal nicely with the flour but not with cream cheese. The robot should be able recognize the materials, just as we can, and it should be able to reason about their physical properties before handling them

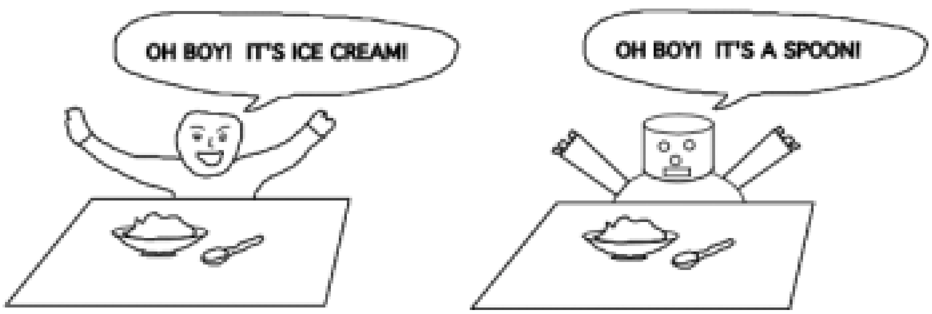


Figure 1. A child can recognize a complex material such as ice cream, based on its optical and mechanical properties. Today s machine vision systems can recognize certain objects, but are rarely useful in recognizing materials.

ASSESSING MATERIAL PROPERTIES

Humans can infer material properties using all the senses. Wool has a certain look and a certain feel, and when wet it has a certain smell. To test whether a wall is solid wood or mere paneling, we can knock on it and listen to the sound. We may squeeze a pear to decide whether it is ripe, and then verify our judgment with the taste and texture when we bite into it.

The use of multiple information sources is formalized in mineralogy. The mineralogist in the field can characterize a mineral by numerous means, including scratching it, fracturing it, heating it, and tasting it. Here are the properties of carnallite, as described in a field guide for rocks and minerals: Soft (2.5), no cleavage, but has conchoidal fracture. Transparent with vitreous or greasy luster. Extremely phosphorescent. As it is deliquescent it disintegrates quickly if exposed to air. Very soluble in water. Has a bitter salty taste. Fuses easily, turning the flame violet (potassium).

The present paper is about vision, so it is interesting to consider the vocabularies that have been developed for the visual appearance of materials in various fields. In identifying rocks and minerals, a mineralogist will describe the luster (meaning the optical quality of the surface), with words like resinous (like plastic), adamantine (like diamond), greasy, pearly, silky, vitreous (glassy), metallic, submetallic, dull, earthy, or chatoyent (like a cat's eye). When broken, the fracture may be described as uneven, conchoidal (shell-like), hackly (like cast iron), or splintery. In addition, rocks and minerals have habits, which are descriptions of their typical forms. They include prismatic, massive (no particular form), acicular (needle-like), reniform (kidney-like spherules), bladed, dendritic, granular, fibrous, encrusting, colloform, porous, concretionary, botryoidal (like grape-bunches), foliated (leaves or layers), scaly, felted, hairlike, stalactitic, nodular, columnar, plumose (feathery), microcrystalline, platy (flat thin plates), reticulated, lamellar, mammillary, saccharoidal (like sugar), ameboid, oolitic, or pisolitic.

Few of us have a visual vocabulary that is as formalized as the mineralogist s, but we are all experts at stuff perception from an early age. What would childhood be without mud, snow, or peanut butter? If we are to emulate human vision with computer systems, we have a long way to go. Because machine vision has concentrated on objects, to the neglect of materials apperance we have the situation shown in the cartoon in figure 1. The human boy is delighted to see the ice cream, which he immediately recognizes based on its visual appearance. The robot, however, has spent its life learning to identify objects, and hasn t a clue about ice cream. It probably can t even mis-identify the ice cream because it lacks the concept of a material in the first place. The best it can do is tell us about the spoon.

A science of material perception would be helpful in many fields that depend on images. For example, when a satellite sends back images of the Martian surface, planetary scientists look at the pictures and try to figure out what the materials are, and what physical processes have acted upon them. Figure 2 shows a recent set of Mars photographs that indicate a history of erosion, perhaps by water formerly on the surface. At present, the main way to interpret such images is to show them to a trained individual and ask what does this look like to you?

