CVA Special Issue: Surface Detail in Computer Models

Kristin J. Dana Oana G. Cula Jing Wang

Department of Electrical and Computer Engineering
Department of Computer Science
Rutgers University
94 Brett Road,
Piscataway, NJ 08854

1 Introduction

Traditionally in vision, the reflectance of light at a surface has been modelled with rather simple linear combination of Lambertian and specular models. Such models are applicable for man-made scenes and where the surfaces are either "matte" or shiny, and rather uniform. An office scene with a desk, chalk, pen, stapler may be modeled well with these simple models. But real world scenes have a far more complex interaction of light and surfaces. Consider Figure 1 which shows examples of real-world objects and their fine scale surface appearance. As high resolution imagery becomes more prevalent, these fine scale details are relevant. Note that these details are not simply painted patterns on a surface, rather the material properties and the fine scale geometry cause a complex interaction of light at the surface.This means that it's not trivial to predict the appearance of the surface when the light source and camera moves. So why are these details important? In computer graphics, we want to render realistically so computationally modelling the appearance of surface under variations of viewing and illumination direction is required. In computer vision, surface details provide an important cue for recognition. Using these fine scale details to assist recognition is

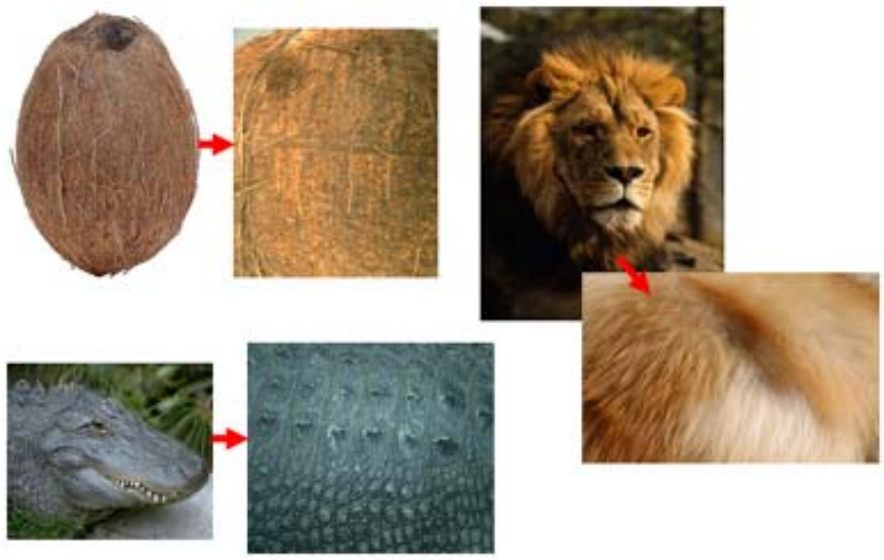

Fig. 1. Examples of real-world surfaces. Notice the fine scale detail in the closer views. This detail is not readily captured in traditional models of appearance in vision and graphics. However, these details are important for both rendering, recognition and general computational characterization.

largely an open problem with only a few researchers exploring this approach.

One of the more interesting surfaces for industry applications and that is human skin. Certainly in graphics rendering humans is of great interest to the entertainment industry for movie making, to the cosmetic industry for appearance prediction, for multimedia industry to enhance virtual reality systems. While great strides have been made in graphics, skin texture that looks as realistic as a photo is still an open problem. For recognition, human identification is a problem of growing importance especially in security related applications. While the literature is filled with papers on face recognition and face identification, an interesting spin is to measure and model facial detail. Fast, high quality, high resolution digital cameras allow us to consider imaging at a resolution that was not common before. A high resolution skin image is filled with texture, pores (as in Figure 2), as well as wrinkles, freckles, moles and other surface irregularities. We are interested

Fig. 2. Skin texture at increasing resolution. The nonuniform spatial detail is increasingly apparent at higher resolution.

in modeling and measuring these components to understand and predict their appearance under various imaging conditions. If we consider the standard problem of face recognition. The resolution that is usually considered is such that the overall face structure is the only cue for appearance. When the faces are diverse in appearance that approach is fine. However, consider the case of similar faces, that are hard to discern visually. Fine scale details can provide an additional cue which constrains the identification.

In this paper we discuss recent work in measuring and modeling surface detail. For modeling, we consider two approaches. The first approach is completely image-based and no geometry is modeled explicitly. In general, image-based techniques are very useful in modeling complex surfaces because fine-scale geometry is difficult to measure (there are limits to the resolution of laser range finders and these devices only work well with certain material types). The second modeling approach combines image-based modeling with inference of geometric textons. In modeling, we present measurement methods for capturing surface detail. Our

goal in this discussion is to establish a unifying framework for our recent methods and illustrate how these methods can improve computational modeling for industrial applications.

2 Modeling Surface Detail

Traditional computer graphics and computer vision represented texture with a single image. This representation is a considerable oversimplification of the complex interaction of light with real surfaces. An important consideration is that texture or surface appearance changes drastically when the viewing and illumination direction is changes. The reason for this change is not only the surface reflectance function but also the fine scale geometry or roughness that causes occlusions, shadows and foreshortening that depends on view/illumination. Work in the literature that addresses texture change with illumination and viewing direction has grown tremendously in recent years. Some early work includes [16,18,28,25,8,10,9,16,18] and more recent work includes [3,2,19,24,4,27,24,26,17,29,31,30,5,7].

Terminology for texture that depends on imaging parameters was introduced in [11,12]. Specifically, the term bidirectional texture function (BTF) is used to describe image texture as a function of the four imaging angles (viewing and light directions). The BTF can be interpreted as a spatially varying bidirectional reflectance distribution function (BRDF). The BRDF is defined as the the radiance reflected from a scene point divided by the irradiance and can be written as $r(\theta_i, \phi_i, \theta_v, \phi_v)$ where θ_i, ϕ_i are the polar and azimuthal angles of the illumination direction, respectively, and θ_v, ϕ_v are the polar and azimuthal angles of the viewing direction. The dependence of r on two directions is the reason this reflectance function is bidirectional. Models of the BRDF can be quite complex in order to capture the large variation of surfaces in real world scenes. But for a textured surface, reflectance also varies spatially (see Figure 3). To capture this additional dimen-

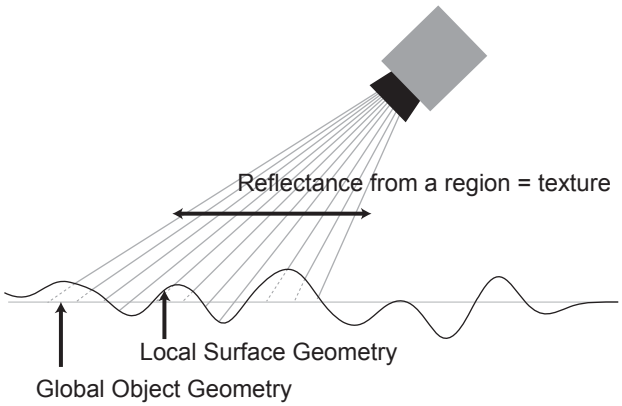

Fig. 3. Surface appearance or texture is the reflected light in a spatial region. We are interested in the case when the local geometry is not smooth and has some roughness. In general this fine-scale geometry is difficult to measure of model so image-based modeling techniques are useful.

sion, accurate models are needed for the BTF (bidirectional texture function) which can be written as $r(x, y, \theta_i, \phi_i, \theta_v, \phi_v)$ where x, y are Cartesian coordinates on the surface.

Surface detail is difficult to model directly because the local surface geometry is unknown. Also, even if measurement of the complex fine scale geometry is accomplished, a reflectance model applied to that geometry in a point-wise manner is likely to be too simplistic. For real world surfaces, light incident at a surface point is not generally reflected only from that point. Instead interreflections, partial transmission and scattering cause reflections from multiple surface points for each illuminated surface point as illustrated in Figure 4. Because of these modeling issues, it is popular to use image-based models which represent a surface by a collection of images that represent samples of the BTF. BTF databases that are made publicly available facilitate this modeling approach. Databases of BTF measurements are comprised of large sets of image for each class of surfaces. Examples of such databases are The CUReT (Columbia-Utrecht Reflectance and Texture) database [11] [12], the Yale texture measurements [17], and the Rutgers Skin Texture Database [7].

