



## 3D Texture Recognition Using Bidirectional Feature Histograms

OANA G. CULA

*CS Department, Rutgers University, Piscataway, NJ 08854, USA*

*oanacula@caip.rutgers.edu*

KRISTIN J. DANA

*ECE Department, Rutgers University, Piscataway, NJ 08854, USA*

*kdana@caip.rutgers.edu*

*Received April 3, 2002; Revised May 14, 2003; Accepted August 14, 2003*

**Abstract.** Textured surfaces are an inherent constituent of the natural surroundings, therefore efficient real-world applications of computer vision algorithms require precise surface descriptors. Often textured surfaces present not only variations of color or reflectance, but also local height variations. This type of surface is referred to as a 3D texture. As the lighting and viewing conditions are varied, effects such as shadowing, foreshortening and occlusions, give rise to significant changes in texture appearance. Accounting for the variation of texture appearance due to changes in imaging parameters is a key issue in developing accurate 3D texture models. The bidirectional texture function (BTF) is observed image texture as a function of viewing and illumination directions. In this work, we construct a BTF-based surface model which captures the variation of the underlying statistical distribution of local structural image features, as the viewing and illumination conditions are changed. This 3D texture representation is called the bidirectional feature histogram (BFH). Based on the BFH, we design a 3D texture recognition method which employs the BFH as the surface model, and classifies surfaces based on a single novel texture image of unknown imaging parameters. Also, we develop a computational method for quantitatively evaluating the relative significance of texture images within the BTF. The performance of our methods is evaluated by employing over 6200 texture images corresponding to 40 real-world surface samples from the CURET (Columbia-Utrecht reflectance and texture) database. Our experiments produce excellent classification results, which validate the strong descriptive properties of the BFH as a 3D texture representation.

**Keywords:** texture, 3D texture, bidirectional texture function, BTF, bidirectional feature histogram, BFH, appearance, appearance-based, recognition, texon, image texon

### 1. Introduction

The vast information embedded in visual imagery is organized at many levels: scenes are comprised of objects, objects are constructed from surfaces, and surfaces are described by reflectance and texture. Within this schematic of the visual world, surface descriptions form the foundation of image understanding. Indeed accurate surface models support appearance prediction which is central to many computer vision algorithms.

Descriptions of textured surfaces have suffered from oversimplification in much of the literature over the past couple of decades in computer vision. Texture viewed as a single image is a poor representation of the intricate interaction of light with real surfaces. Increasingly, recent work on texture representations deal with the complex changes of surface appearance with changes in illumination and viewing direction (Chantler, 1995; Koenderink et al., 1999; Leung and Malik, 1999; van Ginneken et al., 1998; Suen and

Healey, 1998; Dana and Nayar, 1998, 1999a, 1999b; Suen and Healey, 2000; McGunnigle and Chantler, 2000; Cula and Dana, 2001a, 2001b, 2002; Liu et al., 2001; Tong et al., 2002; Cula et al., 2003). Terminology for texture that depends on imaging parameters was introduced in Dana et al. (1997, 1999). 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  $\theta_i, \phi_i$  are the polar and azimuthal angles of the illumination direction, respectively, and  $\theta_v, \phi_v$  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. To capture this additional dimension, 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.

To underscore the importance of accounting for variations with imaging parameters, consider Fig. 1 which shows the same surface under different viewing and illumination directions. Clearly the “image texture” changes dramatically in this image set, while the underlying surface texture remains the same. It is tempting to use geometric models to describe 3D textured surfaces. While the surface texture often exhibits fine scale geometry, the geometry is very difficult to model and sometimes impossible to measure. Even the most sophisticated laser range finders are poor at ascertaining very fine scale geometric changes. Furthermore, geometry is only part of the picture since the local reflectance is also quite important. Consider surface samples that exhibit 3D texture such as skin, orange peel, velvet, bark, linen, fur, etc. For these widely

varying textured surfaces, no simple model can capture the rich variation in local reflection and the contribution of the fine scale geometry in an effective manner. These issues have led to the increasing popularity of exemplar-based or appearance-based approaches.

Recent work in surface texture addresses the dependence of appearance on imaging conditions. The CURET database (Dana et al., 1997, 1999) provides a starting point in empirical studies of textured surface appearance and it has been used in numerous recent studies (Koenderink et al., 1999; van Ginneken et al., 1998; Leung and Malik, 1999, 2001; Suen and Healey, 1998; Dana and Nayar, 1998, 1999a, 1999b; Suen and Healey, 2000; Liu et al., 2001; Cula and Dana, 2001a, 2001b; Varma and Zisserman, 2002). This database contains texture and reflectance measurements from over 60 different samples, each observed with over 200 different combinations of viewing and illumination directions.

Methods for 3D texture recognition using appearance-based representations derived from these image sets are discussed in Leung and Malik (1999), Suen and Healey (1998), Dana and Nayar (1999a), Varma and Zisserman (2002) and Penirschke et al. (2002). In Dana and Nayar (1999a) the texture representation is the set of conditional histograms of the response to a small multiscale filter bank. Principal component analysis is performed on the histogram of filter outputs and recognition is done using the SLAM library (Nene et al., 1994; Murase and Nayar, 1995). The preliminary recognition results in this experiment were encouraging and motivated the work presented in this paper. In Suen and Healey (1998) the individual texture images are represented using multiband correlation functions that consider both within and between color band correlations. Principal component analysis is used to reduce dimensionality of the representation and this color information is used to aid in recognition. However, the use of texture color in Suen and Healey (1998) greatly assists the recognition since many of the samples in

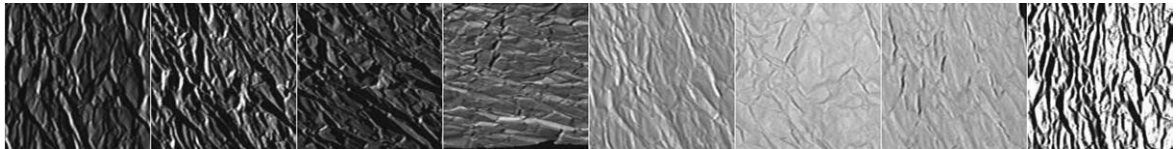


Figure 1. Several instances of the same crumpled paper sample, captured under different imaging conditions. The appearance changes significantly with variations in viewing and illumination directions.

the test dataset are well separated in color space. In Leung and Malik (1999) a 3D texton is created by using a multiscale filter bank applied to an image set for a particular sample. The filter responses as a function of viewing and illumination directions are clustered to form appearance features that are termed 3D textons. In this paper we refer to this method as the *3D texton method*. The methods of Varma and Zisserman (2002) and Penirschke et al. (2002) are recent 3D texture recognition methods which have the property of rotation invariance. The classifier in Penirschke et al. (2002) can classify textures imaged under unknown illumination angle. The methods of Varma and Zisserman (2002) are similar to that of Cula and Dana (2001a) because image features are defined and then the set of feature histograms are used to represent the texture. The image features in Varma and Zisserman (2002) are the maximal filter response over a set of orientations and therefore is a rotational invariant description.

The classical standard framework for 2D texture representation consists of a primitive and a statistical distribution of this primitive over space. So the pertinent question is: how does one extend a standard texture representation to account for 3D texture, where image texture changes with imaging parameters? To accomplish this extension, 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 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. We accomplish a similar goal to that of the 3D texton method, namely precise recognition of textured surfaces. However our approach uses less a priori information and a fundamentally simpler computational representation.

## 2. BTF-Based 3D Texture Representation

We design a 3D textured surface representation based on the BTF, which captures the complex dependency of the texture appearance on the imaging conditions. The proposed representation necessitates the creation of a reduced collection of representatives among the local structural features encountered within texture images.

For each texture image in the BTF a feature histogram is computed to approximate the distribution of key features over the image. The surface descriptor is the collection of feature histograms corresponding to the set of texture images within the BTF. Due to the dependency of texture images on both illumination and viewing directions, our texture representation is referred to as a bidirectional feature histogram. The feature histogram space is highly dimensional and complex, thus dimensionality reduction is achieved by employing principal component analysis (PCA). The resulting surface descriptor is a manifold in the reduced eigenspace, parameterized by the imaging parameters.

### 2.1. Textons and Image Texton Library

Within a texture image there are certain canonical structures, i.e. short edges with different orientations and phases, or spot-like features of various sizes, therefore we seek a reduced set of local structural features. Early psychophysical studies brought evidence that discrimination of black and white textures is based on pre-attentively discriminable conspicuous local features called textons (Julesz, 1981).

Based on the classic concept of textons, and also based on modern generalizations (Leung and Malik, 2001), we obtain a finite set of local structural features that can be found within a large collection of texture images from various samples. We follow the hypothesis that this finite set, called the image texton library, closely represents all possible local structures. Figure 2 schematically illustrates our approach for constructing the image texton library.

A widely used computational approach for encoding the local structural attributes of textures is based on multichannel filtering (Bovik et al., 1990; Jain et al., 1999; Randen and Husoy, 1999; Leung and Malik, 2001). To obtain a computational description of the local feature we employ 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. 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. For simplicity in our discussions we will refer to a single scale, but it is important to keep in mind that the processing is done in parallel for all three scales. As in many approaches in texture literature (Leung and Malik, 2001; Aksoy and Haralick, 1999; Puzicha

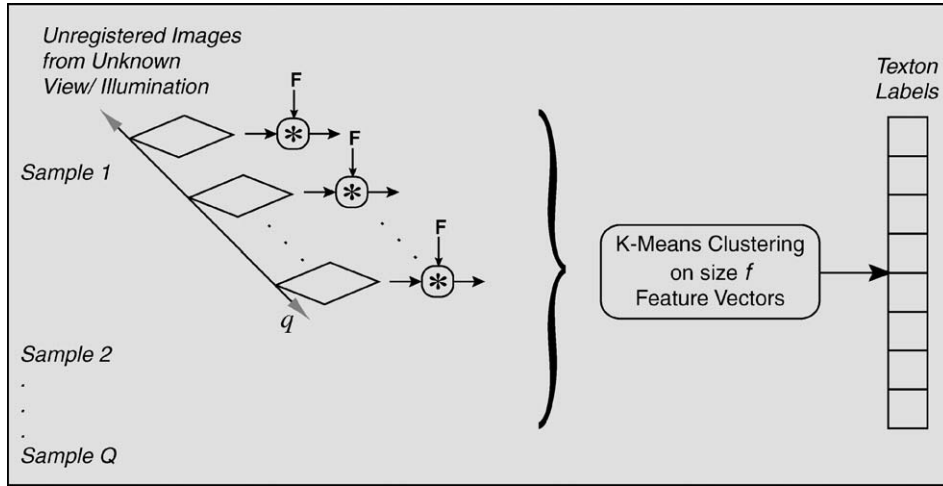


Figure 2. Creation of the image texon library. The set of  $q$  unregistered texture images from the BTF of each of the  $Q$  samples are filtered with the filter bank  $F$  consisting of  $3 \times f$  filters, i.e.  $f$  filters for each of the three scales. The filter responses for each pixel are 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 texon library.

et al., 1999; Ma and Manjunath, 1996), we cluster the feature space to determine the set of prototypes among the population. Specifically, we invoke  $k$ -means algorithm, which is based on the first order statistics of the data, and finds a predefined number of centers in the data space, while guaranteeing that the sum of squared distances between the initial data points and the centers is minimized.

The resulting image texon library is required to be generic enough to represent a large set of 3D texture samples, and comprehensive enough to allow discrimination among different 3D texture samples. In order to verify these two properties, we design two recognition experiments. In the first experiment we use the same 3D texture samples for creating the image texon library and for recognition. The second experiment uses two disjoint 3D texture sample sets for the construction of the image texon library and for recognition. The results are discussed in Section 5.

## 2.2. Texon Histograms

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 texon library, and it is computed as described in Fig. 3. Each texture image is filtered using the same filter bank  $F$  as the one used for creating the texon 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 texon. The spatial distribution of the representative local structural features over the image is approximated by computing the texon 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 texon histogram is a function of the imaging conditions.

Note that in our approach, neither the image texon nor the texon 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 texon 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.

Prior attempts that use histograms to capture the dependency of the texture appearance on the imaging

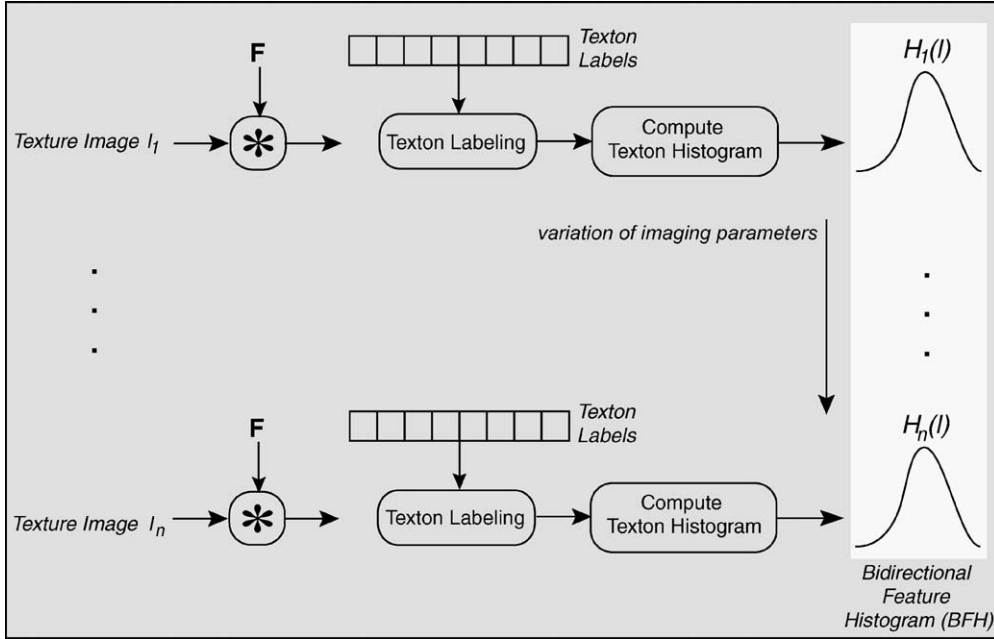


Figure 3. 3D texture representation. Each texture image  $I_j$ ,  $j = 1 \dots n$ , is filtered with filter bank  $F$ , and filter responses for each pixel are concatenated over scale to form feature vectors. The feature vectors are projected onto the space spanned by the elements of the image texon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

conditions include (Dana and Nayar, 1998; van Ginneken et al., 1999). In Dana and Nayar (1998), an analytical model for the bidirectional intensity histogram is formulated, while in van Ginneken et al. (1999) texture histograms as a function of surface illumination and viewing directions are obtained using surface imaging simulations. However, it is unlikely that a single, concise, theoretical model can fully account for the varied appearance of real world surfaces. The bidirectional feature histogram provides an encompassing representation, but it requires exemplars for constructing the surface descriptor.