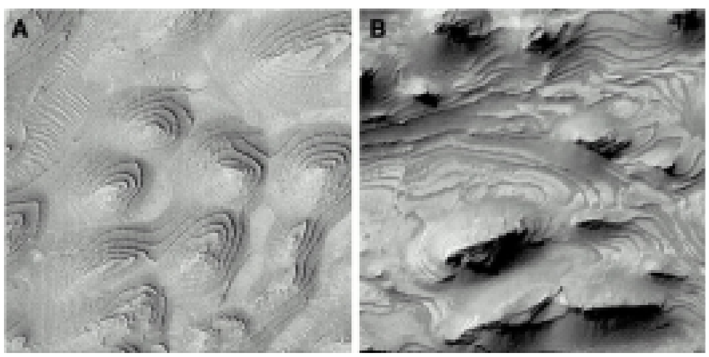


Figure 2. Mars terrain photos obtained by satellite. The appearance of the surface suggests a history of erosion.

WHAT CAN VISION DETERMINE ABOUT MATERIALS?

Optical and mechanical properties are perhaps the major ones we infer by looking at pictures. Consider the object shown on the left in Fig. 3. You may recognize it as a bottle opener. Even if you don’t recognize the object, you can make a pretty good guess as to its shape and its material. It appears to be made of a shiny metal like steel. You can guess that if you pick it up it will feel hard, rigid, and slightly cold. You can guess that if you use it to open a bottle, it will be strong enough to hold its shape. Of course, you might be wrong: it might be made of plastic that has been plated with a silvery coating. But most of the time you guess right about materials, just as most of the time you guess right about object shape and identity.

The image of an object results from a combination of the surface shape, the surface reflectance, the distribution of lights in the environment, and the observer s point of view. Untangling these multiple sources to retrieve shape and reflectance is an amazing feat.

Figure 3: Left: a metal bottle opener. By looking at it, one can estimate its shape, as well as the optical properties of its surface. Right: a chrome-plated sphere shown in two settings. The pattern within each image is entirely different, but each gives the impression of a mirror-like material.

As in most of vision, material perception is automatic and effortless, and therefore may seem initially like an easy problem. However, it is hard. One of the inherent difficulties is shown by the two pictures of spheres on the right. Even though these are both the same shape, and even though they're made of the same material, they produce entirely differ patterns of light in the image. A chrome-plated sphere simply reflects the world around it, as a distorted image. Thus, every time you see one it looks completely different (at the pixel level) from all the other chrome spheres you have seen in the past. Yet there is something about it that looks the same

Now most people are completely unaware that every time they see a chrome sphere they are seeing a picture of the room, with a little picture of themselves right in the middle. The sphere just looks like shiny metal. Even though every pixel can be traced to a point in the surrounding room, the metallic quality seems to lie in the sphere itself[2]. There is no way to extract this property through a local operation, since it depends on the full pattern. And of course, there is no way to proceed with a template match as one might do in recognizing, say, the letter A. Instead there is some sort of quasi-textural quality that is common to the many pictures of chrome spheres.

The problem is not unique to the mirror-like chrome surface. Consider the three spheres shown in fig. 4. The first has a mirror surface. If such a surface is roughened on a fine scale, the reflected image becomes blurred. In the simplest model, we can consider that each point on the sphere is taking a weighted average of the pixel values in the world over some angle. The simplest idea is that our little sphere is surrounded by a big outer sphere, which is the world as seen from the little sphere. The big sphere s pixel values are convolved with a blurring function and mapped to the appropriate location on the little sphere s surface and projected to form the image. (The big sphere is sometimes called an environment map). The same convolution model holds for matte surfaces as well. In the case of a Lambertian surface, each point on the sphere takes an average over a full hemisphere. (Note that the Lambertian hemisphere is centered on the normal to the sphere s surface, while the blurring function for the roughened metal is centered on an angle twice as large, so the two processes cannot be reduced to a single convolution).

The point is that the appearance of every sphere depends on the environment in which it is viewed, and sometimes this dependence is so great that it would seem hopeless to make sense of the sphere s reflectance properties without knowing the environment first. Yet for humans, this task seems fairly easy.

Figure 4. Three spheres, photographed in the same room with the same lighting.

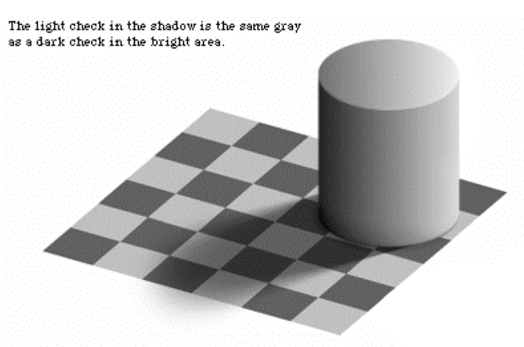
The problem of recognizing a surface s reflectance qualities is in some ways similar to the problem of recognizing a texture. Suppose we have a Gaussian process that generates images of white noise. Each noise image will be completely different from the other noise images at a pixel level, but will look much the same at a textural level. Each image is a sample from the same random process, and its appearance depends on the parameters of that random process. The same is true for a sphere photographed in a randomly selected environment. The sphere-plus-environment process generates the image. The world around us has certain statistical qualities, and the sphere s reflectance parameters map these into image qualities.