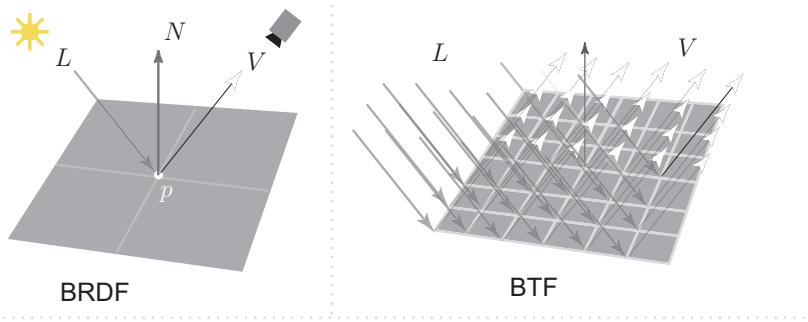

Fig. 4. A typical reflectance model predicts the amount of reflection from a particular point in all directions due to light incident on that point. However, light incident at a point may be interreflected and may be partially transmitted, scattered and then reflected. The result is the reflected light leaves *multiple* points. Using BTF-based models allows one to consider the surface patch-wise. We are concerned with the light exiting a patch due to the light incident on a patch. Therefore interreflections and subsurface scattering are implicitly included.

2.1 Image-based Models

One model for the BTF is the Bidirectional Feature Histogram [6]. Clearly, a statistical representation makes sense when one models texture for recognition purposes. The standard framework for texture representations consists of a primitive and a statistical distribution (histogram) of this primitive over space. So how does one account for changes with imaging parameters (view/illumination direction)? Either the primitive or the statistical distribution should be a function of the imaging parameters. Using this framework, the comparison of our approach with the 3D texton method[19] is straightforward. The 3D texton method uses a primitive that is a function of imaging parameters, while our method uses a statistical distribution that is a function of imaging parameters. In our approach the histogram of features representing the texture appearance is called bidirectional because it is a function of viewing and illumination directions. The advantage of our approach is that we don't have to align the images obtained under different imaging parameters.

The primitive used in our BTF model is obtained as follows. We start by taking a large set of surfaces, filter these surfaces by oriented multiscale filters and then cluster the output. The hy-

pothesis is that locally there are a finite number of intensity configurations so the filter outputs will form clusters (representing canonical structures like bumps, edges, grooves pits). The clustering of filter outputs are textons. A particular texture sample is processed using several images obtained under different imaging parameters (i..e. different light source directions and camera directions). The local structures are given a texton label from an image texton library (set up in preprocessing). For each image, the texton histograms are computed. Because these histograms are a function of two directions (light source and viewing direction), they're called bidirectional feature histograms or BFH. The recognition is done in two stages: (1) a training stage where A BFH is created for each class using example images and (2) a classification stage. In the classification stage we only need a single image and the light and camera direction is unknown and arbitrary. Therefore we can train with one set of imaging conditions but recognize under a completely different set of imaging conditions.

2.1.1 Image Textons

Within a texture image there are generic structures such as edges, bumps and ridges. Figure 5 illustrates the pre-processing step of constructing the image texton library. We use a multiresolution filter bank F, with size denoted by $3 \times f$, and consisting of oriented derivatives of Gaussian filters and center surround derivatives of Gaussian filters on three scales as in[19]. Each pixel of a texture image is characterized by a set of three multi-dimensional feature vectors obtained by concatenating the corresponding filter responses over scale. K-means clustering is used on these concatenated filter outputs to get image textons. By using a large set of images in creating the set of image textons, the resulting library is generic enough to represent the local features in novel texture images that were not used in creating the library.

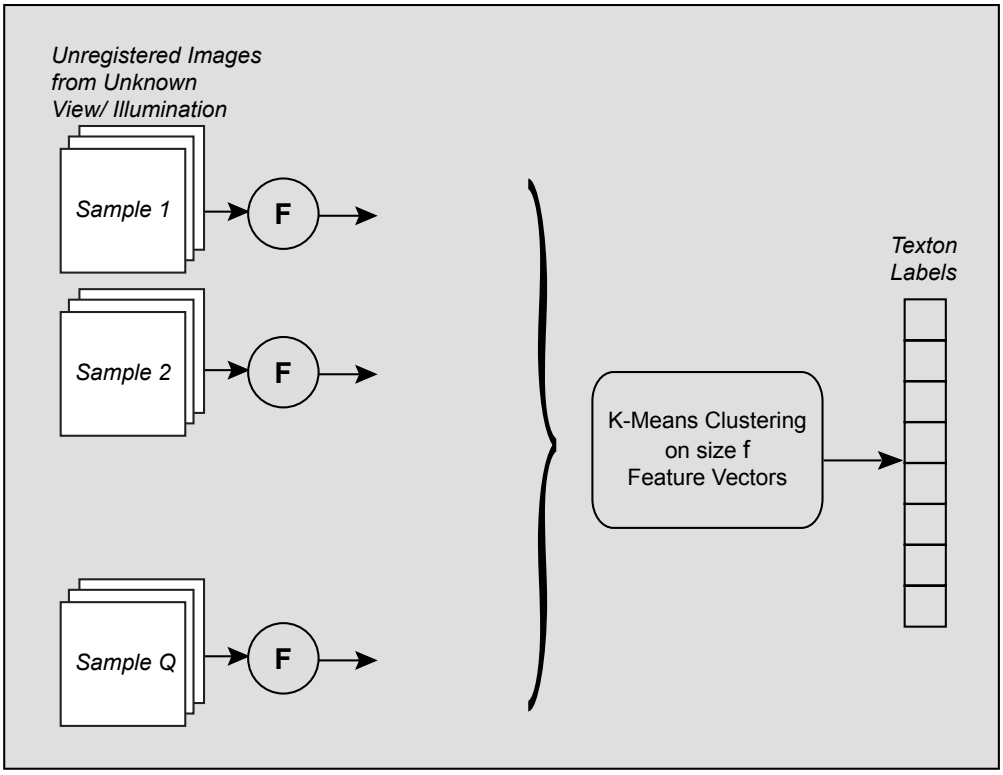

Fig. 5. Creation of the image texton library. The set of q unregistered texture images from the BTF of each of the Q samples are filtered with the filter bank F consisting of $3 \times f$ filters, i.e. f filters for each of the three scales. The filter responses for each pixel are concatenated over scale to form feature vectors of size f. The feature space is clustered via k-means to determine the collection of key features, i.e the image texton library.

2.1.2 Training Stage

The histogram of image textons is used to encode the global distribution of the local structural attribute over the texture image. This representation, denoted by $H(l)$, is a discrete function of the labels l induced by the image texton library, and it is computed as described in Figure 6. Each texture image is filtered using the same filter bank F as the one used for creating the texton library. Each pixel within the texture image is represented by a multidimensional feature vector obtained by concatenating the corresponding filter responses over scale. In the feature space populated by both the feature vectors and the image textons, each feature vector is labeled by determining the closest image texton. The spatial distribution of the representative local structural features over the image is approximated by computing the