The dimensionality of histogram space is given by the cardinality of the image texon library, which should be inclusive enough to represent a large range of textured surfaces. 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 Murase

and Nayar (1995), 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.

By performing eigenanalysis in the texon histogram space, we did not impose any assumption of Lambertian reflectance of the surface. PCA is employed not directly on the reflectance values, but on the set of texon histograms, therefore our assumption is that the corresponding projected points in the lower-dimensional eigenspace still retain the descriptive properties of the original bidirectional feature histogram. Our experimental results support this assumption.

### 2.3. Compact Representation of the BTF

Let us consider that each BTF consists of  $N$  texture images, each captured under certain imaging conditions. When  $N$  is very large, as is the case with the CURET database, where  $N = 205$ , the space of the imaging conditions is densely sampled, giving rise to texture images which might present great similarities.

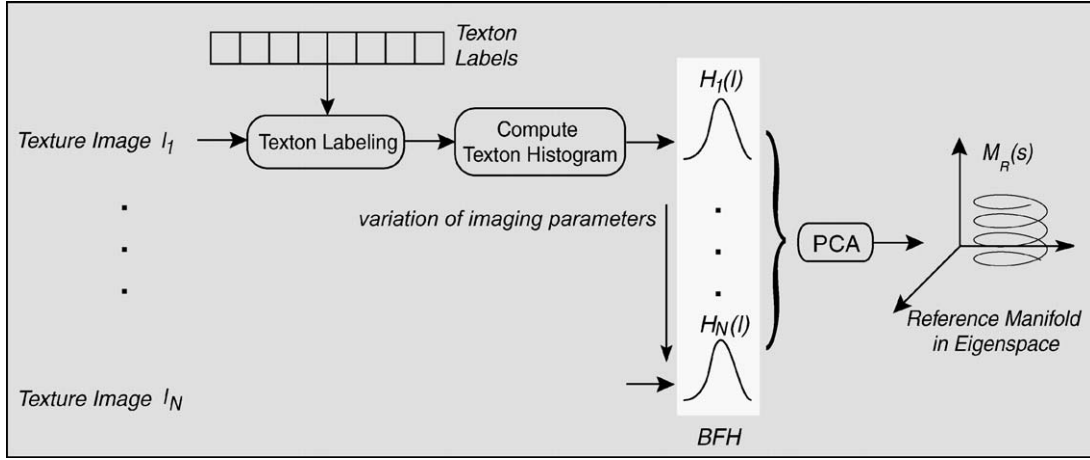


Figure 4. The construction of the reference manifold  $M_R(s)$ . The bidirectional feature histogram (BFH) is computed using all  $N$  texture images  $I_j$ ,  $j = 1 \dots N$ , available within the sampled BTF. PCA is employed to reduce the dimensionality, and the resulting representation in the eigenspace is the reference manifold  $M_R(s)$ , indexed by both the viewing and the illumination directions.

We hypothesize that a subset of this comprehensive representation carries redundant information about the local structure of the surface. In other words, we assume that there exists a subset of texture images which is sufficient for characterizing the 3D textured sample.

As discussed in Section 2.2, our proposed BTF-based texture representation is obtained by employing the set of texon histograms to determine the parametric eigenspace, onto which the bidirectional feature histogram is projected. Our intuition is that there exists an intrinsic topology within the set of points in the eigenspace, and this topology is defined by the viewing and illumination parameters. That is, we assume that the projected points lie on a continuous manifold parameterized by the imaging conditions. Let  $M(s)$  denote this manifold, where parameter  $s$  is indexed by the imaging parameters  $\{\theta_v, \phi_v, \theta_i, \phi_i\}$ , i.e. the polar and azimuthal angles of viewing direction, and the polar and azimuthal angles of illumination direction, respectively.

To identify the core subset of texture images for a certain textured surface we design a quantitative evaluation method for measuring the significance of individual texture images within the overall representation. The three main steps involved in determining the reduced representation of the BTF are depicted in Figs. 4, 6 and 7.

First, as illustrated in Fig. 4, we construct a comprehensive representation of the BTF, called the reference manifold  $M_R(s)$ , which is based on all  $N$  texture images available for representing the surface. Ideally, this

surface appearance manifold is a hypersurface parameterized by four parameters:  $\theta_v, \theta_i, \phi_v$  and  $\phi_i$ , i.e. the reference manifold is a 4-dimensional surface embedded in the high dimensional eigenspace. In practice, for the sake of simplicity, we choose to reduce the 4-dimensional manifold to a one dimensional curve, parameterized by  $s$ , where  $s$  is indexed by the imaging parameters  $\{\theta_v, \phi_v, \theta_i, \phi_i\}$ . While constructing the reference manifold, a predefined ordering of the projected points is necessary. The ordering that we use is defined in the 3-dimensional space of imaging angles:  $\theta_v, \theta_i$ , and  $\phi = |\phi_v - \phi_i|$ . The use of the difference in azimuthal angles  $\phi$  instead of individual azimuthal angles  $\phi_v$  and  $\phi_i$  is strictly correct when the imaged textured surface is isotropic. Because all surfaces employed in our experiments are approximately isotropic, for ordering purposes we employ  $\phi$ . The diameter of the set of points in the 3D space of angles is determined. Then, starting from one of the diametral points, the manifold path is constructed by choosing at each step the closest unassigned point. The metric employed is the Euclidean distance. There are two reasons for this ordering. First, we want the curve representing the manifold to have relatively small variations between two neighboring knot points. Second, we want to objectively compare from one sample to another how removal from the manifold of a point corresponding to a certain combination of imaging parameters affects the overall topology of the manifold.

Second task involved in constructing the compact representation of the BTF is illustrated in Fig. 6. It

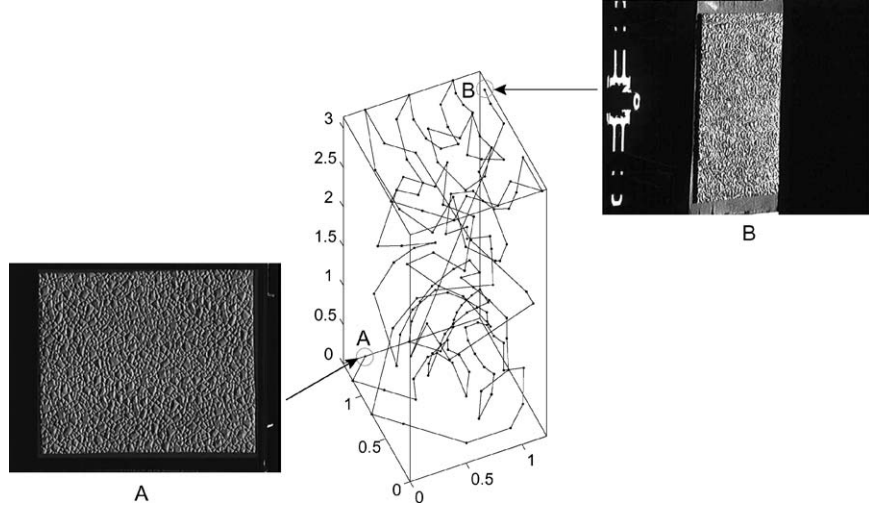


Figure 5. The points in the 3D space of imaging angles:  $\theta_v$ ,  $\theta_i$ , and  $\phi = |\phi_v - \phi_i|$ . The endpoints are denoted by 'o'. The texture images corresponding to the endpoints points A and B of the manifold, are exemplified for sample 4 (rough plastic). The imaging conditions under which texture image A has been captured are  $\theta_{v_A} = 11.25^\circ$ ,  $\phi_{v_A} = 0^\circ$ ,  $\theta_{i_A} = 78.75^\circ$ ,  $\phi_{i_A} = 0^\circ$ , while those under which texture image B has been imaged are  $\theta_{v_B} = 67.5^\circ$ ,  $\phi_{v_B} = 180^\circ$ ,  $\theta_{i_B} = 67.5^\circ$ ,  $\phi_{i_B} = 0^\circ$ .

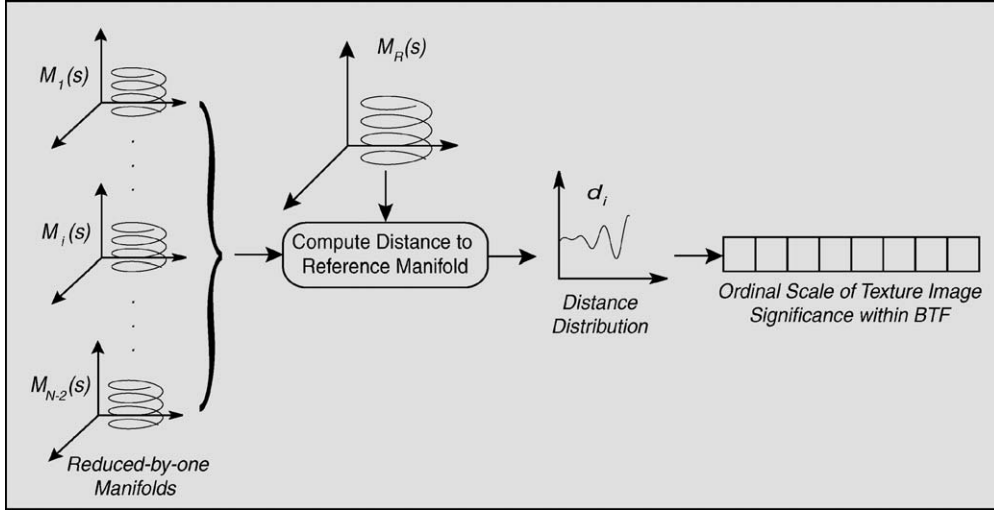


Figure 6. The second step involved in computing the compact representation of the BTF. The set of distances between the reference manifold  $M_R(s)$  and the set of reduced-by-one manifolds  $M_i(s)$ ,  $i = 1 \dots N - 2$ , is used to determine the order of significance of various texture images within the BTF.

consists of creating the set of reduced-by-one manifolds, denoted by  $M_i(s)$ , each obtained by removing from  $M_R(s)$  the point at position  $i$ , where  $i = 1, \dots, N - 2$ , given that the  $N$  points on the manifold are indexed from 0 to  $N - 1$ . The manifold endpoints are not removed because they typically represent texture images captured under very distinct imaging conditions, as illustrated in Fig. 5.

Each of the resulting reduced-by-one manifolds  $M_i(s)$  has  $N - 1$  knot points. The set of  $M_i(s)$  is further analyzed to compute the set of distances  $d_i$  between each  $M_i(s)$  and the reference  $M_R(s)$ .

In practice, each manifold is sampled at a very high rate, and it is stored as a list of points in a multidimensional space. We define the distance between two manifolds as the average of the set of point-to-point

distances. Therefore, in order to compute a meaningful distance between two manifolds we need to sample both manifolds at the same rate. Based on the distribution of distances  $d_i$  between  $M_i(s)$ ,  $i = 1, \dots, N - 2$ , and  $M_R(s)$ , we define an ordinal scale for the relative significance of individual texture images within the BTF. The most important and least important images within the BTF are identified by considering how the manifold changes when the image is removed. The most important texture image is the one that when removed generates the largest distance  $d_i$  between  $M_i(s)$  and the reference  $M_R(s)$ . Similarly, removal of the least important image generates the smallest distance  $d_i$ .

Note that the second step of the proposed algorithm, intended to determine an importance ordering for the texture images within the sampled BTF, necessitates a uniform or quasi-uniform distribution of the sampled points on the hemisphere of all possible imaging conditions. If this property is violated, e.g. the BTF is sampled multiple times at the same point or at very close points, then the algorithm is not accurate anymore, because the texture images corresponding to these points would all be labeled as insignificant. One way to overcome this issue is to modify the current approach such that every time a point is eliminated from the manifold and the distance between the corresponding reduced-by-one manifold  $M_i(s)$  and the reference  $M_i(R)$  is very

reduced, then this point is permanently eliminated from the manifolds, including the reference. However, this modified approach greatly affects the efficiency of the algorithm. The CURET database is constructed such that the property of uniform sampling of the viewing and illumination directions over the hemisphere is ensured, therefore we take advantage of this property, and we utilize the initial approach. Another important property of the CURET database is that the imaging conditions corresponding to the points in which the BTF is sampled, are known. This property allows us to impose the ordering of knot points on the manifold, which is required for objective comparison of the significance of various imaging parameters from one texture sample to another.