In unusual settings, we can be fooled. For example, a shiny surface will not look very shiny if photographed in an environment with broad diffuse illumination. Portrait photographers use large diffuse light sources when they want to minimize the shine of the skin. This trick also minimizes the visibility of wrinkles.

CONFIGURATION AND CONTEXT

To fully characterize the reflectance properties of a point on a surface, one must determine the bi-directional reflectance distribution function (BRDF). This is a function of four continuous variables that describes the relative intensity of light emerging in every direction for each possible incident angle. It can be a huge and hopelessly complex function. Fortunately, many materials can be approximately described with very simple BRDF s. Perhaps the simplest example is the Lambertian surface, an ideal matte surface that reflects light uniformly in all directions regardless of the angle of incidence.

A Lambertian surface has only one parameter, which is its albedo: the percentage of incident light that is reflected. Thus, describing the albedos in a scene amounts to assigning a scalar to each point. However, even though albedo may be defined as a pointwise physical property, it cannot be computed from an image using a pointwise calculation. The larger context is essential. Consider the image[3] on the left of fig. 5. The cylinder casts a fuzzy shadow onto the checkerboard. The actual gray shade of the white square in the shadow is the same as the gray shade of the dark squares outside the shadow (i.e., the pixel intensities are the same in the image). This powerful illusion results from the operation of the lightness constancy mechanisms, which seek to determine the albedo of the surface in the scene, and to ignore the albedo of the pixels in the image. In order to do this, the visual system must take into account a great number of configural cues about what is paint and what is shadow[4,5].



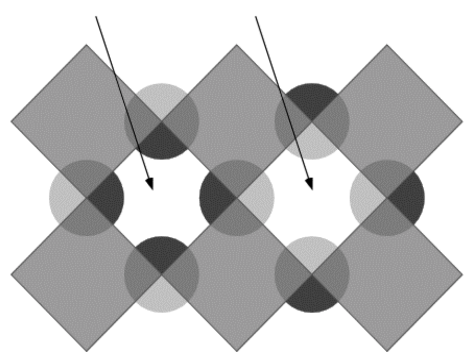


Figure 5. Left: the light square in the shadow is the same shade of gray ink as the dark squares outside the shadow. Right: the two diamond-shaped regions marked by the arrows are the same white color.

The importance of context is also shown in the image[5] on the right of fig. 5. The two regions marked by arrows are identical; indeed they are simply the color of the white page. However, they appear quite different, due to the configuration of gray levels and X-junctions around them. Indeed, it is not even possible to compare the two regions in terms of albedo. A new perceptual variable, a kind of haziness, has become apparent. Thus at least two values will have to be used to describe each point in the scene.

When we consider more complex surfaces, the importance of context continues to be apparent. Figure 6 shows a synthetic image of a shiny sphere on the left. On the right is the same sphere with the two specular highlights removed. Now the entire sphere looks matte. Observe that the sense of gloss or matteness spreads across the entire sphere, and is not restricted to the specularities themselves. Again, the perceptual quality at a given location in an image can be influenced by pixels at distant locations.

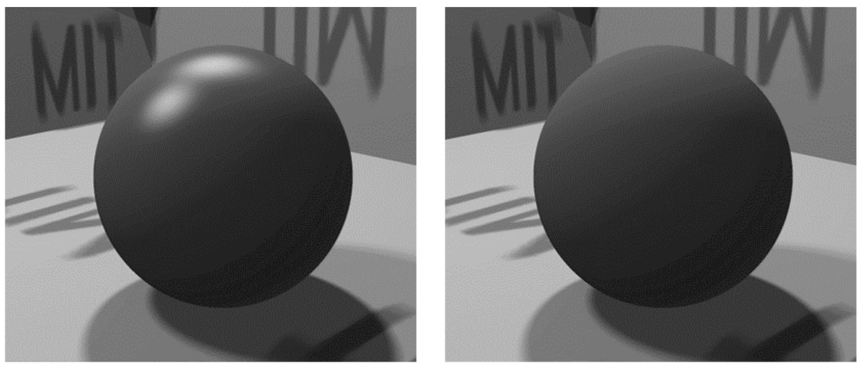


Figure 6. Left: an image of a shiny sphere, generated by computer graphics. Right: the same sphere, with the specularities removed. The appearance of the entire sphere s surface changes. (Images by Roland Fleming, after a demonstration by Jacob Beck).