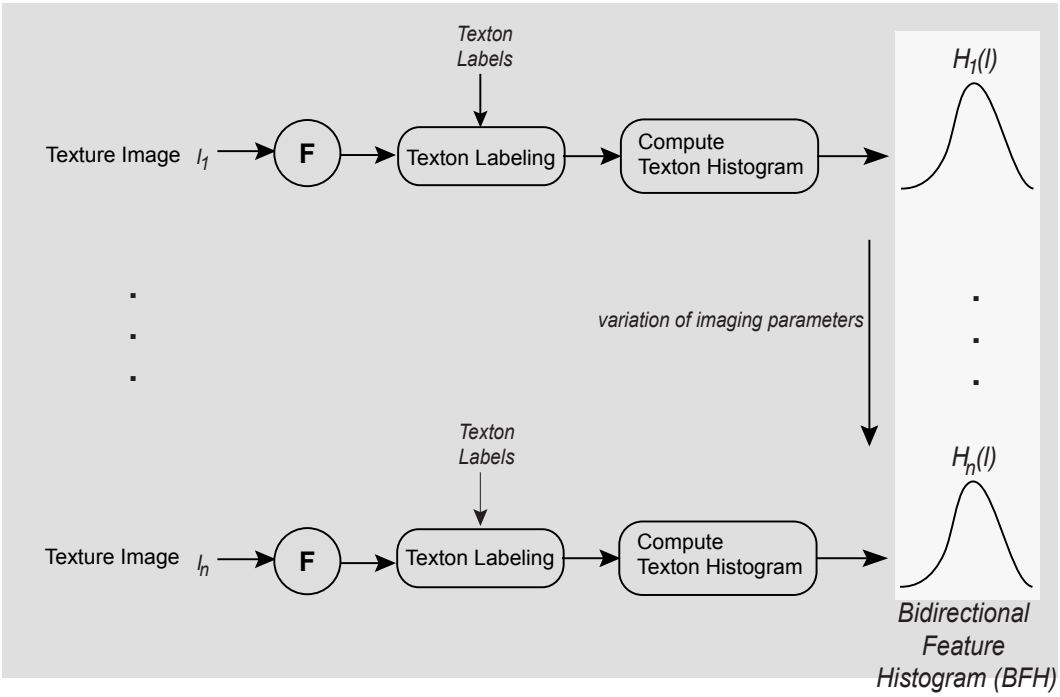

Fig. 6. 3D texture representation. Each texture image $I_j, j = 1 \dots n$, is filtered with filter bank F, and filter responses for each pixel are concatenated over scale to form feature vectors. The feature vectors are projected onto the space spanned by the elements of the image texton library, then labeled by determining the closest texton. The distributions of labels over the images are approximated by the texton histograms $H_j(l), j = 1 \dots n$. The set of texton histograms, as a function of the imaging parameters, forms the 3D texture representation, referred to as the bidirectional feature histogram (BFH).

texton histogram. Given the complex height variation of the 3D textured sample, the texture image is strongly influenced by both the viewing direction and the illumination direction under which the image is captured. Accordingly, the corresponding image texton histogram is a function of the imaging conditions.

Note that in our approach, neither the image texton nor the texton histogram encode the change in local appearance of texture with the imaging conditions. These quantities are local to a single texture image. We represent the surface using a collection of image texton histograms, acquired as a function of viewing and illumination directions. This surface representation is described by the term *bidirectional feature histogram*. It is worthwhile to explicitly note the difference between the bidirectional feature histogram and the BTF. While the BTF is the set of measured

images as a function of viewing and illumination, the bidirectional feature histogram is a representation of the BTF suitable for use in classification.

The dimensionality of histogram space is given by the number of textons in the image texton library. Therefore the histogram space is high dimensional, and a compression of this representation to a lower-dimensional one is suitable, providing that the statistical properties of the bidirectional feature histograms are still preserved. To accomplish dimensionality reduction we employ PCA, which finds an optimal new orthogonal basis in the space, while best describing the data. This approach has been inspired by [22], where a similar problem is treated, specifically an object is represented by set of images taken from various poses, and PCA is used to obtain a compact lower-dimensional representation.

2.1.3 Classification Stage

In the classification stage, the subset of testing texture images is disjoint from the subset used for training. Again, each image is filtered by F, the resulting feature vectors are projected in the image texton space and labeled according to the texton library. The texton distribution over the texture image is approximated by the texton histogram. The classification is based on a single novel texture image, and it is accomplished by projecting the corresponding texton histogram onto the universal eigenspace created during training, and by determining the closest point in the eigenspace. The 3D texture sample corresponding to the manifold onto which the closest point lies is reported as the surface class of the testing texture image.

2.1.4 Example Results

Detailed results can be found in [5]. Here we summarize an example of the results using the CUReT database[12], which is a

collection of BTF/BRDF measurements for 61 real-world textured surfaces, each imaged under more than 200 different combinations of viewing and illumination directions. In this experiment, we consider $Q = 20$ samples to construct the image texton library. For each sample we employ a set of $q = 64$ unregistered texture images corresponding to distinct imaging conditions. These images are uniformly sampled from the complete set of 205 texture images available within the CUReT database for each texture sample. Every texture image is filtered by filter bank F of size $3 \times f = 45$ as given in [5]. Color is not used as a cue for recognition because we are specifically studying the performance of texture models. Moreover, we consider that the use of color greatly assists the recognition since many of the surface samples in the CUReT dataset are well separated in color space. For computational reasons, the filter responses corresponding to a pixel in the image are grouped for each of the three scales to form three feature vectors. Three feature spaces are defined, each of dimension $f = 15$. In each of the feature spaces, k-means algorithm is used to determine the cluster centers. We varied the number of cluster centers K over the set of values $\{100, 150, 200, 250, 300, 350, 400, 450, 500, 550\}$, therefore we created ten image texton libraries for each scale.

For recognition, the same $P = 20$ 3D texture samples as the ones used for creating the image texton libraries. The subset of $p = 64$ texture images, for each sample, used for constructing the image texton library, are also employed during training. The set of texton histograms, corresponding to $P \times p = 1280$ texture images used for training, is employed to compute the universal eigenspace via PCA. In all our experiments, PCA is accomplished by employing the SLAM software library [23] [22]. We construct texture models by projecting all texton histograms onto the universal eigenspace, and by fitting a curve to the projected points corresponding to a certain sample using a quadratic b-spline interpolation. Therefore, each 3D texture sample is represented in the parametric eigenspace by a manifold indexed by the imaging

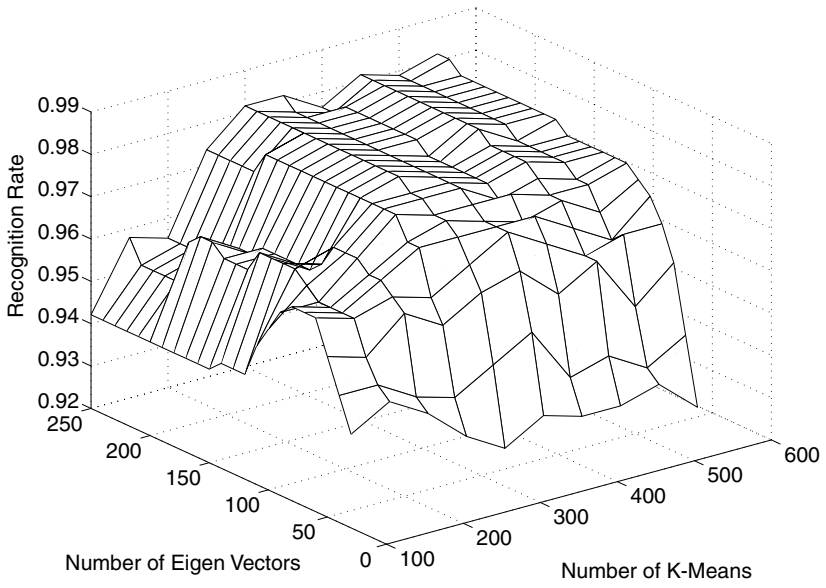

Fig. 7. Recognition rate, i.e. the percentage of correctly classified test images, as a function of the cardinality of the image texton library, as well as of the number of dimensions of the eigenspace.

parameters. We experiment with multiple values for the dimensionality e of the eigenspace, specifically we vary e from 30 to 250, with a step of 10. For classification we use 19 arbitrarily chosen texture images for each 3D textured surface, so the total number of test images is 380. Note that the sets of imaging conditions corresponding to the test and training images are disjoint.