The third step in constructing the compact representation of the BTF consists of creating the set of reduced-by- $i$  manifolds  $M^i(s)$ ,  $i = 1, \dots, N - 2$ , as it is described in Fig. 7.  $M^i(s)$  is obtained by eliminating  $i$  points from the manifold, these points being chosen in accordance with the order of significance of individual texture images within the BTF. That is,  $M^i(s)$  is defined as the manifold from which all first  $i$  least significant points are removed. The distribution of the distances between  $M^i(s)$ ,  $i = 1, \dots, N - 2$  and  $M_R(s)$ , denoted by  $d^i$ , is used in determining which of the  $N - 2$  reduced-by- $i$  manifolds  $M^i(s)$  is sufficiently representative, while still encoding the most significant

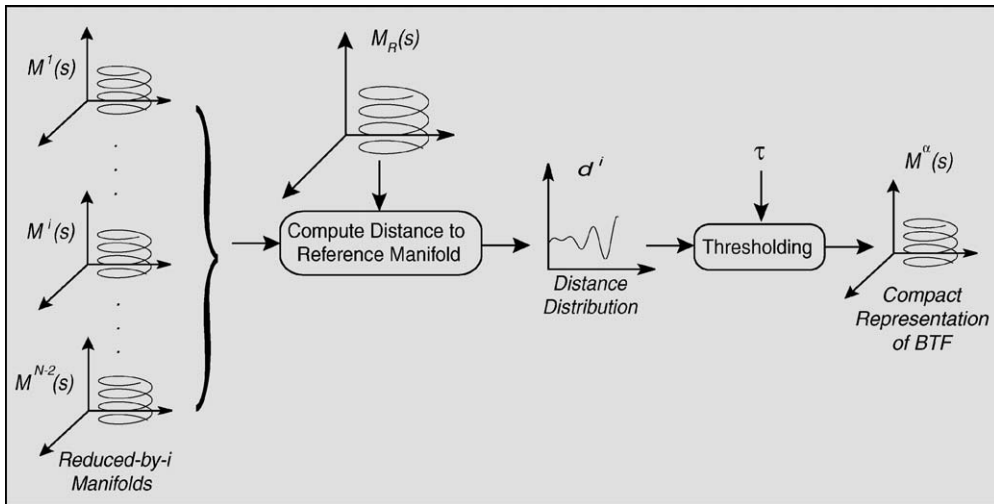


Figure 7. The last task employed during the creation of the compact representation of the BTF. The distances  $d^i$  between the reference manifold  $M_R(s)$  and the reduced-by- $i$  manifolds  $M^i(s)$ ,  $i = 1 \dots N - 2$  are computed.  $M^i(s)$  is obtained by eliminating from  $M_R(s)$  the subset of  $i$  least significant texture images within the BTF. The compact representation of the BTF,  $M^\alpha(s)$ , is the reduced manifold at distance  $d^\alpha$  from  $M_R(s)$ , where  $d^\alpha$  is no larger than a threshold  $\tau$ , defined in Eq. (2).



local structural properties of the textured surface. The compact representation  $M^\alpha(s)$  is determined by imposing that its distance  $d^\alpha$  to the reference manifold should not be larger than a threshold  $\tau$ . More specifically:

$$d^\alpha = \max_{i=1 \dots N-2} \{d^i \mid d^i < \tau\} \quad (1)$$

where

$$\tau = c \cdot d^{\max}, \quad c \in (0, 1] \quad (2)$$

and

$$d^{\max} = \max_{i=1 \dots N-2} \{d^i\}. \quad (3)$$

Consider the imaging parameters corresponding to the most important texture image within the BTF, as it is determined by our method. In our experiments we notice that these important imaging parameters vary from one surface to another, especially when the two surfaces have very distinct local structural features. However, for many samples the most significant texture image is a near frontal view with an oblique illumination that casts shadows and emphasizes structural detail. As an example, consider two surface samples from the CURET database: felt (sample 1) and painted spheres (sample 35). The most significant texture image within the BTF for sample 1 is captured under the imaging parameters:  $\theta_{v_1} = 16.33^\circ$ ,  $\phi_{v_1} = 43.80^\circ$ ,  $\theta_{i_1} = 78.75^\circ$ ,  $\phi_{i_1} = 0^\circ$ , while the most significant texture image within the BTF of sample 35 has as imaging parameters:  $\theta_{v_{35}} = 11.25^\circ$ ,  $\phi_{v_{35}} = 180^\circ$ ,  $\theta_{i_{35}} = 56.25^\circ$ ,  $\phi_{i_{35}} = 0^\circ$ . The most significant images of both samples are illustrated in Fig. 8. Note

how well the geometric detail is depicted in these images. In Section 5 we further discuss the implications of the complexity of fine scale geometry of surface upon the compact representation of the BTF.

### 3. Image Texton Recognition Method

The 3D texture recognition method consists of three main tasks: (1) creation of image texton library, (2) training, and (3) classification. The image texton library is described in Section 2.1 and illustrated in Fig. 2. To extract meaningful textural features we employ a multichannel filter bank  $F$ . Each of the  $q$  texture images, for each of the  $Q$  3D texture samples, is filtered by  $F$  and the filter responses corresponding to a pixel are concatenated to form the feature vector. The feature space is populated with the feature vectors from all  $Q \times q$  images. Considering that texture has repetitive spatial properties, and that different textures exhibit similar local structural features, it is reasonable to assume that the data populating the feature space is redundant. Feature grouping is performed via  $k$ -means algorithm, and the representatives among the population are found, forming the image texton library.

During training, a 3D texture model for each of the 3D texture samples is constructed. Figure 9 illustrates the steps involved in this task. The subset of texture images used for training is arbitrarily sampled from the entire BTF, or it is carefully planned as we suggest in Section 2.3, i.e. the training set is constructed based on the compacted BTF for each surface sample. In our experiments, for comparison reasons, we design the training task in both ways. Each of the  $p$  texture images considered for each of the  $P$  3D texture samples to be classified, is filtered by the same filter bank  $F$  as the one involved for constructing the

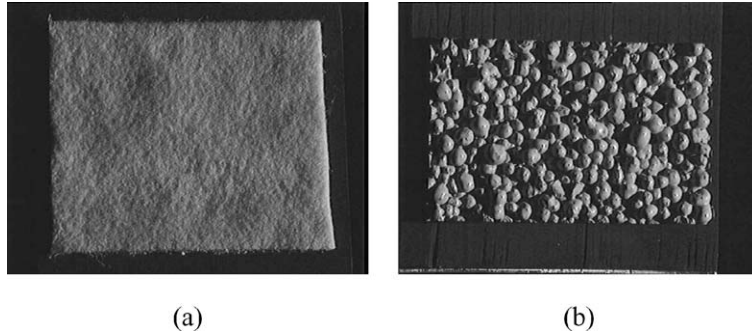


Figure 8. Two surface samples from the CURET database: (a) felt and (b) painted spheres, illustrated by their most significant texture image within the BTF. Notice how well the surface structural detail is depicted in these images.

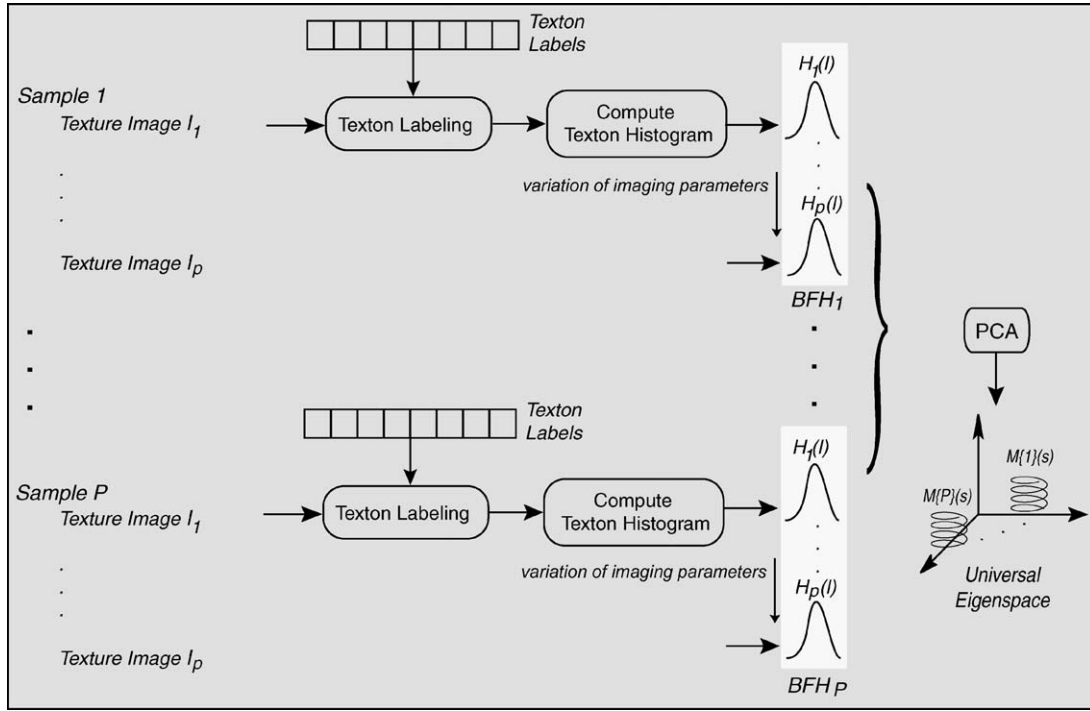


Figure 9. The training step of the image texton method. The set of texture images  $I_j$ ,  $j = 1 \dots p$ , for each of the  $P$  samples, is processed to compute its corresponding  $BFH$ . In the space spanned by the set of  $BFH$ 's for all  $P$  surface samples, PCA is employed and the universal eigenspace is obtained. The projected points corresponding to a certain 3D texture sample, say sample  $k$ , lie on a parametric manifold  $M\{k\}(s)$  indexed by the imaging conditions.

texon library. The resulting multidimensional feature vectors, obtained by concatenating the filter responses for a certain pixel in the image, are labeled relative to the set the image textons. For each texture image a texton histogram is computed. Therefore each 3D texture sample is characterized by the set of  $p$  texture histograms, referred to as a bidirectional feature histogram (BFH). The set of BFH's corresponding to the set of 3D texture samples is employed to compute the universal eigenspace, where each 3D texture sample is modeled as a densely sampled manifold.

In the classification stage, illustrated in Fig. 10, 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

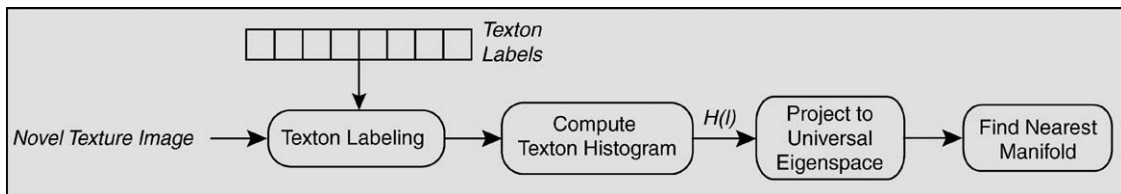


Figure 10. The classification task of the image texton method. A novel texture image, captured under unknown imaging conditions, is processed to compute its texton histogram  $H(l)$ .  $H(l)$  is projected onto the universal eigenspace, and the closest point is found. The surface sample corresponding to the manifold onto which the closest point lies is the class of the query.

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.

To demonstrate the versatility of the image texton library we design two classification experiments. The first experiment uses the same set of textured surfaces for all three tasks involved during the image texton recognition method, while the second experiment uses two disjoint sets of 3D texture samples, one employed in creating the image texton library, and the other involved for recognition. The results are discussed in Section 5.

#### 4. Comparison with 3D Texton Recognition Method

##### 4.1. Summary of 3D Texton Recognition Method

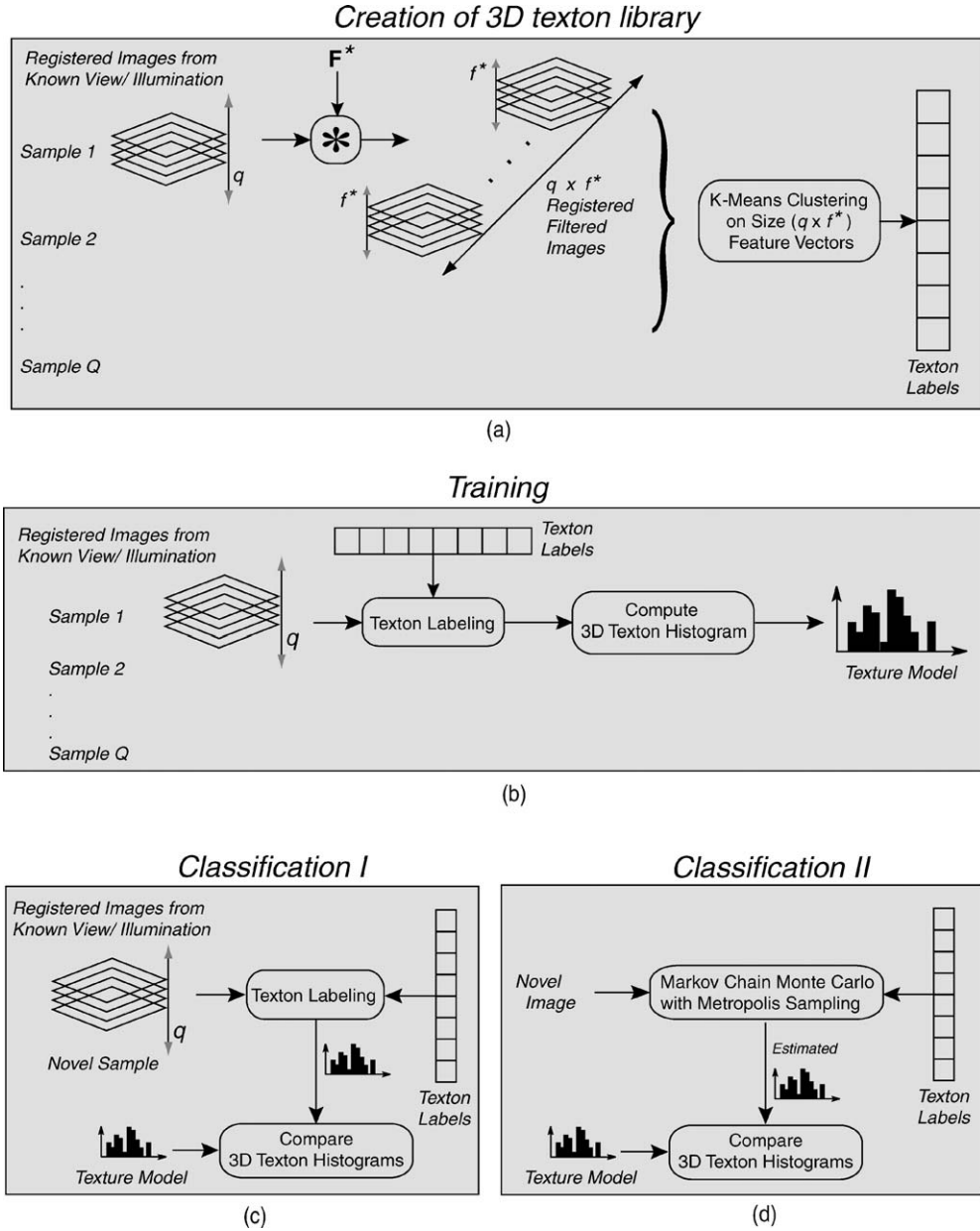
While recognition of 2D textures has been treated extensively, 3D texture recognition has received limited attention. The most notable prior work is the 3D texton method (Leung and Malik, 1999, 2001). For comparison reasons, we briefly summarize this method here. The approach defines and uses the notion of 3D texton, which essentially is a surface feature, parameterized by both viewing and illumination directions. The method consists of 3 main steps: (1) creation of the 3D texton vocabulary, (2) training, and (3) classification. During the first step, depicted in Fig. 11(a), a collection of 3D textons is constructed by using  $Q$  3D texture samples, each represented by  $q$  spatially aligned texture images. Each texture image is filtered by a multichannel filter bank  $F^*$ , of size  $f^*$ . The filter responses corresponding to a pixel, and corresponding to all  $q$  registered texture images of a certain BTF, are grouped together into a highly dimensional feature vector of size  $q \times f^*$ . The resulting feature vector encodes the variations of local structure at a pixel due to changes in the imaging conditions. Because of the repetitiveness of local structures within textured surfaces, and the local similarities between different textures, the set of spatial features is redundant, and it is compacted to a reduced set of prototypes, termed 3D textons. The set of 3D textons is obtained by performing  $k$ -means clustering in the highly dimensional feature space.