Visual cues often tell us about more than just optical qualities. In particular, the mechanical property of a material is often expressed in its image. A good example is fabric, which drapes and folds in different ways depending on how stiff, thick, or elastic it is. Another example is shown in Figure 7, which depicts two viscous white substances: hand cream and cream cheese. In both cases, a blob of the material has been swirled with a spoon. The hand cream returns to a smooth, even surface, while cream cheese retains the peaks and valleys.

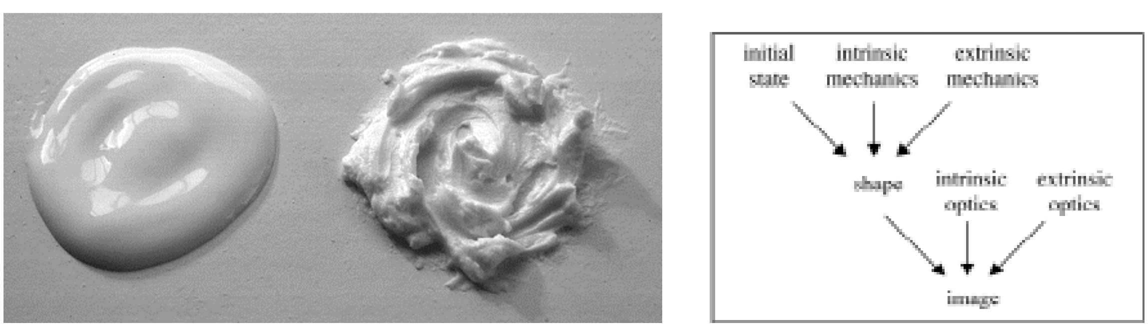


Figure 7. Left: a blob of hand cream and a blob of cream cheese. Both have been swirled with a spoon. Right: a flow diagram showing how an image depends on mechanical and optical properties of the object and the environment.

On the right is a flow diagram showing that how optical and mechanical aspects of the world are combined with the optical and mechanical aspects of the material to produce an image. Let us start with a chunk of material in some initial state. Its shape is then changed by the outside forces that act upon it over time (e.g., the swirling by the spoon), as well as the mechanical properties of the material itself (e.g. elasticity, viscosity, etc.). We can refer to these as the extrinsic mechanics and the intrinsic mechanics. The combination leaves the material in a certain shape. The material is then illuminated by some distribution of light in the environment, and it reflects, refracts, or absorbs the light according to its optical properties. In addition, the observer (or the camera) is positioned at some viewing point in space, looking in a certain direction with a certain focal apparatus. The optical properties inherent in the material are the intrinsic optics, while the lighting and viewing conditions are extrinsic optics. Shape, intrinsic optics, and extrinsic optics combine to form an image. (Of course, there can also be chemical, thermal, and other processes at work, but the mechanical and optical ones are often the most prominent).

It is impossible to think that the observer can retrieve all of the variables that went into making the picture look as it does. In particular, we may have only rough ideas about the extrinsic mechanics (the history of distortions the blob underwent) and the extrinsic optics (the distribution of lights in the environment). However, we can do a reasonably good job of guessing the intrinsic mechanics and intrinsic optics of the materials. This is fortunate, since the intrinsic qualities are usually those that matter most.

Food photography is an interesting sub-industry within the photography business, and it places a particular emphasis on material appearance. There are photographers who specialize in food, and stylists who specialize in optimizing its appearance for the photographers. The food stylist is in charge of giving the mashed potatoes the proper thick, moist, warm look, by cooking, shaping, spraying, or whatever is needed. The photographer arranges the lights and the camera to capture the look. In preparing the mashed potatoes, the stylist establishes by intrinsic optics and mechanics (by cooking and mashing), and then uses extrinsic mechanics to sculpt the potatoes to look a certain way. The photographer is mainly in charge of the extrinsic optics: the lights and the camera. Photography catalogs have various supplies to support the profession of food photography. Here is a sample entry: Trengrove Aqua Gel: Use as is for a variety of effects, ranging from water beads, spilled water, melted ice, ice formations, and melted plastic. Applies easily to beverage containers to give a natural, cool, refreshing look. It is wonderful to think that a natural, cool, refreshing look can come from a jar of gel.