The recognition rate, expressed as the percentage of correctly classified texture images, is analyzed in terms of its dependency on two factors: the size of the image texton library, and the number of eigenvectors. The plotted surface, illustrated in Figure 7, clearly suggests that undersampling the image texton library negatively affects the performance of the method. This behavior is due to grouping distinct local structural characteristics to be represented by the same image texton, affecting the capability of the representation to discriminate between different 3D textured surfaces. The number of eigenvectors also influences the performance, i.e. when the dimensionality of the eigenspace is reduced under a certain level the representation does not preserve the discriminative properties of the image texton histograms, and the performance decreases. However, when the size of the image tex-

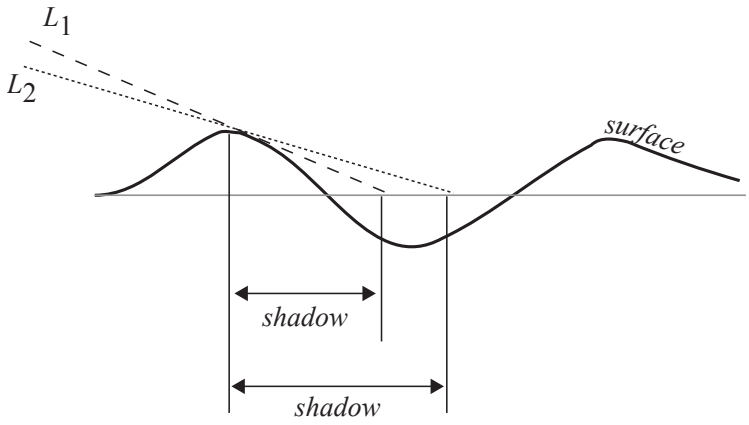

Fig. 8. An illustration of the creation of a shadow on a rough surface. The shadow length is determined by the illumination direction. As the illumination direction changes from L_1 to L_2 the shadow grows accordingly. The location, length, and continuous change of the shadow can be determined using geometric considerations. However, predicting cast shadows (that are not observed in the BTF sample images) is not possible with a purely image-based approach.

ton library is above 300 and the set of image texton histograms is projected onto an eigenspace of dimensionality greater than 70, the performance is excellent. Specifically, for this experiment, the recognition rate is well above 98%.

2.2 Hybrid Models: Image-based + Geometry

While image-based approaches are suitable for recognition tasks, there is a fundamental problem when using these methods for rendering. Only a sample of the images are measured, so image interpolation is required to render a smooth variation of imaging parameters. However, image interpolation cannot properly interpolate geometric effects. For example as the light source moves over a rough surface, the shadow lengths change continuously. Because shadows are nonlinear, i.e. no linear interpolation of shadow images can correctly predict the continuum of images between the interpolation endpoints. Similarly, the effect of occlusions as the viewing direction changes cannot be captured correctly with interpolation.

Figure 8 shows the cross section of a geometric structure and the

cast shadow that appears for the illustrated illumination. Consider that as the illumination changes continuously, the shadow shape changes continuously. Movement of the shadow borders is an important perceptual cue that provides and impression of the 3D nature of the geometric structure, and therefore improves the appearance of the actual surface texture.

The inaccuracies in the predicted texture reduce the realism in rendering. If we consider the continuous change of illumination direction, these inaccuracies become large visual artifacts. Figure 8shows the cross section of a geometric structure and the cast shadow that appears for the illustrated illumination. Consider that as the illumination changes continuously, the shadow shape changes continuously. Movement of the shadow borders is an important perceptual cue that provides and impression of the 3D nature of the geometric structure, and therefore improves the appearance of the actual surface texture. In this paper we present a new model of surface texture that incorporates both reflectance (images) and limited but useful geometric information. We recognize that while reflectance alone is not sufficient for modeling texture, a full geometry based approach is neither practical (it's hard to measure fine-scale geometry) nor sufficient (even if you have geometry you need the texture and shading for accurate modeling). Our model is a hybrid of geometry and reflectance models and has key advantages in a wide range of applications. Notably, for appearance prediction, the representation that can handle the non-linear effects of cast shadows and occlusions. While we have motivated the problem with rendering, the representation has significant implications in the areas of texture classification, point correspondences, as well as texture synthesis.

2.2.1 Geometric Textons

Our method assumes that the local geometry in a surface texture consists of a finite number of geometric configurations called *geometric textons*. Furthermore, we assume that these geometric

textons can be estimated using image observations. We are given N images of a surface texture corresponding to N different combinations of L, V. This image set is the sampled BTF. Standard stereo is not used in order to avoid the task of point correspondences. Also, we don't use photometric stereo because the local surface normal is not sufficient for cast shadows. Instead we estimate the local geometric structure and absolute height using a finite library of geometric primitives. This approach is motivated by the surface structure comprising a fine-scale geometric height variation that has limited range and exhibits some degree of spatial invariance. That is, local structure repeats and we expect the local structure to be well characterized by a finite set of predefined primitives. Thus our method is fundamentally different from [33], which explicitly recovers dense shape using shadow graphs, shape-from-shadow and shape-from-shading methods. The basic approach is as follows. We begin by learning a library of geometric textons. A key property of the geometric textons is that they contain information about absolute height in addition to local surface normals in a small neighborhood (e.g. 7×7). While the overall surface structure can be obtained by the surface normals, absolute height is required to recover cast shadows and occlusions. The next step is to label observations of a novel sample by using the sparsely sampled BTF to estimate the local geometric primitive at each pixel. The estimation of the local primitive is done in a two stage approach. First, primitives with surface normal information but no absolute height information are estimated. We use the term normalized geometric textons when referring to these geometric primitives. Once the normalized geometric texton has been identified, the absolute height is obtained by a comparison to the sampled BTF. After the map of labeled surface points is established, we reconstruct an estimate of the local fine-scale geometry. While this reconstruction is coarse, the key geometric structures which contribute to cast shadows and occlusions are recovered.

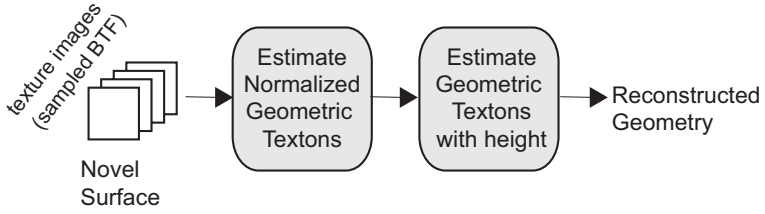

Fig. 9. Estimation of the local geometric texton. The sampled BTF of a novel surface is given. Then, the normalized geometric texton is estimated. Finally, the family of geometric textons with the same surface normal configuration but different heights is then considered and the closest match is used as the estimated geometric texton. To reconstruct geometry overlapping geometric textons are averaged.

2.2.2 Library of Geometric Textons

We assume that a finite library geometric textons can be used to represent the local fine scale geometry. For example these primitives correspond to height edges, ridges, etc. We also assume that the local albedo is constant. We define this library of geometric textons using a set of training samples with known geometry and k-means clustering. As in appearance-based texton methods such as [19][2,6], we use training images to obtain the library, however here we cluster on geometry information instead of intensity information. The training surfaces are a set of surfaces that exhibit sufficiently varied surface structure so that a library can be created with these examples. The training surfaces are not the surfaces of interest. Unlike the training images, the novel surface shown as the input in Figure 9 has no prior geometric information. In fact, for our results we use synthetic rough surfaces for training, generated by simulating two dimensional gaussian random fields that has a gaussian covariance function. By varying the roughness and the effective correlation length, various input surfaces can be achieved. For clustering geometry, we consider a 7×7 patch of surface height values, where the center value is subtracted so that each patch as zero height in the center. In addition we consider the surface normal in the center of the patch. Therefore there is a 52 dimensional vector g at each surface point is given by

$$g = [h(1)\ h(2)\ ...\ h(49)\ n_x\ n_y\ n_z], \tag{1}$$

where h is the height vector for the 7×7 patch and the surface normal is given by n_x, n_y, n_z. In practice the surface normal must be scaled so that it contributes to the clustering result.