The training stage is illustrated in Fig. 11(b). It consists of creating the 3D texture models for all surfaces to be classified. Each 3D texture sample is represented by the 3D texton histogram obtained by first filtering

$q$  registered texture images, captured under the same imaging conditions as the texture images used for creating the 3D texton vocabulary. Then the filter responses of a pixel are aggregated to form feature vectors, which are labeled according to the 3D texton vocabulary. The classification is performed with two methods. The first approach, shown in Fig. 11(c), is non-iterative, and uses a set of novel registered texture images for testing, with known imaging conditions. This set of texture images is processed to determine the corresponding 3D texton histogram, and classification is performed by computing the chi-square distance between all model histograms and the query. The texture class is determined by the closest model histogram. The second classification approach, summarized in Fig. 11(d), is based on a single novel texture image with unknown imaging conditions, and uses an iterative Markov Chain Monte Carlo algorithm with Metropolis sampling to obtain texton labelings and a material classification.

##### 4.2. Comparison of Methods

Every appearance-based texture modeling technique must encode the global distribution of various local structural features over the texture. Furthermore, a suitable representation of 3D texture, presenting fine scale geometry, must also capture the variations of local appearance due to changes in imaging conditions. Both methods, the image texton method, as well as the 3D texton method, have texture representations which satisfy these requirements. However, there are fundamental differences. The primitive of the 3D texton method is the 3D texton, a salient local structural feature of a surface, as a function of the view/illumination angles. The texture representation is the 3D texton histogram, which approximates the distribution of 3D textons over the surface. Therefore the 3D texton-based texture representation accounts for the variations of appearance with the imaging conditions within the local feature, while the global arrangement of local structural characteristics is encoded in the final stage of constructing the texture representation. Alternatively, the image texton method uses the concept of image texton, defined as a key local feature encountered in a single texture image. The local structural feature does not encode any information about the imaging conditions. The global distribution of local features over the texture image is approximated by the image texton histogram. The texture representation is the bidirectional feature histogram, i.e. the collection of image



**Figure 11.** The 3D texton method. (a) For each of the  $Q$  samples, a set of  $q$  registered texture images, with known imaging parameters, is processed to form a  $(q \times f^*)$ -dimensional feature vectors. The feature space is clustered, and the cluster centers form the 3D texton library. (b) In the training stage, each surface sample is represented by a set of  $q$  spatially aligned texture images, with the same imaging conditions as for constructing the 3D texton library. Each surface sample is modeled by a 3D texton histogram. (c) The direct classification method employs a registered set of  $q$  texture images for each sample to be classified. The classification is obtained by comparing the 3D texton histogram computed from the query with all texture models. (d) The second approach for classification uses a single novel image as query, and a MCMC algorithm with Metropolis sampling is employed to estimate the 3D texton histogram.

texton histograms, as a function of both viewing and illumination directions. Therefore, the dependency of texture appearance on the imaging conditions is encoded within the variation of the feature histogram as

a function of the view/illumination angles. The differences between the two approaches have important consequences on the amount of information required about texture images, and on the computational effort

necessary to construct the texture models. Specifically, the 3D texton-based model necessitates a set of *spatially aligned* texture images captured under *known* imaging conditions, while the bidirectional feature histogram needs a set of *unregistered* texture images captured with *unknown* view/illumination angles. Also, the dissimilarity between the two methods influences the process of 3D texture classification: to classify a 3D texture sample using the 3D texton method one would need as input a set of registered texture images with known imaging parameters, while 3D texture classification using the image texton method is based only on a single novel texture image captured under unknown view/illumination directions. The 3D texton method also employs an iterative Markov Chain Monte Carlo technique for classification, which requires as input a single texture image with unknown imaging parameters. However an iterative procedure adds processing time and it raises convergence issues, therefore a direct method is more advantageous.

It is important to note that a quantitative comparison between the two methods it is not appropriate, due to the essential differences in the two approaches. Specifically, in the case of the image texton method, the two sets of imaging conditions corresponding to the training set and the set of testing texture images are disjoint. On the other hand, the 3D texton method with a direct/non-iterative classification procedure uses for training and testing different instances of the same texture image, therefore the imaging conditions of images used for training and testing are identical. When classification is performed iteratively, using a Markov Chain Monte Carlo technique, a quantitative comparison with the image texton method is reasonable, and it shows that the performance of our method is superior to that of the 3D texton method. Our approach generates a recognition rate well above 96%, while the iterative approach for the 3D texton method classification shows a performance of 86%.

Another important issue is the feature grouping in the high-dimensional space of feature vectors. The range of the feature vector component can be quite large, and the average density of points across the data space is likely to be low. The higher the dimensionality of the feature space, the larger the sparseness, therefore a higher-dimensional space represents a greater challenge for clustering. In the case of 3D texton method the feature vector is constructed by filtering the texture image with filter bank  $F^*$  of size  $f^*$ , then grouping the filter responses corresponding to a pixel, and aggregat-

ing the groups of filter responses for a set of spatially aligned images of the same 3D textured surface. The number of feature vector components is  $q \times f^*$ , as mentioned in Section 4.1, where  $q$  is typically 20, and  $f^* = 48$ , therefore the dimensionality of the feature space is 960. For the image texton method the feature vector is obtained by grouping scale-wise the filter responses corresponding to a certain pixel in the image, therefore the dimensionality of the feature space, denoted by  $f$ , is much smaller, i.e.  $f = 15$  in our experiments. Moreover, the dimensionality of the feature space directly influences the size of the computations involved.

Clearly the BFH as a 3D texture representation has some advantages over the texture representation based on 3D textons when a recognition task is to be designed, but is less suitable for synthesis. This is mainly because the image texton method requires less a priori information about the input data. Consequently, the BFH encodes less surface information, therefore it is difficult to use it for rendering the 3D textured surface. On the other hand the 3D texton-based texture representation explicitly encodes the variation of texture appearance as a function of the imaging parameters, allowing rendering of 3D texture more easily.

## 5. Experimental Results

We design three sets of recognition experiments, all using input data from the CURET database, 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.

The first experiment is devised to analyze the dependency of the image texton method performance on two factors: the cardinality of the image texton library, and the dimensionality of the universal eigenspace obtained by performing PCA on the set of image texton histograms corresponding to the set of training texture images. To construct the image texton library, we consider  $Q = 20$  3D texture samples, shown in Fig. 14(a). For each sample we employ a set of  $q = 64$  unregistered texture images corresponding to distinct imaging conditions, given in Table 5 and illustrated in Figs. 22 and 23. These images are uniformly sampled from the complete set of 205 texture images available within CURET database for each texture sample. Every texture image is filtered by filter bank  $F$  of size  $3 \times f = 45$ , illustrated in Fig. 12. To ensure that

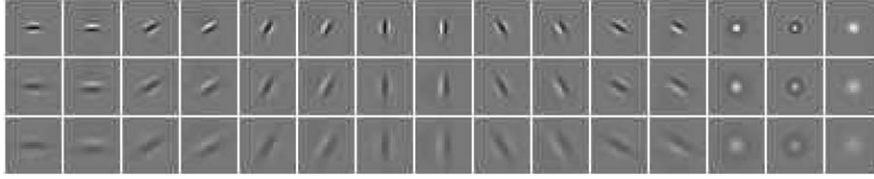


Figure 12. The filter bank employed in our experiments. Each row illustrates  $f = 15$  filters, i.e. 12 oriented derivatives of Gaussian filters with 6 orientations and 2 phases, 2 center-surround derivative of Gaussian filters, and one low-pass Gaussian filter, corresponding to a certain scale. The arrangement of filters on each row is the same as that of filter responses composing the feature vectors. Note that  $F$  is the same as  $F^*$ , with the exception of three low-pass filters that we did not consider.

only texture information is included for analysis, all texture images are manually presegmented and converted to grey level. 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.

We consider 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 (Nene et al., 1994; Murase and Nayar, 1995). 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 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 num-

ber 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 Fig. 13, 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 texton 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%.

The second experiment is more complex and it is designed to analyze the dependency of the image texton method on the number of texture images used for training. This experiment also analyzes the efficiency of representing the BTF through the compacted version, obtained as we propose in Section 2.3. The image texton library employed during the second experiment is one of the versions computed during the first experiment, i.e. the image texton library of size  $K = 100$ . Again, the recognition experiment utilizes the same set of 3D texture samples as for creating the image texton library, illustrated in Fig. 14. For each sample we now consider the entire BTF, with the

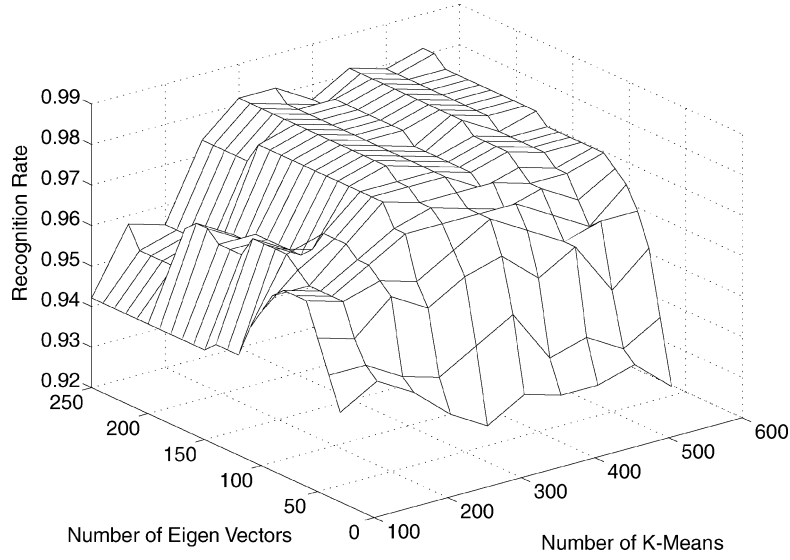
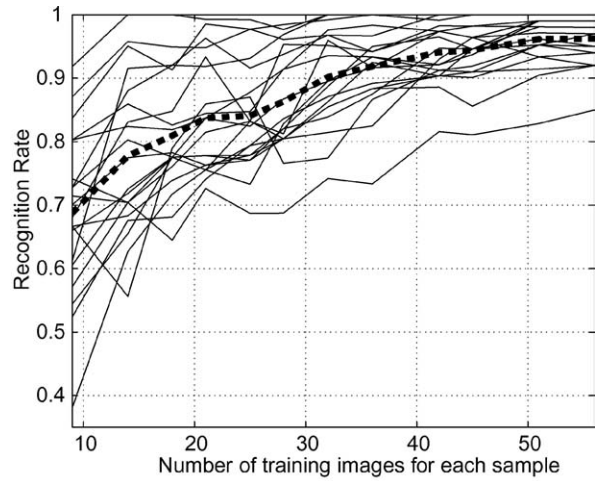


Figure 13. Recognition rate, i.e. the percentage of correctly classified test images, as a function of the cardinality of the image texon library, as well as of the number of dimensions of the eigenspace.



(a)



(b)

Figure 14. (a) The set of 20 3D texture samples used for creating the image texon library. First row: sample 1 (felt), sample 4 (rough plastic), sample 6 (sandpaper), sample 10 (plaster-a), sample 12 (rough paper); second row: sample 14 (roofing shingle), sample 16 (cork), sample 18 (rug-a), sample 20 (styrofoam), sample 22 (lambswool); third row: sample 25 (quarry tile), sample 27 (insulation), sample 30 (plaster-b zoomed), sample 33 (slate-a), sample 35 (painted spheres); fourth row: sample 41 (brick-b), sample 45 (concrete-a), sample 48 (brown bread), sample 50 (concrete-c), sample 59 (cracker-b). The image texon library should be generic enough to represent a large class of 3D texture samples, while it should also be comprehensive enough to allow good segregation between different classes of texture. (b) Recognition rates for individual 3D texture samples (solid lines), and the global recognition rate (thick dashed line), as a function of the number of training images for each sample.

exception of the texture images corresponding to very tilted poses (polar angle of viewing direction  $\theta_v \geq 70^\circ$ ), for which the texture image is almost featureless, therefore their recognition becomes very difficult. The re-

sulting subset consists of  $N = 156$  images for each texture sample, therefore the entire set of texture images processed during the second experiment is of size 3120.