RETURNING TO HABITS

One may think of the mashed potatoes as having a habit: a particular combination of shape and texture that signifies mashed-potato-ness. Let us return for a moment to this concept, which was earlier used in the discussion of rocks and minerals. The notion of habit is often used in characterizing plants. On the left of figure 8 is a drawing of the typical shape of a pin oak in winter. It is not a shape that can be template matched, since the particular branches are different for every tree. It is not a texture either, since there is a top and a bottom, and a textural quality that changes from one part to another.

On the right are two pictures: a melanoma and a normal mole. They seem to be made of a different kind of stuff, and this is apparent because of a combination of their shape, texture, and color. A popular mnemonic for the visual character of a melanoma is ABCD: Asymmetrical shape; Borders that are irregular or indistinct; Color that is variable; Diameter greater than 6 millimeters. This too can be thought of as a habit.

One might argue that the hand cream and cream cheese are characterized by their habits as well. They have no fixed form, and no constant visual texture. The broad center of a hand cream blob has a different visual character than the roundly curved boundaries. When everything is assembled into a coherent image, the hand cream s qualities become apparent.

Figure 8: Left: the habit of a pin oak. Center and right: A melanoma can be visually distinguished from a normal mole. The combination of shape, color, and texture are used to make the diagnosis.

UNDERSTANDING HOW IMAGES ARE MADE

To build a science of material appearance, it is useful to consider work in related areas. I list a few here.

Ecological optics: Given the fundamental physics of materials, one can ask what particular materials are common in the real world, what forms they commonly take, and what patterns of light tend to illuminate them. From these starting points, one can ask how scenes in the world map to images. This is the forward problem of how images are made. Vision is the inverse problem of figuring out what made an image. There has been recent progress in characterizing natural materials in the laboratory [6,7]

3-D computer graphics: A large body of work in computer graphics has been devoted to understanding reflectance and making things look like they're made of the right stuff [8,9]. Plastic is easy, but flesh is hard. It is no accident that the characters in today s blockbuster animations are often plastic toys or hard-shelled insects. It is very difficult to make humans; their skin and hair must look just right, and they must move and deform correctly as well. The degradation of materials is also needed to make them look natural, and this is a complex problem to simulate[10].

The bottom line for most computer graphics is: does it look right to a person? Therefore, much of computer graphics is embedded in a subjective psychophysical feedback loop. This is especially true for the rendering of materials. Since physically correct rendering is impossibly expensive, the researchers try out various tricks and shortcuts, and ask themselves whether it looks right; then they modify the procedure and look again. Unfortunately, the methods of 3-D computer graphics can be hard to translate into an understanding of human vision. The optimization process is occurring a few steps removed from the image itself, and so it is rarely clear what aspects of the 2-D image are responsible for a given perceptual impression.

Traditional painting: the old master painters were highly skilled at portraying the materials in the scenes they painted. They knew how to make a velvet robe look velvety, and how to make a silver goblet look silvery. This is very difficult to do. Note that even an amateur can draw a shape well enough for an object (like a goblet) to be recognized. However, portraying materials realistically is extremely challenging. Our visual systems are exquisitely sensitive to the way materials look, and will not be satisfied by imperfect approximations. Artists are taught certain tricks, but mostly they just look very carefully and over the years learn how to replicate the pattern of light in the image. Unlike researchers in 3-D computer graphics, the artists work directly with the image data, and therefore may have a great deal of knowledge that could be helpful for understanding human vision.

2-D computer graphics: In a program like Photoshop, there are a variety of special effects tools for making beveled objects that appear to made of plastic, metal, rubber, etc. These tools run in real time, and are based on a set of 2-D image processing tricks rather than on true 3-D material rendering. Because these tricks work with low-level operations rather than with sophisticated 3-D rendering systems, they work near the level of the image itself. Therefore they may offer interesting insights into human vision, since vision begins with the image data.

Photography: A studio photographer takes control of light and camera. With experience, a photographer learns how to control the way materials look. A raking light will bring out fine surface texture. A diffuse light will minimize fine texture and will then to make materials look matte rather than shiny. A hair light can be used in a portrait to bring out a glow around the boundary of the head. And so on. By studying the tricks that photographers know, we may learn the tricks that the human visual system uses in interpreting the photographs.