A novel surface is assigned geometric texton labels in the following manner. Assume that we have a sparsely sampled BTF f_s consisting of N images where

$$f_s(x, y, i) = f(x, y, L_i, V_i)\ i \in [1..N-1]. \tag{2}$$

We assume no prior geometry, e.g. from stereo, laser scanning, or other methods. The hypothesis is that the local reflectance distribution observed under several viewing and illumination directions is sufficient to infer the local geometric primitive or texton. Computationally the estimation of the local geometric texton relies on comparing the observed BTF with the prediction obtained by rendering the geometric texton. In order to render the geometric texton, we must choose a shading model and for simplicity Lambertian shading is chosen. However, any shading model, or multiple shading models, could be employed. Also, the surface need not strictly adhere to the shading model, as long as the best match to the observed BTF is the correct geometric texton. Let k denote the number of geometric textons. For each of the k geometric textons and for each of the N imaging parameters L_i, V_i, we render a 7×7 image. The center pixels of these images form a vector v of length N for each of the k geometric textons. At each pixel x, y, we have an N dimensional vector u from the sampled BTF $f_s(x, y)$. The correct geometric texton label l at that pixel is given by

$$l = \mathbf{argmax}_j\ (v_j \cdot u)\ j \in [1..k]. \tag{3}$$

At this point, the geometric textons considered have zero height at the center. These textons are the normalized geometric textons. To get absolute height instead of relative height, we consider linear transformations of the height vector h encoded within the primitive g that best match f_s. To reconstruct the geometry we

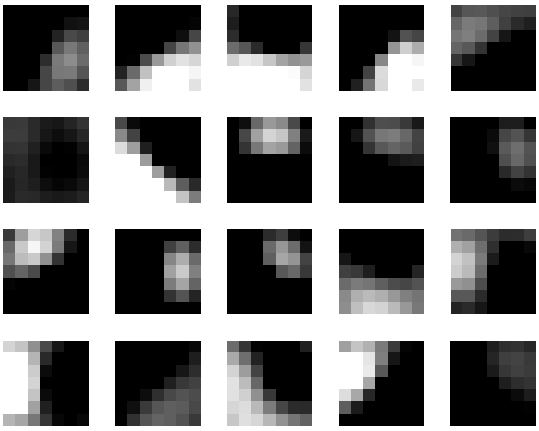

Fig. 10. An illustration of the normalized geometric textons with $k = 20$.

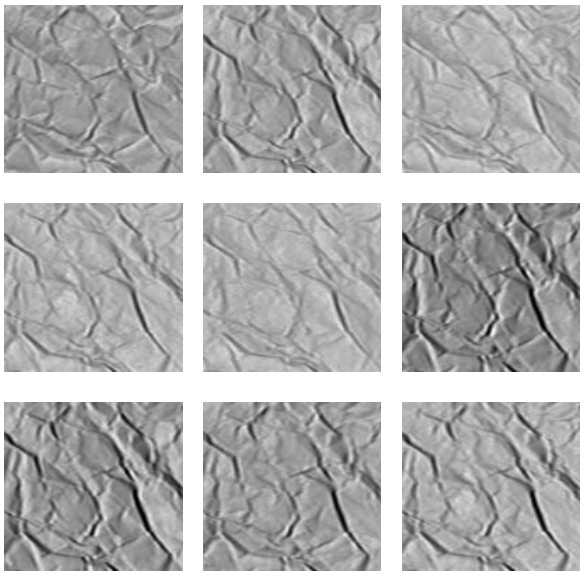

Fig. 11. BTF samples for crumpled paper used in generating the result in Figure 12. For these results, $N = 9$ which means f_s consists of 9 texture images.

put the 7×7 region of height information in the correct position and average the overlapping regions.

2.2.3 Combining Reflectance and Geometry

Consider the goal of generating the full BTF $f(x, y, L, V)$ from the sampled BTF f_s. Using only reflectance information in the form of images, some type of interpolation is needed. For example, if eigenspace methods are used, then a linear combination of the basis images is required to give us $f(x, y, L, V)$. Alternatively, if 3D textons are used as in [19][20], then the appearance vectors

associated with the 3D textons must be interpolated to get the correct texture image. For an arbitrary L, V, let I_r denote the texture image obtained using only reflectance information f_s. Let I_g be the image obtained by rendering the reconstructed geometry. I_r is a good prediction except where there are cast shadows. The location of the cast shadows can be predicted using I_g. Specifically, a binary mask M is created to indicate the location of the cast shadows. In practice, the mask is blurred to remove abrupt intensity changes. The two images I_r and I_g are blended to get the result I given by

$$I = I_r M + I_g(1 - M). \tag{4}$$

This equation explicitly shows that the representation is a hybrid of reflectance and geometric methods and motivates the name hybrid texton method. Note that because I_g is primarily used for cast shadow information, the exact shading model used in the intermediate rendering step is not critical.

2.2.4 Results

To illustrate how this representation can be used, we show the result of predicting texture appearance. Given a sampled BTF f_s, we wish to predict the image using imaging parameters L, V, where L, V is not in the set of N imaging parameters L_i, V_i. For our results, the number of classes $k = 20$, and the number of BTF sample image $N = 9$. The height data from the k geometric textons is illustrated in Figure 10. Notice these local primitives correspond to basic structural elements. We use texture images in the CUReT database [12] from Sample28 (crumpled paper) and Sample11(rough plaster). An example of the 9 BTF images used in the estimation is shown in Figure 11. The corresponding prediction result is shown in Figure 12. In this result, the frontal view image with $L = 67.4$ is not included in the sampled BTF. It is shown as ground truth to evaluate the quality of the rendered result. The BTF images used for interpolation to get I_r correspond to $L = 22.5°$ and $L = 45°$. A simple inter-

polation is sufficient to match the intensities everywhere except the cast shadow region. Since no similar cast shadows exist in the 9 BTF images, the cast shadows cannot be predicted by any other linear combination of these images. The figures show I_r, I_g and the final result. By comparing the rightmost images (top and bottom row) it is clear that the final result predicts the overall appearance including cast shadows.

In summary, this hybrid texton method explicitly estimates concise geometric information in the form of geometric textons that encode local height distribution. This approach allows prediction of the key property of appearance that can only be obtained using geometry, namely cast shadows. While cast shadows may be cursory information in some scenes, for rough surfaces and other 3D texture, these cast shadows are abundant and important. Our approach can also be extended to predict occlusion areas and since the problem of cast shadows and occlusions are essentially the same geometric problem.

3 Measuring Surface Detail

While BTF measurements are useful in texture research, measurement methods are time consuming because the light source and camera must be moved in a hemisphere of possible directions. Our approach to measuring surface detail uses a curved mirror to create a convenient imaging device where multiple views of the same surface point are realized simultaneously. The device has the significant advantage that no parts need to be moved in a hemisphere of directions. Instead, only simple planar motions of the imaging components are needed. Figure 13 illustrates the basic idea. A concave parabolic mirror focuses light to a single point. Therefore, this mirror can be used for convenient orientation of the illumination direction. An incident ray reflecting off the mirror will reach the surface at an angle determined by the point of intersection with the mirror. The light reflected from

the surface point at a large range of angles is also reflected by the mirror and can be imaged by a camera. This mirror is the main component in the imaging device and the complete device enables fast multiview or bidirectional surface imaging at high spatial resolution. We summarized the device here, and the full details of the design are presented in [13].

The imaging device or *texture camera* uses optical components such as a beam-splitter, concave parabolic mirror, CCD camera, translation stages and is illustrated in Figure 14. The beam-splitter allows simultaneous control of the illumination and viewing direction. A concave parabolic mirror section is positioned so that its focus is coincident with the surface point to be measured. The illumination source is a collimated beam of light parallel to the global plane of the surface and passing through a movable aperture. The aperture ensures that only a spot of the concave mirror is illuminated and therefore one illumination direction is measured for each aperture position. In this approach, the problem of changing the illumination direction over a hemisphere is transformed to the easier problem of translating an aperture in a plane. The light reflected at each angle is reflected from the mirror to a parallel direction and diverted by the beam-splitter to the camera. The camera is equipped with an orthographic or telecentric lens that images the light parallel to the optical axis. The image of the mirror is viewed by the camera that is positioned so that its optical axis lies along the Y axis so that a single image corresponds to reflectance measurements from all angles in a partial hemisphere. To obtain a measurement of a surface patch for spatially varying BRDF, the concave mirror is translated along the X-Y plane. This arrangement also has the advantage that all light from the measurement point will reflect away from the sample and thus will not reilluminate the surface point changing the intended illumination pattern.

The approach of using curved mirrors was used by [32] who introduced a method of measuring BRDF using a half-silvered spherical mirror and a fish-eye lens that enables simultaneous measure-

ment of light from all viewing directions without changing camera position. This device is not ideal for BRDF/BTF measurements for a few reasons. First, there is no means of automatically changing the illumination direction over the hemisphere. Second, it is designed for measuring reflectance from a single point and extended samples could cause reflection of light to the mirror and back onto the surface, changing the illumination pattern. Also, the fish eye lens is an imaging component that introduces significant distortion.