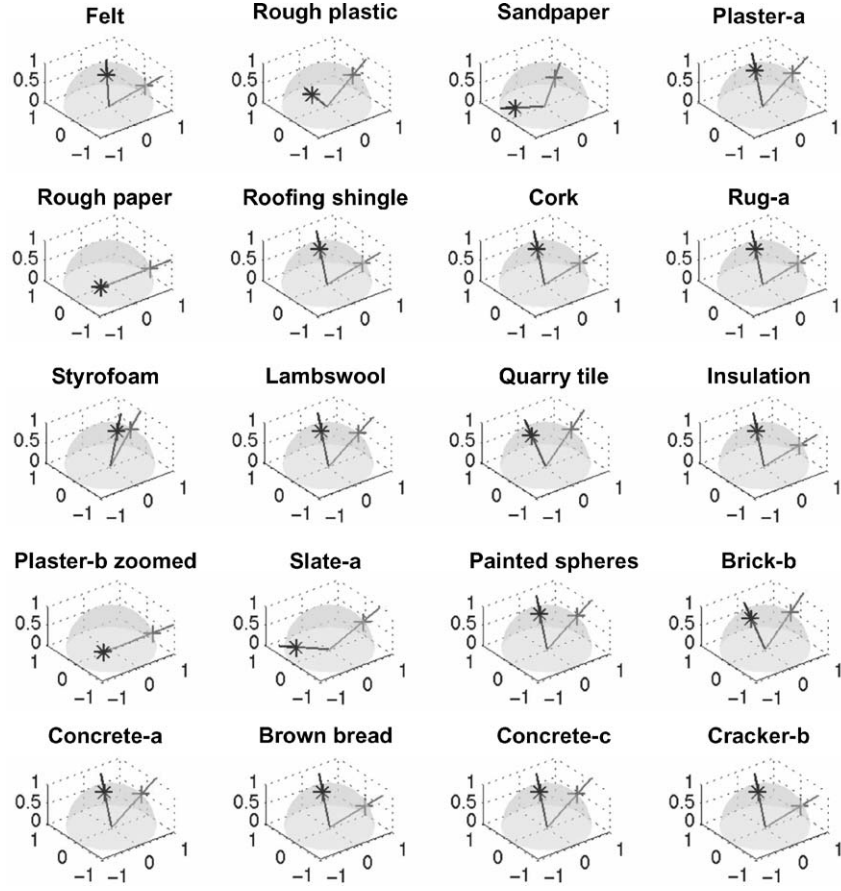


Figure 15. Illustration of the viewing direction (marked by ‘\*’, in blue) and the illumination direction (marked by ‘+’, in red) for the most significant texture images corresponding to each sample involved in recognition during the second experiment. The corresponding imaging conditions are characterized by the angles  $\theta_v, \theta_i, \phi_v, \phi_i$  (in radians) listed in Table 1.

To design the training stage of the image textron recognition method, we compute the compact representation of the BTF for each 3D texture sample. The compacted BTF is obtained by processing the entire BTF to compute the reference manifold  $M_R(s)$ , based on all  $N = 156$  points in the parametric eigenspace of textron histograms. As described in Section 2.3, we determine for each 3D texture sample the subset of significant texture images which best encode the local structural features of the surface. The constraint imposed for obtaining the compact representation of the BTF is consistent across the entire set of surfaces, and it is defined in Eqs. (1)–(3). The distance between the reduced-by- $\alpha$  manifold  $M^\alpha(s)$ , retained as the compacted BTF, and the reference manifold  $M_R(s)$ , does not exceed  $c = 30\%$  of the maximum distance  $d^{\max}$  between any of the reduced-by- $i$  manifolds  $M^i(s)$ ,  $i = 1 \dots N - 2$ , and  $M_R(s)$ .

In the first column of Fig. 16, for six different 3D textured surfaces: sample 4 (rough plastic), sample 10 (plaster-a), sample 20 (styrofoam), sample 22 (lambswool), sample 45 (concrete-a), and sample 50 (concrete-c), we illustrate the profile of the distance  $d_i$  between  $M_i(s)$  and  $M_R(s)$ , normalized by  $d_{\max}$  defined as

$$d_{\max} = \max_{i=1 \dots N-2} \{d_i\}, \quad (4)$$

as a function of the ordinal index  $i$  of the eliminated point on the manifold. It is interesting to observe that the removal of various points from the manifold affects differently the profile of the plot  $d_i$  vs.  $i$ . Also, consider samples 45 and 50, respectively concrete-a and concrete-c. It is noticeable that the profiles of the plots  $d_i$  vs.  $i$  corresponding to these two samples are qualitatively similar, while the two surfaces, illustrated in



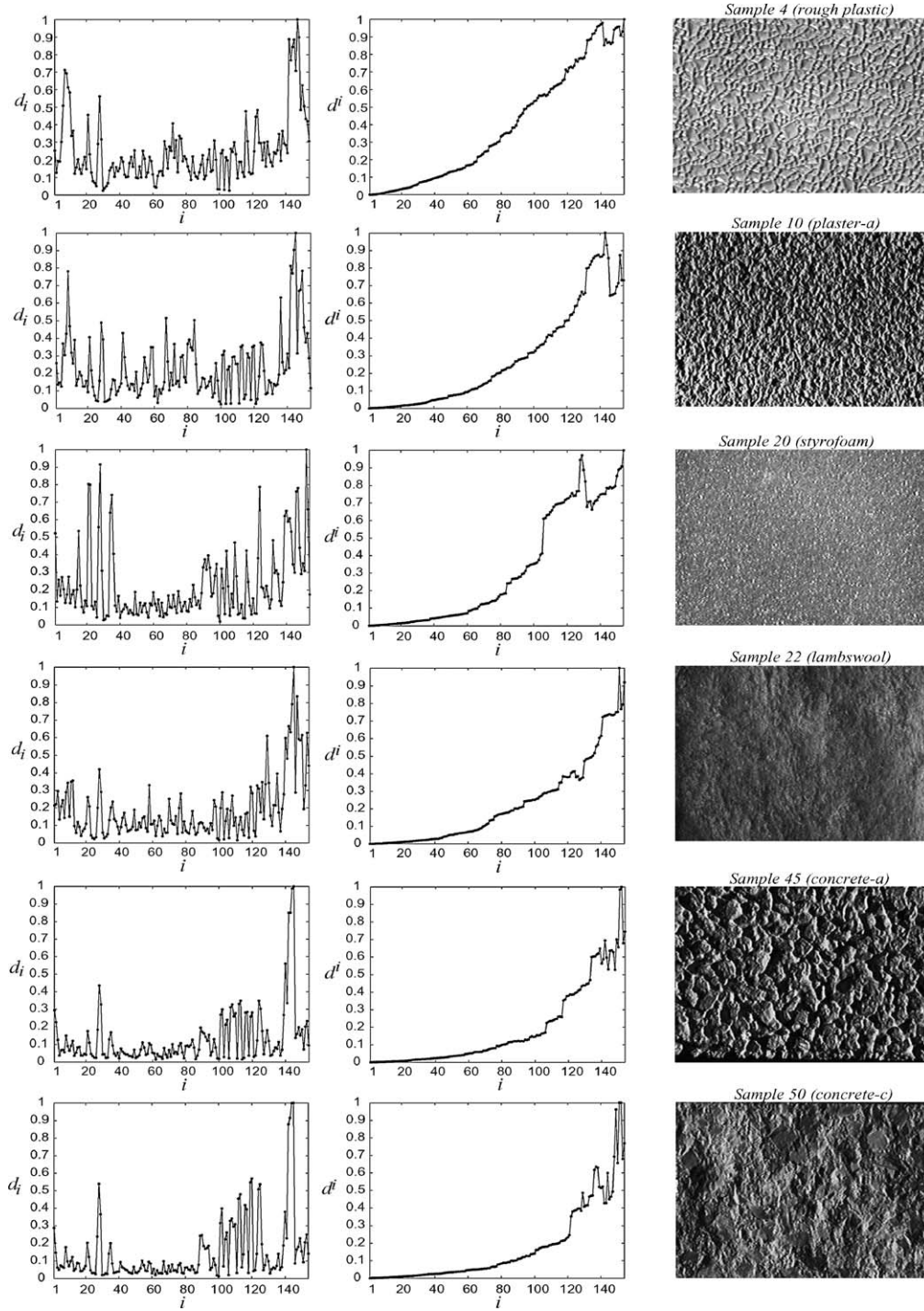


Figure 16. Samples 4 (rough plastic), 10 (plaster-a), 20 (styrofoam), 22 (lambswool), 45 (concrete-a), and 50 (concrete-c). First column: the distance  $d_i$  between the reference manifold  $M_R(s)$  and the reduced manifold  $M_i(s)$ , normalized by  $d_{\max}$ , vs. the ordinal index  $i$  of the eliminated point. Second column: the distance  $d^i$  between the reference manifold  $M_R(s)$  and the reduced manifold  $M^i(s)$ , normalized by  $d^{\max}$ , vs. the number  $i$  of eliminated points. Third column: each 3D textured surface illustrated by the corresponding most significant texture image determined by reducing the manifold. Notice how well the local structure of the surface is captured by the most important texture image within the BTF.

Fig. 16, have similar intrinsic nature, presenting analogous local structural features. These observations show that the differences in the local structure of various surfaces are reflected by the differences between corresponding eigenspace representations. That is, 3D texture samples which have different local structures have compact representations of BTF defined by distinct imaging parameters.

In the second column of Fig. 16, for the same six 3D texture samples, we show the profiles of distance  $d^i$  between  $M^i(s)$  and  $M_R(s)$ , normalized by  $d^{\max}$  defined in Eq. (3), as a function of the number  $i$  of eliminated points from the manifold. Note that for some samples the threshold  $\tau$ , defined in Eq. (2), is reached sooner, while other samples show a slower growth of the profile  $d^i$  vs.  $i$ . As a result different samples have different cardinalities for the compacted BTF. Let  $m$  denote the number of significant texture images within the compacted BTF of a 3D texture sample. We have that  $m = 76$  for sample 4,  $m = 58$  for sample 10,  $m = 60$  for sample 20,  $m = 45$  for sample 22,  $m = 38$  for sample 45, and  $m = 33$  for sample 50. As we expect, the number of significant texture images required within the compact representation of the BTF correlates well with the visual complexity of the surface.

In practice, we store the manifolds as a list of multidimensional points acquired by sampling at high rate the parametric curve in the eigenspace obtained by fitting a quadratic  $b$ -spline interpolation to a set of knot points. The distance between two manifolds is determined by computing the average over the set of point-to-point distances between two corresponding points on the manifolds, given that both manifolds are sampled at the same rate. However, consider the reference manifold  $M_R(s)$ , based on  $N$  knot points, and some reduced manifold  $M^i(s)$  based on  $N - i$  knot points, sampled at the same rate. We employ SLAM to perform PCA, interpolation, and manifold sampling. During sampling, SLAM keeps constant the number of sampled points between two neighboring knot points. Consequently, the distance between two manifolds is dependent on the relative position of the missing points on  $M^i(s)$ , i.e. removal of knot points closer to the end points of the manifold generates a larger distance between the two manifolds, while removal of knot points close to the middle point affects less the configuration of the manifold. Therefore this comparison between  $M_R(s)$  and  $M^i(s)$  is invalid, because it does not capture any information about the texture structure. Our solution for this problem is to replace an eliminated

Table 1. Viewing and illumination directions ( $\theta_v, \theta_i, \phi_v, \phi_i$ , in radians) for the most significant texture images corresponding to each of the samples involved during the second experiment.

Sample	$\theta_v$	$\phi_v$	$\theta_i$	$\phi_i$
Felt	0.285	-2.377	1.371	-0.041
Rough plastic	0.810	-2.702	0.615	-0.685
Sandpaper	1.209	-2.984	0.393	-1.571
Plaster-a	0.196	3.141	0.982	0.000
Rough paper	1.101	-2.433	1.358	-0.281
Roofing shingle	0.196	3.141	1.374	0.000
Cork	0.196	3.141	1.374	0.000
Rug-a	0.196	3.141	1.374	0.000
Styrofoam	0.213	-1.410	0.473	-0.431
Lambswool	0.196	3.141	0.982	0.000
Quarry tile	0.393	3.141	0.785	0.000
Insulation	0.196	3.141	1.374	0.000
Plaster-b zoomed	1.101	-2.433	1.358	-0.281
Slate-a	1.178	3.141	1.178	0.000
Painted spheres	0.196	3.141	0.982	0.000
Brick-b	0.393	3.141	0.785	0.000
Concrete-a	0.196	3.141	0.982	0.000
Brown bread	0.196	3.141	1.374	0.000
Concrete-c	0.196	3.141	0.982	0.000
Cracker-b	0.196	3.141	1.374	0.000

knot point with a point obtained by linearly interpolating its two neighbors. That is, we locally *flatten* the manifold. Consequently, the manifolds  $M_R(s)$ ,  $M_i(s)$  or  $M^i(s)$ ,  $i = 1 \dots N - 2$  are all based on  $N$  knot points. As an example, the set of knot points of the manifold  $M^{N-2}(s)$  consists of  $N$  points lying on the line between the two end points of the manifold, in the multidimensional eigenspace.