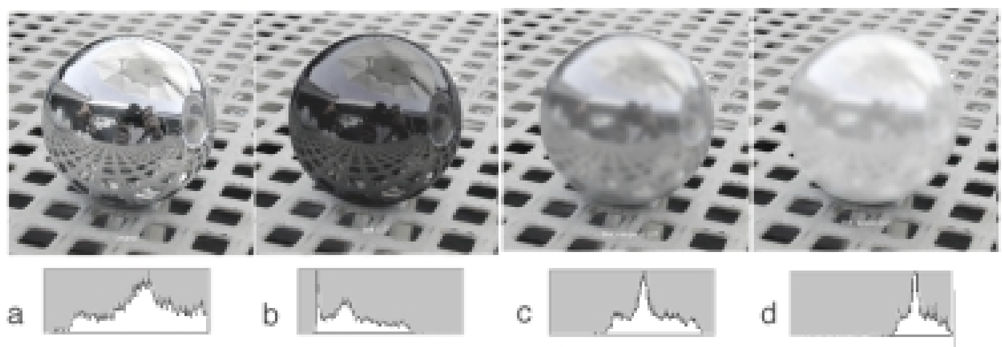


Figure 9: Four images created by modifying an original image. (a) The original image of a chrome-plated sphere. The histogram of pixel intensities (in the sphere image) is shown below. (b) the image after compressing the grayscale range of the sphere image. (c) The result of blurring the image, reducing the range, and adding a constant term. (d) The result of blur, more compression and a larger additive term.
.

INFORMATION IN IMAGE STATISTICS

As noted earlier, the problem of material appearance bears some similarity to the problem of texture perception. Multiple stochastic processes, including the distribution of lights, combine to produce a particular image of a particular object. For this reason, it is interesting to manipulate some simple image statistics of an image and see what happens to the material appearance.

Figure 9(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 world.

In fig. 9(b) the same photograph has been modified by selecting the circular region of the sphere and compressing 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 pool 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. (Note that for a real plastic sphere, the specular reflections would get brighter toward the edges due to the increased efficiency of Fresnel reflection at grazing angles).

In fig. 9(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 will take place as a convolution in the spherical domain rather than being uniformly applied in the image plane. (Equivalently a space-variant convolution could be applied in the image domain.) However this example shows that even a simple uniform convolution produces a reasonable impression of a roughened metal sphere.

Figure 9(d) shows what happens if we take the image of fig. 9(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.

The experiments of fig. 9 suggest at least two domains that might useful for characterizing material appearance. The first, in the image domain, is the shape of the intensity histogram. The second would be in the frequency domain, where blur and sharpness give us cues about the material being viewed. Space-domain and Fourier-domain statistics have been widely used in texture analysis. However, in recent years it has been found that wavelet statistics are more useful than Fourier statistics in characterizing textures. Heeger and Bergen[11] found that two textures that matched both in terms of pixel statistics and wavelet coefficient statistics tended to look similar, and others have built on this notion to develop successful texture synthesis techniques[12,13]. Image statistics have been proposed as useful in understanding the human perception of materials as well[14].

Ron Dror and I have followed a similar strategy in a machine vision system for characterizing the optical qualities of surfaces viewed in real world settings[15]. For simplicity, we have concentrated on images of spheres. Since the environment tends to contain a broad range of luminances and numerous sharp edges, we expect these properties to manifest themselves in the specular reflections of a sphere. If the sphere has a finely roughened surface the specularities will be blurred and the high frequency components will be removed. The diffuse component of the surface will provide a broad, gradually varying additive component. A shiny black ball will have almost no diffuse component whereas a shiny white ball will have a great deal. The white diffuse component will tend to serve as a floor that raises the intensities throughout the ball and simultaneously lowers the contrast ratio of the specular reflections. The specular component itself will be larger or smaller depending on the refractive index of the material, and whether the material is metal or dielectric.

Figure 10 shows a flow chart of how we actually have applied these basic ideas for analyzing and classifying the surfaces of spheres. Ideally we would work in a coordinate system appropriate to the spherical surface. However, for simplicity in the present work, we take an annulus from our image and unwrap it into rectangular coordinates. The unwrapped annulus is then analyzed in terms of the statistics of pixel intensity distributions and the statistics of wavelet coefficient distributions. These measurements serve as features that are handed to a pattern classification system such as one using support vector machines.

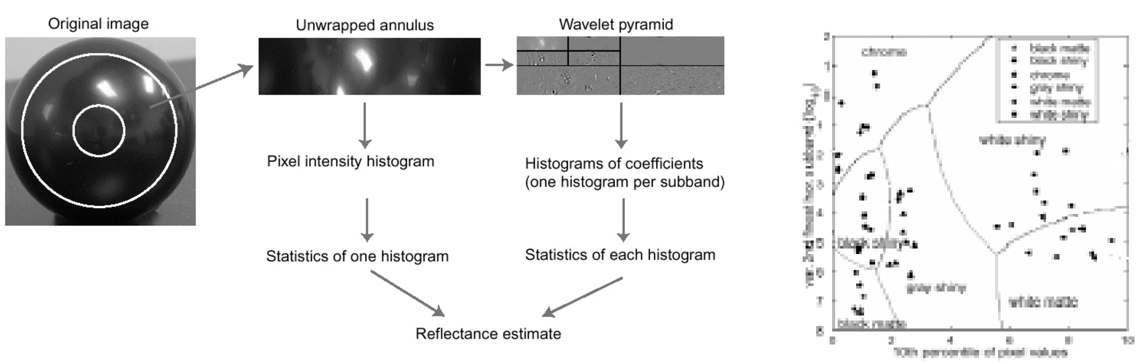


Figure 10: A flow diagram for estimating the reflectance parameters of a sphere photographed under unknown lighting. Right: classification results using two measurements.