Another BRDF measurement device described in [14], [21] uses a hemi-elliptical mirror in a hand-held device designed for industrial coating evaluation. Our device differs most significantly from this device in its method of illumination angle control. Specifically, in that device the illumination direction is changed using an additional gimbal mirror which is cumbersome to control automatically. Indeed the commercial version of this device only enables illumination angle variation in a plane instead of a partial hemisphere. Angular variations of a gimbal mirror are more difficult than translational motion of an aperture. Also, when scanning a surface area for BTF measurements, the design in [14], [21] requires translation of both mirrors and sensors. In our design, translating a single mirror is sufficient for scanning a small surface region.

Another device employing concave mirrors for BRDF measurement is described in [1]. This device uses two mirrors to achieve a similar functionality to the device described here. The device is consequently more complex and more difficult to prototype. Also, scanning a surface requires traversal of two mirrors instead of one. The design has two focal points and the distance between the two focal points limits the size of the object to be measured.

A "kaleidoscope" approach to BTF measurement is given in [15]. This device measures texture appearance for a discrete number of imaging parameters since the mirrors are not curved. The advantage is that scanning is not necessary for acquiring the ap-

pearance of a region. The disadvantage of this device is that only a few discrete samples of the full imaging parameters space can be explored.

The prototype device has an off axis parabolic mirror (Janos Technology A8037-175) that is a section of a full parabolic mirror. Figure 16 shows a closer view of this parabolic mirror and a test sample (blue glossy cardboard).

Consider the dimensions of the sample that can be measured with this device. We move the sample when scanning, so the width and height of the sample are limited only by the travel range of the stage. In the current prototype this range is approximately 25 mm in both dimensions. The fine-scale depth variation limit is determined empirically to be approximately 1-2 mm. Beyond this depth variation, the surface point is no longer in the mirror focus.

Additional equipment includes a beam splitter (K54-823 Edmund Scientific), fiber optic illuminator (Schott Fostec, DCR II), and an iris diaphragm with 0.8 mm minimum aperture and 25.4 mm maximum aperture (K53-915 Edmund Scientific). Figure 15 shows an image of the main components of the current system prototype. Not shown in the figure are the X-Z and X-Y-Z mechanical stages (Velmex Inc.). These stages are used for automatic scanning by placing the illumination aperture on an X-Z stage and the mirror on an X-Y-Z stage. The light source is DC-regulated and equipped with a quartz halogen bulb. Light collimation is implemented with a convex lens series so that the incident illumination is sufficiently bright and nearly parallel. Note that several variations of the illumination source can be made within the context of the original design. For instance, spectral filters on the illumination source would enable BRDF measurements as a function of spectral wavelength. In addition, polarizers may be used at both the source and detector for measurements as a function of polarization angle. Also, scanning a non-planar sample will require an additional system component to estimate the depth of the object. While this prototype is mounted on an op-

tical board for convenience of testing and modification, portable implementations are foreseeable extensions.

3.0.5 Imaging Results

To obtain a texture image, the desired viewing direction and illumination direction are specified. The illumination aperture is positioned to achieve the correct illumination direction. A single pixel from each camera image is identified that corresponds to the correct viewing direction. This pixel is acquired for each mirror position along a 2D grid in the $X - Y$ plane.

To illustrate the image texture acquired by the texture camera, we show representative examples from various surface samples (Figures 17 and 18). For each of these examples the spatial sampling is dense with an effective pixel size of 0.075 mm. The images in Figure 17 correspond to a fixed viewing direction and three different illumination directions. Notice that the illumination change affects the position of shadows, specularities and other features of appearance. The images in Figure 18 correspond to a fixed illumination direction and three different viewing directions. These images also reveal the changes in the overall appearance of each surface sample with viewing direction.

4 Discussion

The methods discussed in this paper comprise a modeling and measurement framework for computational surface detail. There are a wide variety of specific industrial applications for these methods. For example, the BRDF/BTF of textiles and coatings can be measured for quality control and inventory characterization. Measurements of BRDF/BTF of materials can be used in design and planning, e.g. in automotive interior design. Instead of expensive mock-up for modeling and evaluation, rendering realistic interiors would speed the design process while lowering costs.

Similarly, architectual interior design can use BRDF/BTF measurements and models for visualization purposes. E-commerce would be greatly enhances if the consumer could view the appearance of fabrics under a variety of pose and illuminations. Evaluations of human skin texture can be done in a quantitative manner using these methods. Such quantification would help the consumer product industry in validating marketing claims. Also the pharmaceutical industry can utilize quantitative evaluation of skin texture during drug development in dermatology. Computational representations of surface detail enable a true digital representation of the object that captures all the essential observable features of an object. As such, these models are fundamentally important for industries adapting to an emerging digital society.

5 Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. 0092491 and Grant No. 0085864.

References

[1] R. R. Carter and L. K. Pleskot. Imaging scatterometer. *US Patent 5912741*, June 1999.

[2] O. G. Cula and K. J. Dana. Compact representation of bidirectional texture functions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1:1041–1067, December 2001.

[3] O. G. Cula and K. J. Dana. Recognition methods for 3d textured surfaces. *Proceedings of SPIE Conference on Human Vision and Electronic Imaging VI*, 4299:209–220, January 2001.

[4] O. G. Cula and K. J. Dana. Image-based skin analysis. *Proceedings of Texture 2002 - The 2nd international workshop on texture analysis and synthesis*, pages 35–40, June 2002.

[5] O. G. Cula and K. J. Dana. 3D texture recognition using bidirectional feature histograms. *International Journal of Computer Vision*, 59(1):33–60, August 2004.

[6] O. G. Cula and K. J. Dana. 3D texture recognition using bidirectional feature histograms. *to appear in International Journal of Computer Vision*, 59(1):33–60, August 2004.

[7] O. G. Cula, K. J. Dana, F. P. Murphy, and B. K. Rao. Skin texture modeling. *International Journal of Computer Vision*, 62(1/2):97–119, April/May 2005.

[8] K. J. Dana and S. K. Nayar. Histogram model for 3d textures. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 618–624, June 1998.

[9] K. J. Dana and S. K. Nayar. 3d textured surface modeling. *IEEE Workshop on the Integration of Appearance and Geometric Methods in Object Recognition*, pages 46–56, June 1999.

[10] K. J. Dana and S. K. Nayar. Correlation model for 3d texture. *International Conference on Computer Vision*, pages 1061–1067, September 1999.

[11] K. J. Dana, B. van Ginneken, S. K. Nayar, and J. J. Koenderink. Reflectance and texture of real world surfaces. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 151–157, June 1997.

[12] K. J. Dana, B. van Ginneken, S. K. Nayar, and J. J. Koenderink. Reflectance and texture of real world surfaces. *ACM Transactions on Graphics*, 18(1):1–34, January 1999.

[13] K.J. Dana and Jing Wang. Device for convenient measurement of spatially varying bidirectional reflectance. *Journal of the Optical Society of America A*, 21:pp. 1–12, January 2004.

[14] K. J. Davis and Diane C. Rawlings. Directional reflectometer for measuring optical bidirectional reflectance. *US Patent 5637873*, June 1997.

[15] J.Y. Han and K. Perlin. Measuring bidirectional texture reflectance with a kaliedoscope. *ACM Transactions on Graphics*, 22(3):pp. 741–748, July 2003.

[16] J. J. Koenderink, A. J. van Doorn, K. J. Dana, and S. K. Nayar. Bidirectional reflection distribution function of thoroughly pitted surfaces. *International Journal of Computer Vision*, 31(2-3):pp. 129–144, 1999.

[17] M. L. Koudelka, S. Magda, P. N. Belhumeur, and D. J. Kriegman. Acquisition, compression and synthesis of bidirectional texture functions. *3rd International Workshop on Texture Analysis and Synthesis (Texture 2003)*, pages 59–64, 2003.

[18] T. Leung and J. Malik. Recognizing surfaces using three-dimensional textons. *International Conference on Computer Vision*, 2:1010–1017, 1999.

[19] T. Leung and J. Malik. Representing and recognizing the visual appearance of materials using three-dimensional textons. *International Journal of Computer Vision*, 43(1):29–44, 2001.