From the second column of Fig. 16 it is detectable that the profile of the plot  $d^i$  vs.  $i$  is not monotonic. The nonmonotonicity of  $d^i$  is always observed for large values of  $i$ , i.e. after a large number of points have been already eliminated from the manifold. The reduced manifolds, based on less than about 15% of the original number of knot points, become very unstable, i.e. they lose important topological information about the original manifolds. As a result,  $M^i(s)$  is closer to  $M_R(s)$  than  $M^j(s)$ , when  $i > j$ . As an example, consider the plot of  $d^i$  vs.  $i$  for sample 1 (felt), depicted in Fig. 17(a). Note that  $d^{153} > d^{154}$ , that is the reduced manifold based on the last three points is further away

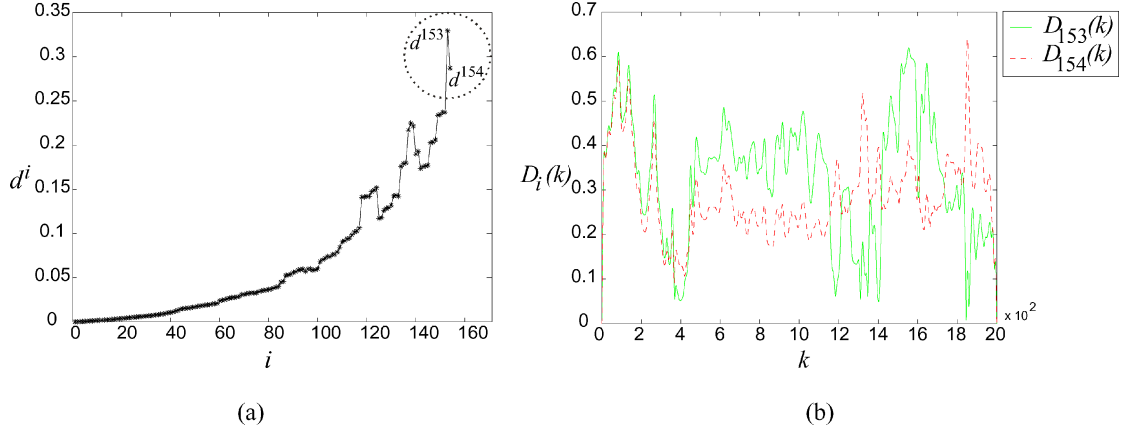


Figure 17. (a) The plot of distance  $d^i$  vs.  $i$ , for sample 1 (felt). The circled region shows that the distance  $d^{153}$  between  $M_R(s)$  and  $M^{153}(s)$  is larger than the distance  $d^{154}$  between  $M_R(s)$  and  $M^{154}(s)$ . (b) The plots of point-to-point distances  $D_i(k)$  between each of the reduced-by- $i$  manifolds  $M^i(s)$ , where  $i = 153, 154$ , and the reference  $M_R(s)$ , as a function of the index  $k$  of the point on the manifold. Note that for most values of  $k$ ,  $D_{153}(k) > D_{154}(k)$ , i.e. the points on  $M^{153}(s)$  are further away from  $M_R(s)$ , than the points on  $M^{154}(s)$ .

from the reference than the reduced manifold based only on the endpoints. The point-to-point distances  $D_i(k)$ ,  $i = 153, 154$ , between  $M^i(s)$ ,  $i = 153, 154$ , and the reference manifold  $M_R(s)$ , as a function of the index  $k$  of the point, are plotted in Fig. 17(b). It can be observed that for many values of  $k$ , we have that  $D_{153}(k) > D_{154}(k)$ . Again, this behavior is present when the reduced-by- $i$  manifolds rely on too few initial knot points and do not preserve anymore the global structural configuration captured by the reference manifold.

The third column of Fig. 16 illustrates the most significant texture image corresponding to each sample. We expect the significant image to be the image which best conveys the physical structure of the surface. Indeed, visual observations support this conjecture. For each 3D texture sample, the most significant image shown in the third column of Fig. 16 clearly depicts detailed local surface structure. This observation suggests that the compact representation of the BTF indeed consists of the subset of texture images which best convey the surface texture.

The construction of a compact representation of the BTF is important for an efficient design of the training set employed for recognition. We study the performance variation of the image texture recognition method as a function of the number of important texture images used for training. We construct 12 variants of training sets, by varying the size of the compact representation of the BTF for each sample. The number  $j$  of significant texture images considered for each 3D texture sample is increased from 1 to 12, with a step of 1, i.e.  $j = 1 \dots 12$ . To determine the global training set we take the union over all 20 surface samples of all individual sets of  $j$  most significant imaging conditions for each sample. Explicitly, consider that each texture image is captured under particular imaging conditions, uniquely determined by the quadruple of imaging parameters  $s = (\theta_v, \theta_i, \phi_v, \phi_i)$ . If a texture image, characterized by a certain  $s'$ , belongs to any of the compacted BTFs, then the global training set will contain the texture image characterized by  $s'$  for all 3D texture samples. For reference, this procedure used for determining the global training set is

Table 2. The number  $j$  of significant texture images considered for each of the 3D texture samples (first row), and the corresponding cardinality  $p_j$  of the training set for each sample (second row). The total number of training  $n_{\text{train}}$  and testing  $n_{\text{test}}$  images are listed on third and fourth rows, respectively.

$j$	1	2	3	4	5	6	7	8	9	10	11	12
$p_j$	9	14	18	21	25	28	32	36	42	45	51	56
$n_{\text{train}}$	180	280	360	420	500	560	640	720	840	900	1020	1120
$n_{\text{test}}$	2940	2840	2760	2700	2620	2560	2480	2400	2280	2220	2100	2000

Table 3. Viewing and illumination directions ( $\theta_v, \theta_i, \phi_v, \phi_i$ , in radians) for the most significant texture images corresponding to each of the samples involved in recognition during the third experiment.

Sample	$\theta_v$	$\phi_v$	$\theta_i$	$\phi_i$
Polyester	1.209	-2.984	0.393	-1.571
Terrycloth	0.196	3.141	0.982	0.000
Leather	0.285	-2.377	1.371	-0.041
Velvet	0.437	-2.701	0.196	-1.571
Pebbles	0.196	3.141	0.982	0.000
Frosted glass	0.393	3.141	0.785	0.000
Aluminum foil	0.196	3.141	0.982	0.000
Rough tile	0.810	-2.702	0.615	-0.685
Rug-b	0.196	3.141	0.982	0.000
Sponge	0.196	3.141	0.982	0.000
Rabbit fur	0.196	3.141	0.982	0.000
Limestone	0.196	3.141	0.982	0.000
Brick-a	0.196	3.141	0.982	0.000
Human skin	0.196	3.141	0.982	0.000
Salt crystals	0.196	3.141	0.982	0.000
Linen	0.628	-1.795	0.662	-1.136
Stones	0.196	3.141	0.982	0.000
White bread	0.196	3.141	0.982	0.000
Wood-a	0.717	-1.271	0.877	-0.767
Tree bark	0.196	3.141	0.982	0.000

referred to as *the merging procedure*. By varying  $j$  from 1 to 12 we obtain various values for the size of the global training set, denoted by  $p_j$ , as listed on second row of Table 2. Therefore we have 12 variants for the second recognition experiment, where each experiment uses a subset of  $p_j$  texture images for training, and a subset of  $N - p_j = 156 - p_j$  texture images for testing. The testing set is the complement of the training set relative to the entire available BTF, consisting of  $N = 156$  texture images. The sizes of the overall training

sets and the corresponding overall testing sets, i.e.  $n_{\text{train}}$  and  $n_{\text{test}}$ , respectively, are specified on the third and fourth rows of Table 2. Note that, while in this work we employ information about the imaging conditions in order to design the training sets, generally the image texture recognition method does not require knowledge about the viewing/illumination parameters. Usually the training sets are arbitrarily chosen from the sampled BTF, as in Cula and Dana (2001b, 2002) and Cula et al. (2003). The reason for using the imaging conditions during training is to test the efficiency of representing the textured surface based on the compacted BTF.

Figure 14(b) depicts the profiles of the global and individual recognition rates, as percentages of correctly recognized images, indexed by the number of training images considered for each sample. As we enrich the training set with more texture images, the recognition rate increases. By employing  $j = 12$  most significant images within each BTF, and by applying the merging procedure we obtain a global training set consisting of  $p_j = 56$  texture images for each sample. The total set employed for training consists of  $n_{\text{train}} = 1120$  images, while during testing a total set of  $n_{\text{test}} = 2000$  novel texture images are classified. This particular configuration of the system achieves an excellent global recognition rate of over 96%. It is notable that four 3D texture samples (roofing shingle, rug-a, lambswool, cracker-b) were 100% correctly classified, while other eight samples (felt, sandpaper, rough paper, cork, insulation, painted spheres, brown bread, concrete-c) reached a recognition rate well above 97%. Moreover, all samples, with one exception, were correctly classified in a proportion larger than 90%. The exception corresponds to sample 4 (rough plastic), for which we achieved a recognition rate of 85%. This type of surface has a rich local structure, which leads to extreme variations of appearance as the imaging parameters are varied. Figure 18 presents a series of eight texture images of sample 4, obtained by sampling the BTF employed in our experiments. The performance obtained

Table 4. The number  $j$  of significant texture images considered for each of the 3D texture samples (first row), and the corresponding cardinality  $p_j$  of the training set for each sample (second row), for the third experiment. The total number of training  $n_{\text{train}}$  and testing  $n_{\text{test}}$  images are listed on third and fourth rows, respectively.

$j$	1	2	3	4	5	6	7	8	9	10	11
$p_j$	8	11	17	22	27	34	39	44	48	51	56
$n_{\text{train}}$	160	220	340	440	540	680	780	880	960	1020	1120
$n_{\text{test}}$	2960	2900	2780	2680	2580	2440	2340	2240	2160	2100	2000

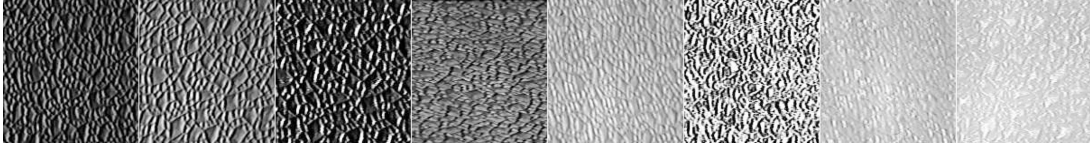
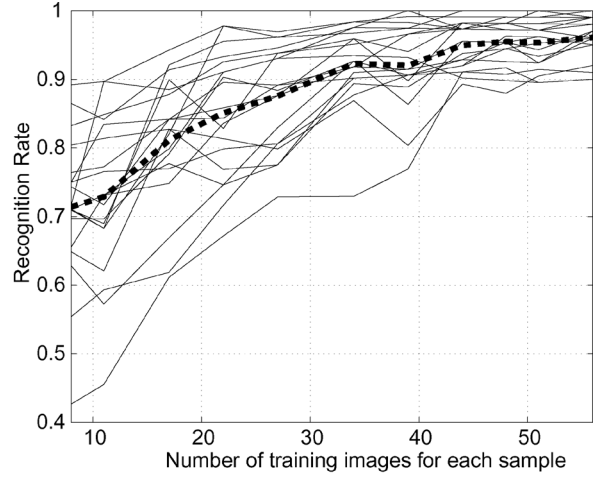


Figure 18. Sample 4 (rough plastic) imaged under various viewing and illumination conditions. The appearance changes so dramatically with the camera/illumination directions, that the images seem to be from different surfaces. These texture images show that, because of surface structure, a large number of texture images are necessary to represent the compacted BTF for this 3D texture sample.



(a)



(b)

Figure 19. (a) The set of 20 3D texture samples employed during the third recognition experiment. First row: sample 2 (polyester), sample 3 (terrycloth), sample 5 (leather), sample 7 (velvet), sample 8 (pebbles); second row: sample 9 (frosted glass), sample 15 (aluminum foil), sample 17 (rough tile), sample 19 (rug-b), sample 21 (sponge); third row: sample 24 (rabbit fur), sample 36 (limestone), sample 37 (brick-a), sample 39 (human skin), sample 43 (salt crystals); fourth row: sample 44 (linen), sample 47 (stones), sample 52 (white bread), sample 54 (wood-a), sample 58 (tree bark). (b) Recognition rates for individual 3D texture samples (solid lines), and the global recognition rate (thick dashed line), as a function of the number of training images for each sample.

for sample 4 suggests that the image texton library involved in computing the 3D texture representation does not comprise key structural features which best convey the surface texture. It is also consistent that sample 4, among the complete set of 20 samples involved in the second experiment, necessitates the largest number of texture images for the compacted BTF, i.e. 76 texture images. This observation shows that our proposed texture representation performs well when encoding the structural complexity of the surface.

In order to verify the validity of our method for determining the compacted BTF, we design an alternative recognition experiment. In this case we do not construct the training set by using the most significant texture images of each 3D texture sample, instead the training set is randomly sampled from the available BTF. We compare the performance of the alterna-

tive recognition experiment, denoted by  $E_r$ , for which the training set consists of 56 randomly chosen texture images for each sample, with the performance of the recognition experiment, denoted by  $E_{12}$ , for which the training set of the same size is designed by applying the merging procedure to the set of the reduced BTF's consisting of 12 most important texture images for each sample. While  $E_{12}$  achieves a global recognition rate of 96%, the alternate experiment  $E_r$  shows a significantly poorer performance of about 86% recognition rate.