We have tested this approach using both synthetic and real images. The synthetic images were made using the Ward model (a variant of the well-known Phong model) and a set of panoramic illumination images collected by Debevec[16]. We chose 6 sets of Ward parameters, corresponding to black matte, black shiny, gray shiny, white shiny, white matte, and chrome. Each sphere type was rendered under the same set of 9 lighting distributions, taken from indoor and outdoor scenes. The lighting conditions were quite varied, and so the images of a given sphere type appeared quite different from each other. Our hope was that a set of features could be found to that would allow the system to ignore the lighting and classify according to reflectance parameters.

On the right of fig. 10 is a plot of the categorization possible using only two features, namely, the 10th percentile of the intensity histogram and the variance of one of the wavelet subbands. Even with these two basic features, the classification is quite good. Using 6 features, the classification becomes almost perfect, and is comparable to that shown by humans for this particular set of stimuli.

We performed a similar experiment using pictures of real spheres in real scenes. We used nine spheres with different reflectances, and photographed them in various indoor and outdoor locations. Note that the spheres, being real, did not obey Ward s model or any other simple reflectance model. This is all right since the classification procedure does not depend on a model. (This is one reason for preferring a classification task rather than a parameter estimation task). After training on a subset of the images, we tested the system with images of familiar spheres in novel environments. Again, with 6 features, the classifier did about as well as humans.

There is a major limitation, of course: we have only worked with spheres. It should be straightforward to generalize to other surfaces in which the surface normals are known everywhere, since the pixel values can be mapped onto a virtual sphere, producing an equivalent sphere image (assuming distant illumination, and assuming the surface normals include a large range to cover the sphere.) However, this is still quite restrictive, since in passive vision applications one rarely has knowledge of surface normals. Nonetheless we find the results quite encouraging, since they indicate that simple image features can be profitably employed in characterizing reflectance in unknown lighting.

ANALYSIS — BY - SYNTHESIS

Analysis-by-synthesis is an approach that has been used to estimate parameters in simple polyhedral images, including shape, lighting, and albedo[17]. In pilot work in my lab, Marshall Tappen is considering a toy world that includes smooth blobby shapes with Phong/Ward surfaces. The photograph of fig. 11(a) shows an example: a bath-oil capsule in the shape of a penguin, with a white shiny surface. We are exploring an analysis-by-synthesis approach to modeling this and similar scenes. We assume that the contour of the penguin is known; in the present case we extracted it by hand, but for simple scenes like this one it could be automated. We assume a distant point source, and a Ward reflectance model plus an ambient illumination term.

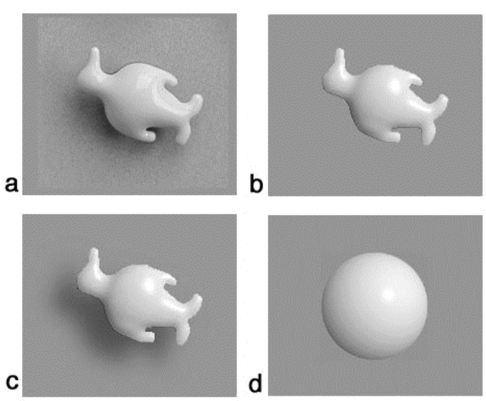


Figure 11. (a) A photograph of a penguin shaped bath-oil capsule. (b) synthesized penguin image, based on estimated shape and Ward reflectance parameters. (c) synthesized penguin with synthesized drop shadow. (c) sphere rendered using the same lighting and reflectance parameters.

Given the contour, we wish to estimate the blob s shape, the light source direction, and the reflectance parameters. We inflate the contour into a blobby shape by one of various methods. In this case we use a grassfire algorithm to compute distance from the contour, and then apply a smoothing algorithm. This produces an initial guess for the shape. We then use heuristics to make initial estimates of the light source direction and ambient illumination. Given these starting points, we synthesize the predicted penguin image, and run an optimization routine to find the shape and reflectance parameters that give minimal squared error. The result is shown in fig. 11(b). Although we do not have ground truth, the estimated shape and reflectances appear to be good.