[20] X. Liu, Y. Yu, and H.Y. Shum. Synthesizing bidirectional texture functions for real world surfaces. *ACM SIGGRAPH*, pages 97–106, 2001.

[21] P.R. Mattison, M.S. Dombrowski, J. Lorenz, K. Davis, H. Mann, P. Johnson, and B. Foos. Hand-held directional reflectometer: an angular imaging device to measure brdf and hdr in real-time. *Proceedings of SPIE - the International Society for Optical Engineering, Scattering and Surface Roughness II*, 3426:240–251, July 1998.

[22] H. Murase and S. K. Nayar. Visual learning and recognition of 3-d objects from appearance. *International Journal of Computer Vision*, pages 5–24, 1995.

[23] S. A. Nene, S. K. Nayar, and H. Murase. Slam: A software library for appearance matching. *Technical Report CUCS-019-94*, Proceedings of ARPA Image Understanding Workshop, November 1994.

[24] A. Penirschke, M. Chantler, and Maria Petrou. Illuminant rotation invariant classification of 3D surface textures using lissajous's ellipses. *Proceedings of Texture 2002 - The 2nd international workshop on texture analysis and synthesis*, pages , pp. 103–108, 2002.

[25] P. Suen and G. Healey. Analyzing the bidirectional texture function. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 753–758, June 1998.

[26] Frank Suykens, Karl Berge, Ares Lagae, and Philip Dutre. Interactive rendering with bidirectional texture functions. *Computer Graphics Forum*, 22(3):463–463, 2003.

[27] X. Tong, J. Zhang, L.Liu, X.Wang, B. Guo, and H.Y. Shum. Synthesis of bidirectional texture functions on arbitrary surfaces. *ACM Transactions on Graphics*, 21(3):665–672, 2002.

[28] B. van Ginneken, M. Stavridi, and J. J. Koenderink. Diffuse and specular reflectance from rough surfaces. *Applied Optics*, 37:130–139, January 1998.

[29] Manik Varma and Andrew Zisserman. Texture classification: Are filter banks necessary? *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2:691–698, 2003.

[30] Manik Varma and Andrew Zisserman. Estimating illumination direction from textured images. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1:179–186, 2004.

[31] M. A. O. Vasilescu and Demetri Terzopoulos. Tensortextures. *ACM SIGGRAPH*, 2003.

[32] G. J. Ward. Measuring and modeling anisotropic reflection. *ACM SIGGRAPH*, 26(2):265–272, 1992.

[33] Yizhou Yu and Johnny T. Chang. Shadow graphs and surface reconstruction. *Proceedings of the European Conference on Computer Vision*, pages 31–45, May 2002.

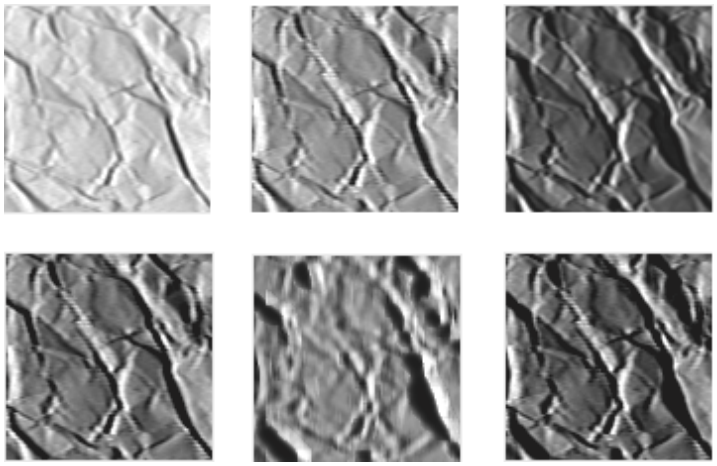

Fig. 12. Predicted texture appearance using hybrid texton method using crumpled paper. Top row: texture image with $L = 22.5°$ (left), $L = 45°$ (center), and $L = 67.5°$ (right). The viewing direction is frontal for each of these images. The sampled BTF f_s for this example did not contain the rightmost image with $L = 67.5°$. The goal in this result is to predict this texture image. Bottom row: I_r (left), I_g (center), predicted image $I(L, V)$ with $L = 67.5$. Compare the rightmost image of both rows to compare ground truth (top row) with the predicted image (bottom row).

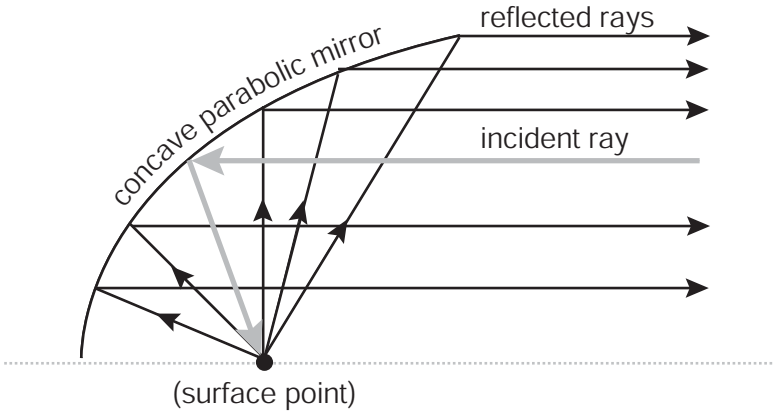

Fig. 13. The focusing property of a concave parabolic mirror is exploited to simultaneously measure reflected rays from a large range of angles over the hemisphere. The same mirror is used to direct the incident illumination ray to the sample at the desired angle.

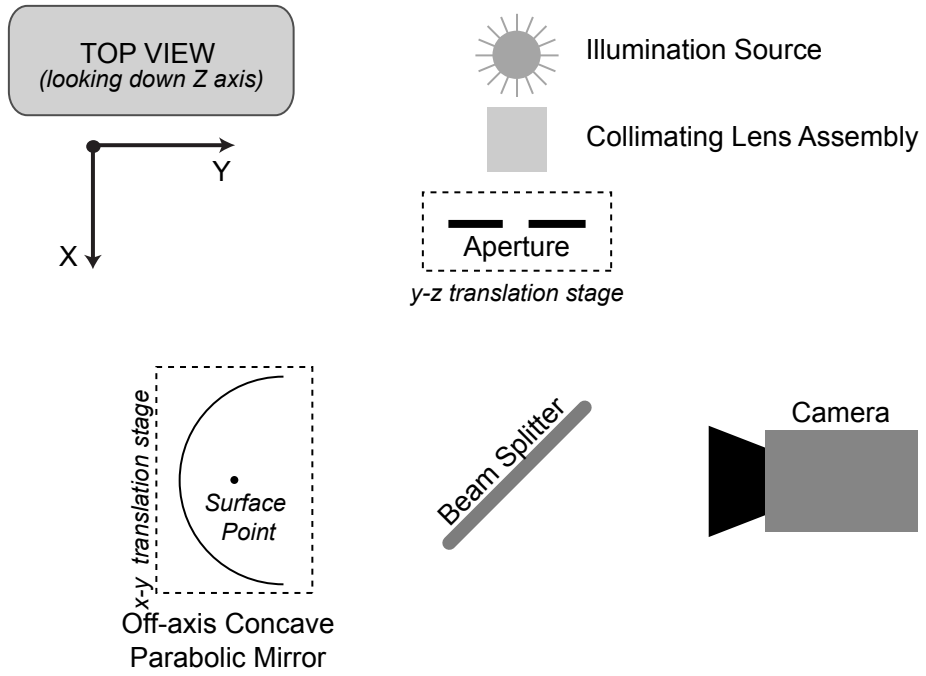

Fig. 14. BRDF/BTF Measurement Device. The surface point is imaged by a CCD video camera observing an off-axis concave parabolic mirror to achieve simultaneous observation of a large range of viewing directions. Illumination direction is controlled by an aperture, i.e. translations of the aperture in the X-Z plane cause variations in the illumination angle incident on the surface point. The device achieves illumination/viewing direction variations using simple translations of the illumination aperture instead of complex gonioreflectometer equipment. Measurements of bidirectional texture are accomplished by translating the mirror in the $X - Y$ plane.