When the cardinality of the training set is decreased, as expected, the performance is also gradually diminished. However, let us consider the case when  $j = 1$ , i.e. the global training set is constructed by applying the merging procedure to the compacted BTF containing only the most significant texture image. Table 1

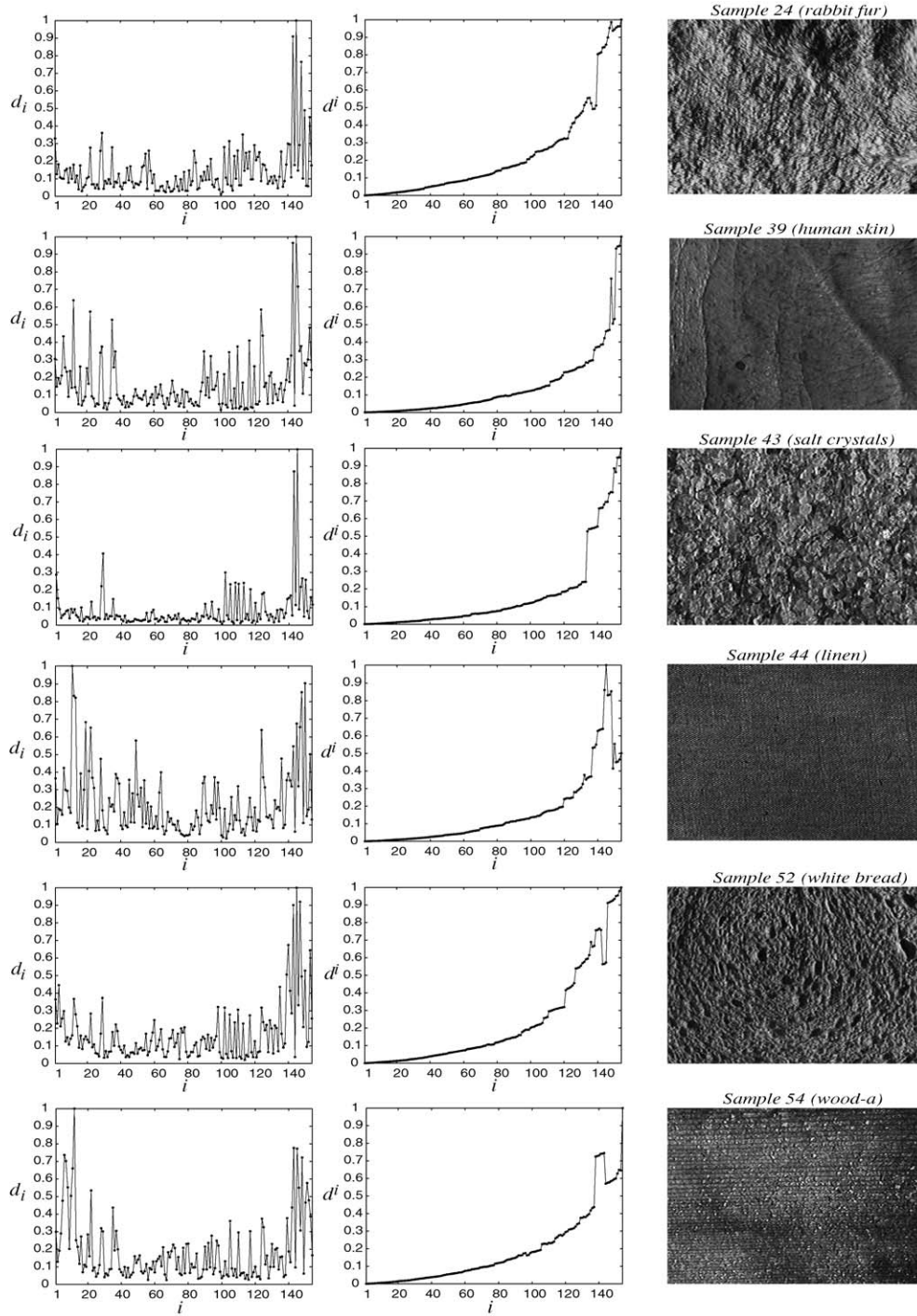


Figure 20. Samples 24 (rabbit fur), 39 (human skin), 43 (salt crystals), 44 (linen), 52 (white bread), and 54 (wood-a). First column: the distance  $d_i$  between  $M_R(s)$  and  $M_i(s)$ , normalized by  $d_{\max}$ , vs.  $i$ . Second column: the distance  $d^i$  between  $M_R(s)$  and  $M^i(s)$ , normalized by  $d^{\max}$ , vs.  $i$ . Third column: each texture surface is illustrated by the corresponding most significant texture image determined by reducing the manifold. Again, the local structure is well exhibited by the most significant images.

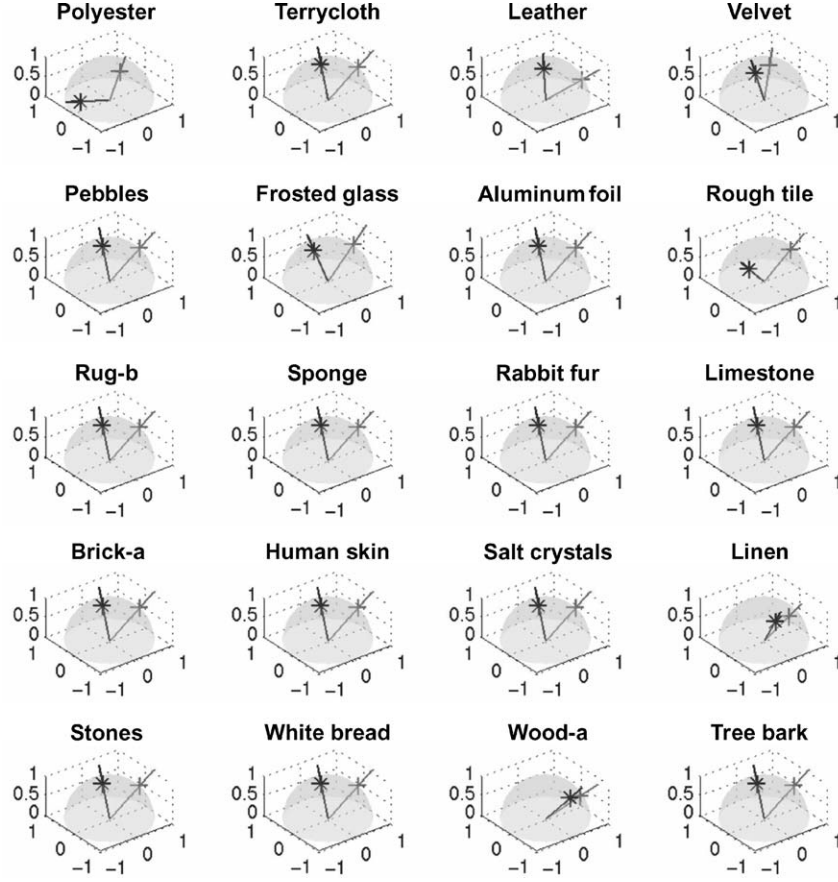


Figure 21. Illustration of the viewing direction (marked by ‘\*’, in blue) and the illumination direction (marked by ‘+’, in red) for the most significant texture images corresponding to each sample involved in recognition during the third experiment. The corresponding imaging conditions are characterized by the angles  $\theta_v, \theta_i, \phi_v, \phi_i$  (in radians) listed in Table 3.

enumerates the four imaging parameters  $\{\theta_v, \phi_v, \theta_i, \phi_i\}$  corresponding to the most significant texture images within the BTF for all  $Q = 20$  samples. These imaging conditions are illustrated on the hemisphere of viewing/illumination directions in Fig. 15. The global training set is of size  $p_j = 9$ , and consequently, the total number of train images is  $n_{\text{train}} = 180$ . Testing is performed for  $n_{\text{test}} = 2980$  novel images. Given this very restricted representation, the global recognition rate is still close to 70%. This result shows the efficiency of our approach for planning the training set by considering the compact representation for each BTF.

Finally, a third experiment is designed to verify the property of generality of the image texton library. To examine if the image texton library is generic enough to represent local structural characteristics of an extensive range of real-world 3D textured surfaces, we design a

recognition experiment which employs for recognition a set of 3D texture samples distinct from the set of samples utilized for computing the texton library. The collection of 3D texture samples involved for classification is illustrated in Fig. 19(a), and consists of  $Q = 20$  samples: sample 2 (polyester), sample 3 (terrycloth), sample 5 (leather), sample 7 (velvet), sample 8 (pebbles), sample 9 (frosted glass), sample 15 (aluminum foil), sample 17 (rough tile), sample 19 (rug-b), sample 21 (sponge), sample 24 (rabbit fur), sample 36 (limestone), sample 37 (brick-a), sample 39 (human skin), sample 43 (salt crystals), sample 44 (linen), sample 47 (stones), sample 52 (white bread), sample 54 (wood-a), sample 58 (tree bark). The setup of the third recognition experiment is similar to the one of the second experiment. We employ the same texton library, and we proceed similarly by using  $N = 156$  texture images



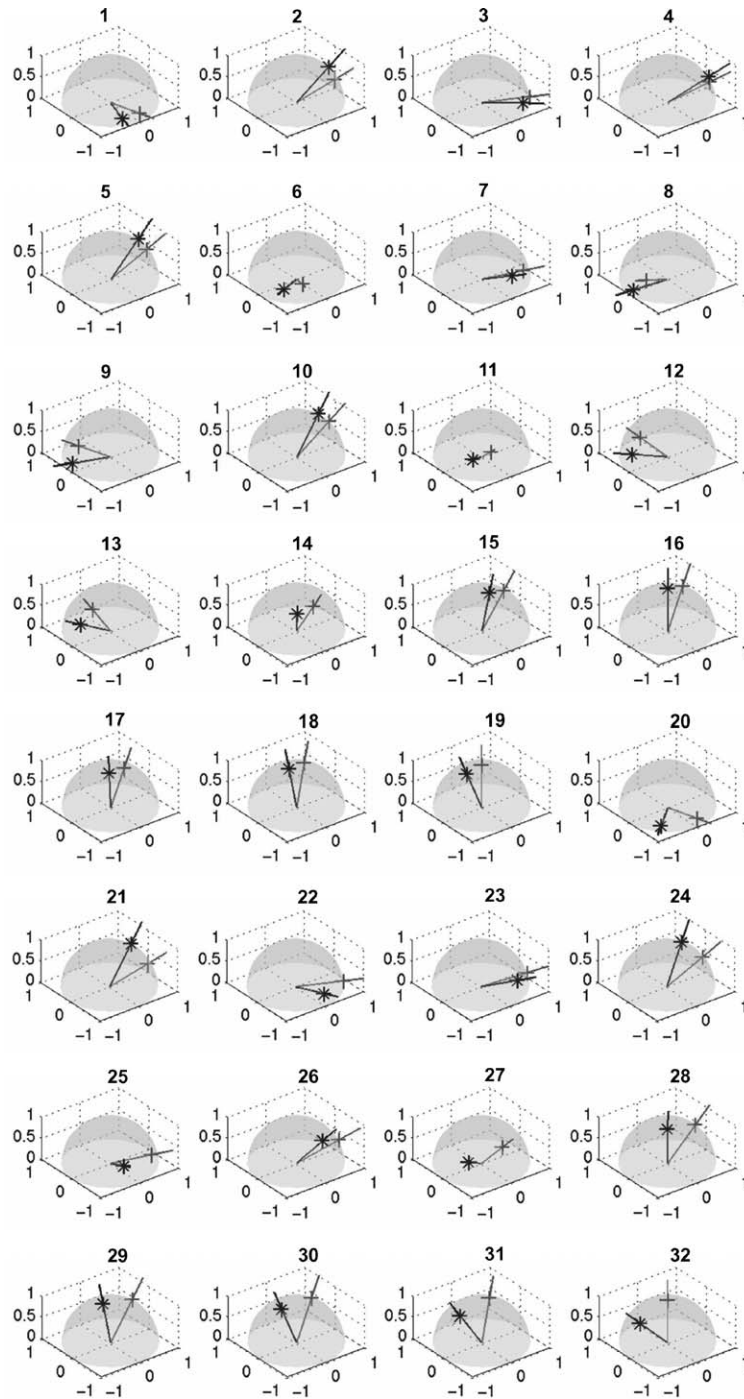


Figure 22. Illustration of the viewing (marked with blue ‘\*’) and illumination (marked with red ‘+’) directions 1–32, as listed in Table 5.

within each BTF. Each BTF is analyzed to compute the corresponding compacted representation. Table 3 and Fig. 21 enumerate the imaging conditions corresponding to the most significant texture image for all 20 sam-

ples involved in recognition. The first column of Fig. 20 illustrates for samples 24 (rabbit fur), 39 (human skin), 43 (salt crystals), 44 (linen), 52 (white bread), and 54 (wood-a), the plots of distance  $d_i$  between  $M_i(s)$  and



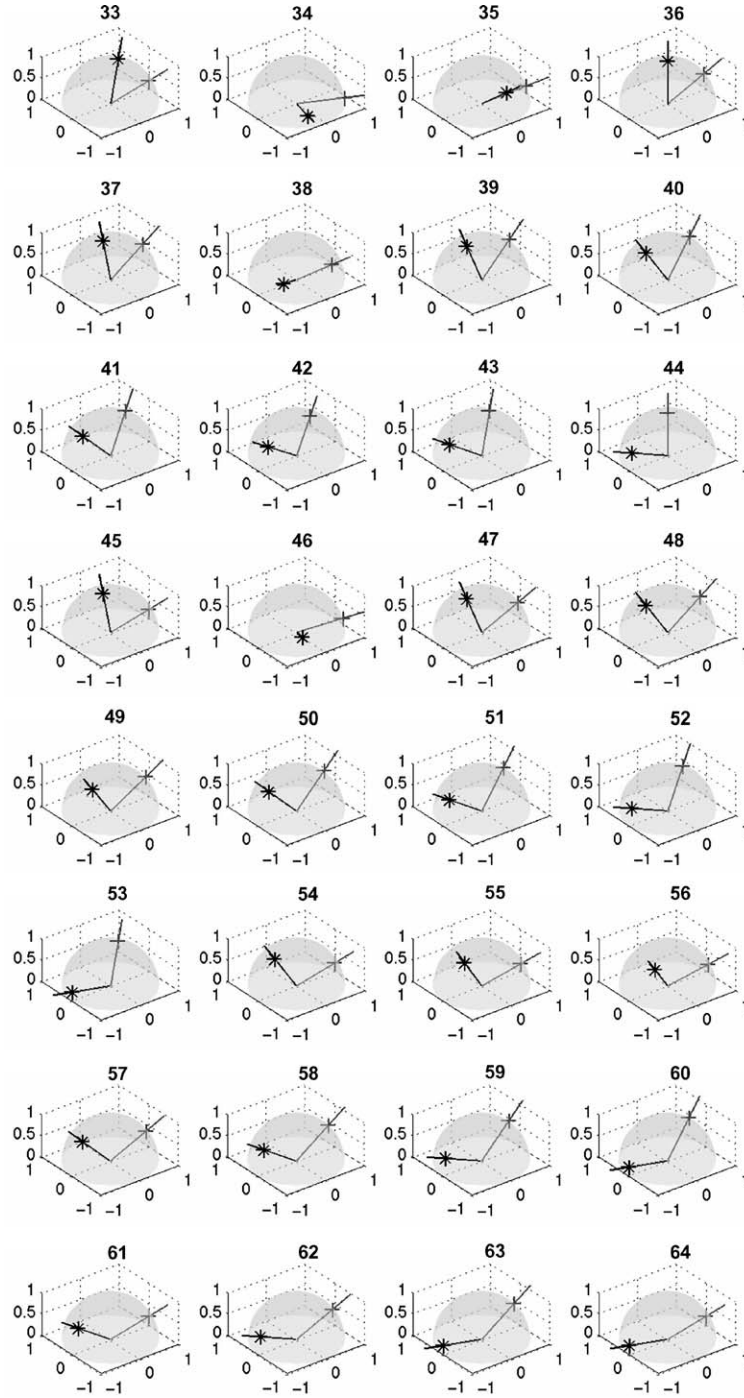


Figure 23. Illustration of the viewing (marked with blue ‘\*’) and illumination (marked with red ‘+’) directions 33–64 as listed in Table 5.