To make a more realistic replica of the scene, we added a drop shadow, as shown in fig. 11(c). This was done with a simple trick often used in 2-D graphics programs: we replicated the contour, filled it with gray, blurred it, and displaced it. We estimated the shadow parameters in a feedback loop. (Note: the fact that the shadow is blurred shows that the prior assumption of point source illumination is incorrect; indeed, the shadow gives information on the size of the extended source. We will impose more consistency as the work progresses). Finally, to give a better intuition of the recovered reflectance parameters, we used them to render the sphere of fig. 11(d).

Although the penguin is just a first step, it offers encouraging possibilities for machine vision. Within a restricted domain of scenes, it is possible to retrieve both shape and reflectance parameters.

CONCLUSIONS

The perception of materials is an important area that is still in its infancy The importance of material appearance is clear, and there are many fields devoted to controlling the way materials look (e.g., fabrics, food, paint), and in using images to convey material properties (e.g., computer graphics, art, photography). Mechanical and optical properties are the main ones that humans derive from image information. Recent work suggests that concepts used in texture analysis may be usefully applied to the problem of material appearance.

ACKNOWLEDGMENTS

This work was supported by Unilever, by NIH Grant EY11005-04, and by a grant from NTT to the MIT AI Lab. Peter Kaplan, Jan Koenderink, Whitman Richards, Ron Dror, and Marshall Tappen provided useful discussions.

REFERENCES

1. E. H. Adelson and J. R. Bergen, The Plenoptic Function and the Elements of Early Vision, in M. Landy and J. Movshon, eds., *Computational Models of Visual Processing*, Cambridge, MA: MIT Press, 1991.
2. S. A. Shafer and B. A. Maxwell, Color as a Carrier of Physical Information, in *Color Perception: Philosophical, Psychological, Artistic and Computation Perspectives*, pp. 52-71, S. Davis ed., New York: Oxford University Press, 2000.
3. E. H. Adelson, at http://www.illusionworks.com, 1995.
4. E. H. Adelson, Perceptual Organization and the Judgment of Brightness, *Science*, 262:2042-2044 (1993).
5. E. H. Adelson, Lightness perception and lightness illusions," in *The Cognitive Neurosciences*, M. Gazzaniga, ed., pp. 339-351, MIT Press, Cambridge, MA, 2000.
6. R. Lu, J. J. Koenderink, and A. M. L. Kappers, Surface Roughness from Highlight Structure, *Appl. Opt.*, **38**, pp. 2886-2894, 1999
7. .K. J. Dana, S. K. Nayar, B. van Ginneken, and J. J. Koenderink, Reflectance and Texture of Real-World Surfaces , ACM Transactions on Graphics, **18**, pp. 1-34, 1999.
8. R. L. Cook and K. E. Torrance, A Reflectance Model for Computer Graphics, *Computer Graphics*, **15**, pp. 307-316, 1981.
9. A. S. Glassner, *Principles of Digital Image Synthesis,* vol. 2, Morgan Kaufmann, San Francisco, 1995.

10. J. Dorsey and P. Hanrahan, Digital Materials and Virtual Weathering, *Scientific American*, pp. 46-53, February 2000.
11. D. J. Heeger and J. R. Bergen, Pyramid-based Texture Analysis/Synthesis, *Computer Graphics (SIGGRAPH)*, 1995.
12. E. P. Simoncelli and J. Portilla, Texture Characterization via Joint Statistics of Wavelet Coefficient Magnitudes, in Fifth IEEE Int l Conf. On Image Processing, Chicago, IL, Oct. 4-7, 1998.
13. J. S. De Bonet, Multiresolution sampling procedure for analysis and synthesis of texture images, *Computer Graphics (SIGGRAPH)*, pp. 361-368, 1997.
14. S. Nishida and M. Shinya, Use of Image-based Information in Judgments of Surface Reflectance Properties, *J. Opt. Soc. A*, **15,** 2951-2965, 1998.
15. R. O. Dror, E. H. Adelson, and A. S. Willsky, Estimating Surface Reflectance Properties from Images under Unknown Illumination , *Proceedings of SPIE vol. 429: Human Vision and Electronic Imaging VI*, in press, 2001.
16. Available on http://www.debevec.org.
17. E. H. Adelson and A. P, Pentland, The Perception of Shading and Reflectance, in D. Knill and W. Richards (eds.), *Perception as Bayesian Inference*, pp. 409-423, New York: Cambridge University Press, 1996.