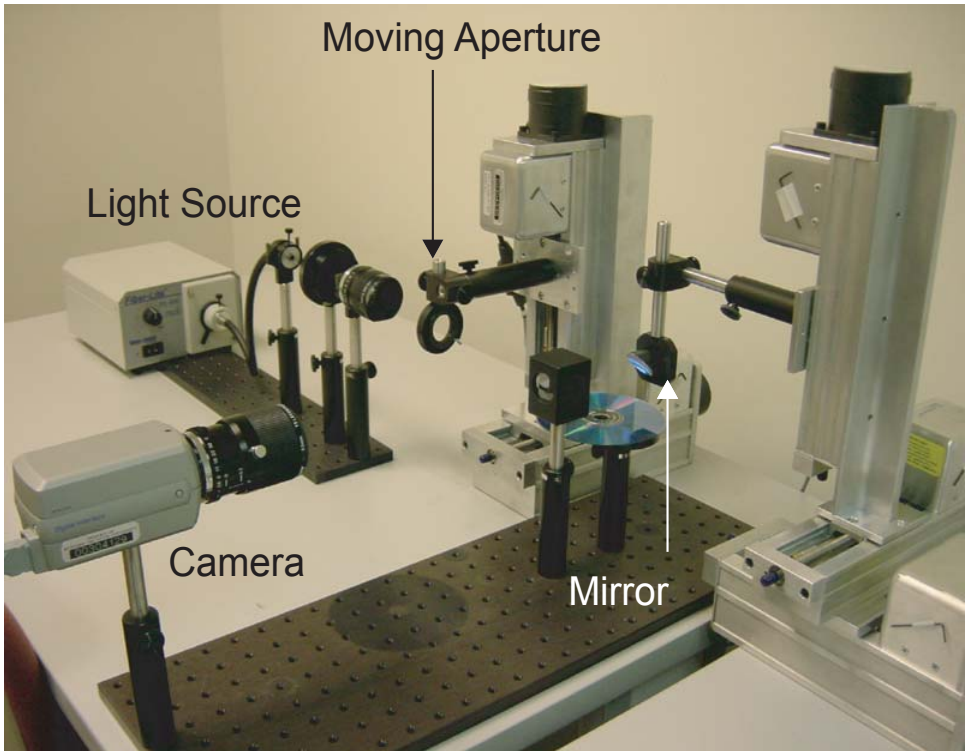

Fig. 15. Device prototype including camera, illumination source, collimating lens assembly, illumination aperture, beam splitter and off-axis concave parabolic mirror. A CD is the imaged sample placed at the focus of the mirror.

Fig. 16. Off-axis concave parabolic mirror used in prototype.

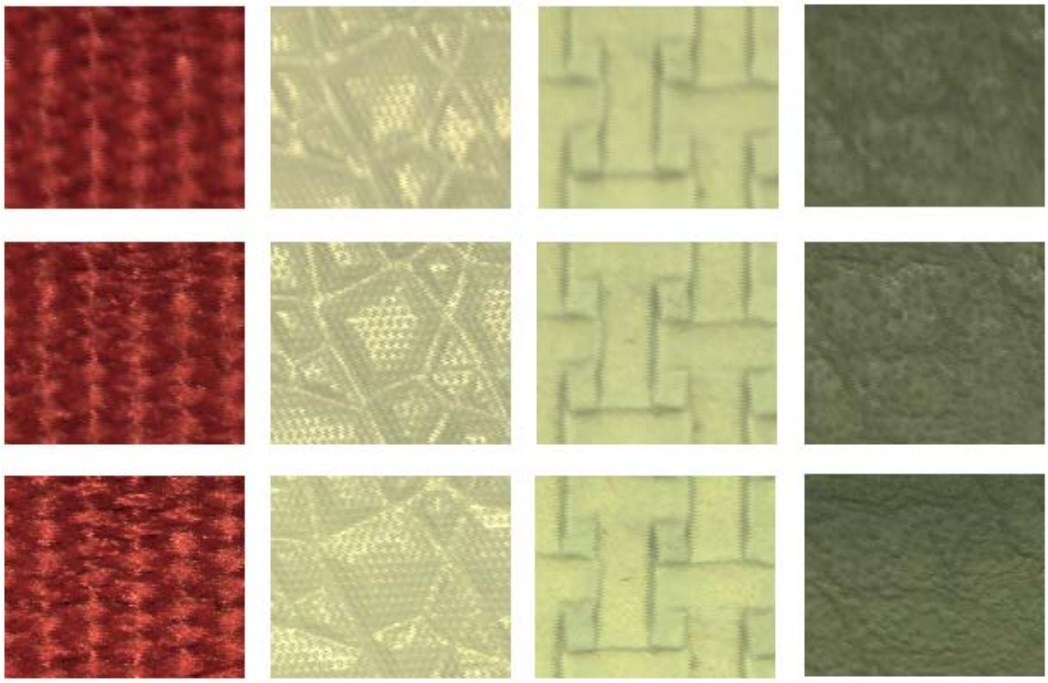

Fig. 17. Texture images for sample surfaces: (from left to right) canvas, rough plastic, rubber mat, and leather. Each image corresponds to a frontal view (i.e. $\theta_v = 0$). The width is 10.5 mm and the height is 8.25mm. The top row corresponds to $\theta_i = 12.1°$. For the middle row $\theta_i = 0$. For the last row $\theta_i = -15.3°$. For each image $\phi_i = 0$, i.e. the illuminating ray intersects the mirror in the $y - z$ plane.

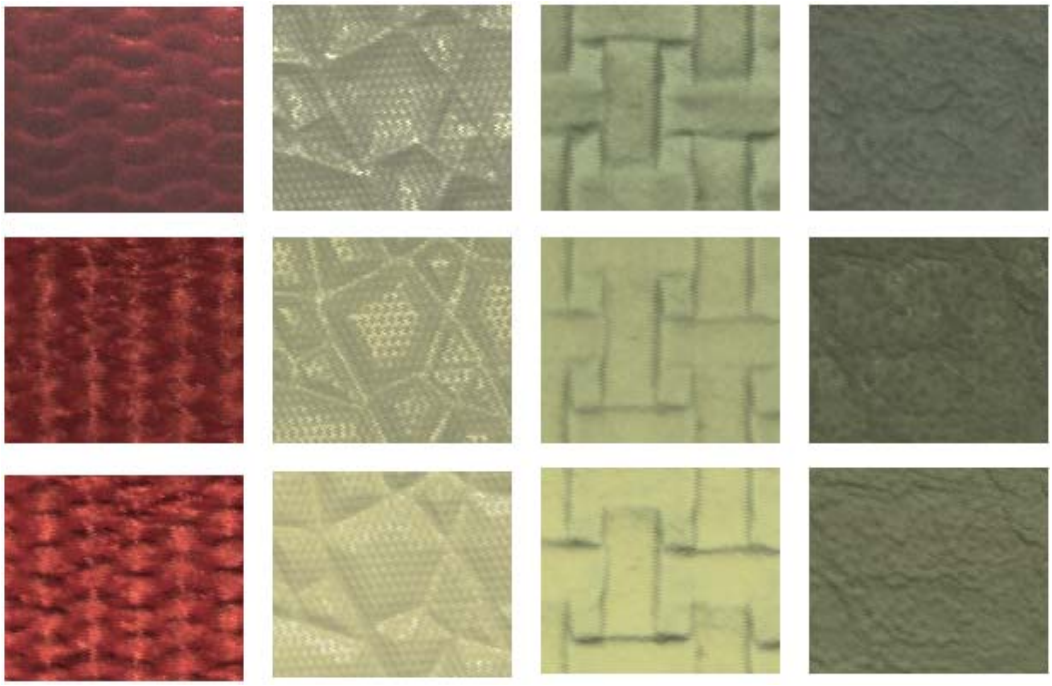

Fig. 18. Texture images for sample surfaces: (from left to right) canvas, rough plastic, rubber mat, and leather. Each image corresponds to a frontal illumination (i.e. $\theta_i = 0$). The width is 10.5 mm and the height is 8.25mm. The top row corresponds to $\theta_v = 22°$. For the middle row $\theta_v = 0$. For the last row $\theta_v = -22°$. For each image $\phi_v = 0$, i.e. the point of interest on the mirror is in the $y - z$ plane. Note that the brightness of the first row has been manually enhanced so that structure is better visible.