$M_R(s)$ , normalized by  $d_{\max}$ , as a function of the ordinal index  $i$  of the eliminated point on the manifold. Again, the profiles of these plots show that the imaging conditions corresponding to the most important texture

images within the BTF vary from one sample to another. Also, by visually examining the texture images illustrated in Fig. 20 it is noticeable that the surfaces have distinct local structural features. This observation

Table 5. Viewing and illumination directions ( $\theta_v, \theta_i, \phi_v, \phi_i$ , in radians) corresponding to the set of texture images used for creating the image texon library.

No.	$\theta_v$	$\phi_v$	$\theta_i$	$\phi_i$	No.	$\theta_v$	$\phi_v$	$\theta_i$	$\phi_i$
1	1.389	-1.970	1.374	-1.571	33	0.196	0.000	1.374	0.000
2	0.982	0.000	1.374	0.000	34	1.308	-1.978	1.370	-0.764
3	1.308	-1.163	1.371	-0.765	35	0.909	-1.517	1.357	-0.281
4	1.051	-0.417	1.358	-0.153	36	0.000	0.000	1.178	0.000
5	0.785	0.000	1.178	0.000	37	0.196	3.141	0.981	0.000
6	1.272	-2.496	1.154	-2.096	38	1.141	-2.514	0.955	-1.136
7	1.099	-1.476	1.154	-1.045	39	0.392	3.141	0.785	0.000
8	1.421	-3.015	1.130	-2.738	40	0.589	3.141	0.589	0.000
9	1.374	3.141	0.981	3.141	41	0.785	3.141	0.392	0.000
10	0.589	0.000	0.981	0.000	42	0.987	-3.007	0.284	-0.764
11	1.101	-2.432	0.955	-2.005	43	0.981	3.141	0.196	0.000
12	1.178	3.141	0.785	3.141	44	1.178	3.141	0.000	0.000
13	1.038	-3.013	0.662	-2.860	45	0.196	3.141	1.374	0.000
14	0.694	-2.211	0.589	-1.570	46	1.154	-2.096	1.360	-0.430
15	0.213	-1.410	0.472	-0.430	47	0.392	3.141	1.178	0.000
16	0.000	3.141	0.392	0.000	48	0.589	3.141	0.981	0.000
17	0.274	-2.335	0.284	-0.764	49	0.662	-2.860	0.955	-0.153
18	0.196	3.141	0.196	0.000	50	0.785	3.141	0.785	0.000
19	0.392	3.141	0.000	0.000	51	0.981	3.141	0.589	0.000
20	1.432	-2.365	1.374	-1.570	52	1.178	3.141	0.392	0.000
21	0.589	0.000	1.374	0.000	53	1.374	3.141	0.196	0.000
22	1.285	-1.570	1.370	-0.764	54	0.589	3.141	1.374	0.000
23	1.122	-1.226	1.360	-0.430	55	0.625	-2.857	1.370	-0.041
24	0.392	0.000	1.178	0.000	56	0.733	-2.628	1.360	-0.094
25	1.122	-1.914	1.154	-1.045	57	0.785	3.141	1.178	0.000
26	0.717	-1.267	1.130	-0.403	58	0.981	3.141	0.981	0.000
27	1.047	-2.526	0.785	-1.570	59	1.178	3.141	0.785	0.000
28	0.261	-2.202	0.662	-0.281	60	1.374	3.141	0.589	0.000
29	0.196	3.141	0.589	0.000	61	0.981	3.141	1.374	0.000
30	0.392	3.141	0.392	0.000	62	1.178	3.141	1.178	0.000
31	0.589	3.141	0.196	0.000	63	1.374	3.141	0.981	0.000
32	0.785	3.141	0.000	0.000	64	1.374	3.141	1.374	0.000

shows that the compact representation of the BTF correlates well with the surface local geometric characteristics, that is samples exhibiting distinct structural attributes have corresponding significant texture images captured under different imaging conditions. The distance  $d^i$  between  $M^i(s)$  and  $M_R(s)$ , normalized by  $d^{\max}$ , indexed by the number of eliminated points  $i$  is depicted for all six samples of interest in the second column of Fig. 20. These plots suggest that each surface, depending on its structural complexity, requires

a specific number of significant texture images within its compacted BTF. For these six samples, the third column of Fig. 20 shows the corresponding most important texture images within each BTF. Again, it is notable how well the local structural characteristics of surface are illustrated in these images.

For this third experiment we construct 11 variants of global training sets, obtained by considering  $j = 1 \dots 11$  most significant texture images within each BTF and by applying the merging procedure.

The global training sets have cardinalities as listed in Table 4. We investigate the performance of the system as the total number of training images  $n_{\text{train}}$  and the total number of test images  $n_{\text{test}}$  are varied as enumerated in the third and fourth rows of Table 4. The recognition rate as a function of the size of the global training set for each 3D texture sample is depicted in Fig. 19(b). When only the most significant texture image for each sample is considered, we train the system on a total number of  $n_{\text{train}} = 160$  texture images, while  $n_{\text{test}} = 2960$  novel texture images are used for classification. Though this representation is extremely simple, the global recognition rate is well above 70%. As the set of important images for each sample is expanded, the recognition rate increases, reaching 96% when the global training set contains 56 images for each surface. This global training set is obtained by applying the merging procedure when considering the most important 11 texture images per sample. These excellent classification results confirm the validity of the assumption that different 3D textured surfaces have similar local structural properties. Therefore an image texton library obtained from a limited collection of 3D texture samples also conveys the texture characteristics of other surfaces.

There are limitations of the texton library. For instance in our experiments we identify surfaces like sample 28 (crumpled paper) or sample 38 (ribbed paper), which exhibit structural characteristics not well represented by the texton library. As a result the texture models of these two samples are not well discriminated. This observation shows that the breadth of the image texton library is a considerable factor in designing efficiently the texture model we propose in this work.

## 6. Conclusion

Due to the inherent microgeometry of the real world surfaces, the surface appearance varies significantly when the viewing and lighting directions change, due to effects such as shadowing, shading or local occlusions. Consequently, an efficient surface representation must account for these changes. In this work we introduce a BTF-based representation for 3D textured surfaces, termed *bidirectional feature histogram (BFH)*. The construction of the BFH consists of two main tasks: first, the image texton library is created from a heterogeneous collection of surface samples such that basic image features, termed image textons, are well represented; second, a set of texture images is

processed to obtain the surface model, that is a collection of image feature histograms, as a function of the imaging conditions. Based on the BFH, we have designed a quantitative evaluation for the level of relative significance of texture images within the BTF, allowing each BTF to be reduced to a core subset of texture images which best conveys the surface characteristics. We have developed a 3D texture recognition method which constructs a BFH-based surface model, and it classifies the surface based on a single novel texture image of unknown imaging parameters. Our recognition experiments, which utilized over 6200 texture images corresponding to 40 real-world surface samples from the CURET database, show excellent classification results. The results are especially interesting when one considers that there is no need for a priori information about the imaging parameters during surface classification.

The methods presented here have important implications for computer vision. For 3D texture recognition or classification, invariance to illumination or viewing direction is essential for robust performance. Our approach has the key advantages that a single image can be used for recognition without the need for iterative methods and the viewing/illumination direction of the novel images need not be known. Furthermore, the images in the training set need not be registered. An additional impact of our work is the identification of a smaller subset of images from the CURET database that can be used to sufficiently represent the texture class. The CURET database contains over 200 images per sample and it is useful to determine which images can be considered redundant using a quantitative measure. Indeed this analysis can be used for BRDF/BTF measurement planning in subsequent experiments.

The work described also has implications for computer graphics. Histograms of textural features have become an important building block for texture synthesis techniques. Compact representations of surface texture such as the one described here can be modified to support efficient and realistic rendering of natural objects with complex surface texture.

## Acknowledgment

We thank Thomas Leung for providing the filter bank used in his experiments. This material is based upon work supported by the National Science Foundation under Grant No. 0092491 and Grant No. 0085864.

## References

- Aksoy, S. and Haralick, R.M. 1999. Graph-theoretic clustering for image grouping and retrieval. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 63–68.
- Bovik, A.C., Clark, M., and Geisler, W.S. 1990. Multichannel texture analysis using localized spatial filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(1):55–73.
- Chantler, M. 1995. Why illuminant direction is fundamental to texture analysis. *IEEE Proceedings Vision, Image and Signal Processing*, 142(4):199–206.
- Cula, O.G. and Dana, K.J. 2001a. Compact representation of bidirectional texture functions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, vol. I, pp. 1041–1047.
- Cula, O.G. and Dana, K.J. 2001b. Recognition methods for 3D textured surfaces. In *Proceedings of SPIE Conference on Human Vision and Electronic Imaging VI*, vol. 4299, pp. 209–220.
- Cula, O.G. and Dana, K.J. 2002. Image-based skin analysis. In *Proceedings of Texture 2002—The 2nd International Workshop on Texture Analysis and Synthesis*, pp. 35–40.
- Cula, O.G., Dana, K.J., Murphy, F.P., and Rao, B.K. 2004. Skin texture modeling. *International Journal of Computer Vision* (to appear).
- Dana, K.J. and Nayar, S.K. 1998. Histogram model for 3D textures. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 618–624.
- Dana, K.J. and Nayar, S.K. 1999a. 3D textured surface modeling. In *IEEE Workshop on the Integration of Appearance and Geometric Methods in Object Recognition*, pp. 46–56.
- Dana, K.J. and Nayar, S.K. 1999b. Correlation model for 3D texture. *International Conference on Computer Vision*, pp. 1061–1067.
- Dana, K.J., van Ginneken, B., Nayar, S.K., and Koenderink, J.J. 1997. Reflectance and texture of real world surfaces. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 151–157.
- Dana, K.J., van Ginneken, B., Nayar, S.K., and Koenderink, J.J. 1999. Reflectance and texture of real world surfaces. *ACM Transactions on Graphics*, 18(1):1–34.
- Jain, A., Prabhakar, S., and Hong, L. 1999. A multichannel approach to fingerprint classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(4):348–369.
- Julesz, B. 1981. Textons, the elements of texture perception and their interactions. *Nature*, 290:91–97.
- Koenderink, J.J., van Doorn, A.J., Dana, K.J., and Nayar, S.K. 1999. Bidirectional reflection distribution function of thoroughly pitted surfaces. *International Journal of Computer Vision*, 31(2/3):129–144.
- Leung, T. and Malik, J. 1999. Recognizing surfaces using three-dimensional textons. *International Conference on Computer Vision*, 2:1010–1017.
- Leung, T. and Malik, J. 2001. Representing and recognizing the visual appearance of materials using three-dimensional textons. *International Journal of Computer Vision*, 43(1):29–44.
- Liu, X., Yu, Y., and Shum, H. 2001. Synthesizing bidirectional texture functions for real-world surfaces. In *Proceedings of SIGGRAPH*, pp. 97–106.
- Ma, W.Y. and Manjunath, B.S. 1996. Texture features and learning similarity. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 425–430.
- McGunnigle, G. and Chantler, M.J. 2000. Rough surface classification using first order statistics from photometric stereo. *Pattern Recognition Letters*, 21:593–604.
- Murase, H. and Nayar, S.K. 1995. Visual learning and recognition of 3-D objects from appearance. *International Journal of Computer Vision*, 14(1):5–24.
- Nene, S.A., Nayar, S.K., and Murase, H. 1994. SLAM: A software library for appearance matching. Technical Report CUCS-019-94 Proceedings of ARPA Image Understanding Workshop.
- Penirschke, A., Chantler, M., and Petrou, M. 2002. Illuminant rotation invariant classification of 3D surface textures using Lissajous's ellipses. In *Proceedings of Texture 2002—The 2nd International Workshop on Texture Analysis and Synthesis*, pp. 103–108.
- Puzicha, J., Hoffman, T., and Buchmann, J. 1999. Histogram clustering for unsupervised image segmentation. *International Conference on Computer Vision*, 2:602–608.
- Randen, T. and Husoy, J.H. 1999. Filtering for texture classification: A comparative study. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(4):291–310.
- Suen, P. and Healey, G. 1998. Analyzing the bidirectional texture function. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 753–758.
- Suen, P. and Healey, G. 2000. The analysis and recognition of real-world textures in three dimensions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(5):491–503.
- Tong, X., Zhang, J., Liu, L., Wang, X., Guo, B., and Shum, H.-Y. 2002. Synthesis of bidirectional texture functions on arbitrary surfaces. In *Proceedings of SIGGRAPH*, pp. 665–672.
- van Ginneken, B., Koenderink, J.J., and Dana, K.J. 1999. Texture histograms as a function of irradiation and viewing direction. *International Journal of Computer Vision*, 31(2/3):169–184.
- van Ginneken, B., Stavridi, M., and Koenderink, J.J. 1998. Diffuse and specular reflectance from rough surfaces. *Applied Optics*, 37:130–139.
- Varma, M. and Zisserman, A. 2002. Classifying images of materials. In *Proceedings of the European Conference on Computer Vision*, pp. 255–271.