Hybrid Textons: Modeling Surfaces with Reflectance and Geometry *

Jing Wang and Kristin J. Dana
Electrical and Computer Engineering Department
Rutgers University
Piscataway, NJ, USA
{*jingwang,kdana*}*@caip.rutgers.edu*

Abstract

The appearance of surface texture as it varies with angular changes of view and illumination is becoming an increasingly important research topic. The bidirectional texture function (BTF) is used in surface modeling because it describes observed image texture as a function of imaging parameters. The BTF has no geometric information, as it is based solely on observed texture appearance. Computational tasks such as recognizing or rendering typically require projecting a sampled BTF to a lower dimensional subspace or clustering to extract representative textons. However, there is a serious drawback to this approach. Specifically, cast shadowing and occlusions are not fully captured. When recovering the full BTF from a sampled BTF with interpolation, the following two characteristics are difficult or impossible to reproduce: (1) the position and contrast of the shadow border, (2) the movement of the shadow border when the imaging parameters are changed continuously. For a textured surface, the nonlinear effects of cast shadows and occlusions are not negligible. On the contrary, these effects occur throughout the surface and are important perceptual cues to infer surface type. In this paper we present a texture representation that integrates appearance-based information from the sampled BTF with concise geometric information inferred from the sampled BTF. The model is a hybrid of geometric and image-based models and has key advantages in a wide range of tasks, including texture prediction, recognition, and synthesis.

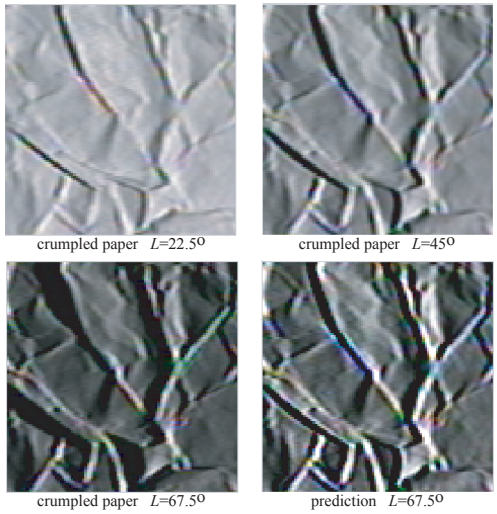

Figure 1: Cast shadows cannot be obtained via interpolation. Top Row: Image of crumpled paper with the illumination $L = 22.6^\circ$ (left) and $L = 45^\circ$ (right). Bottom Row: Image of crumpled paper with $L = 67.5^\circ$(left) and the predicted image for $L = 67.5^\circ$ obtained by a linear combination of the images in the top row. Predicting texture images that are not in the sampled BTF is typically done by interpolation or a linear transformation of the BTF sample images. But as this figure illustrates, interpolation cannot account for cast shadows. A dual effect occurs with occlusions. The artifacts in prediction appearance are especially apparent when the illumination or viewing direction changes in a continuous manner.

1 Introduction

The appearance of surface texture as it varies with angular changes of view and illumination is becoming an increasingly important research topic. Since objects are comprised of surfaces and scenes are comprised of objects, surface appearance is a fundamental issue in developing algorithms in vision and graphics. Ubiquitous high resolution imaging has made the quest for more advanced surface models a timely priority. Also, recent refinement of standard vision and graphics algorithms means performance expectations have increased. Simple shading and texture models are no longer satisfactory for real world scenes.

At a point on the surface, the reflectance depends on two directions, the incident light L and the camera direction V. In a region on the surface, the reflectance varies spatially and we assume it has certain uniform statistics so that the observed region is an image texture. This image texture also varies with L and V, so that it is natural to think of it as a bidirectional texture function (BTF) denoted by $f(x, y, L, V)$ where x, y are the local coordinates on a surface patch. For fixed imaging parameters L_0 and V_0, the

*This research is supported by the National Science Foundation under Grant No. 0092491 and Grant No. 0085864.

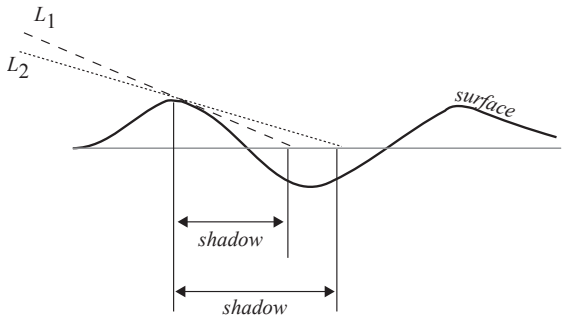

Figure 2: An illustration of the creation of a shadow on a rough surface. The shadow length is determined by the illumination direction. As the illumination direction changes from L_1 to L_2 the shadow grows accordingly. The location, length, and continuous change of the shadow can be determined using geometric considerations. However, predicting cast shadows (that are not observed in the BTF sample images) is not possible with a purely image-based approach.

BTF value $f(x, y, L_0, V_0)$ is an image. Note that the BTF has no geometric information. Complex light interactions that occur on non-smooth surfaces such as shadowing, occlusion and local foreshortening are part of the observed image texture.

Models for bidirectional texture and bidirectional reflectance sometimes are obtained by analyzing physical models of the surface as in [13][14][15][7][17][3]. Considering the complexity and variety of real-world surfaces, physical models of the BTF may be quite complicated. Current physically based models are more suitable for reflectance prediction as opposed to texture. As such, appearance-based or image-based approaches have gained favor in the literature. Texture recognition methods using representations based on measured BTFs include [4][9][1][2][18]. Texture synthesis or rendering methods using appearance-based texture representations include [12][10][16][8][11][6].

The underlying commonality in most existing texture representations is the use of a collection of images and interpolate new views using a compression of these images. Generating the texture between the set of sampled imaging parameters requires some type of interpolation. This is a fundamental limitation of image-based approaches especially if the number of images (BTF samples) is small. Consider a simple example for clarity. Figure 1 shows the two images of texture obtained with the illumination direction at $22.5°$ and $45°$. The predicted texture at $67.5°$ is obtained by a linear combination of the input textures at $L = 22.5°$ and $L = 45°$. Also shown for comparison is the actual texture at $L = 67.5°$. The inaccuracies in the predicted texture are clearly visible. Often in a static image, the inaccuracies, while visible, are not particularly disturbing. But if we consider the continuous change of illumination direction, these inaccuracies become large visual artifacts.

Figure 2 shows the cross section of a geometric structure and the cast shadow that appears for the illustrated illumination. Consider that as the illumination changes continuously, the shadow shape changes continuously. Movement of the shadow borders is an important perceptual cue that provides and impression of the 3D nature of the geometric structure, and therefore improves the appearance of the actual surface texture.

In this paper we present a new model of surface texture that incorporates both reflectance (images) and limited but useful geometric information. We recognize that while reflectance alone is not sufficient for modeling texture, a full geometry based approach is neither practical (it's hard to measure fine-scale geometry) nor sufficient (even if you have geometry you need the texture and shading for accurate modeling). Our model is a hybrid of geometry and reflectance models and has key advantages in a wide range of applications. Notably, for appearance prediction, the representation that can handle the non-linear effects of cast shadows and occlusions. While we have motivated the problem with rendering, the representation has significant implications in the areas of texture classification, point correspondences, as well as texture synthesis.

2 Method

2.1 Geometric Textons

Our method assumes that the local geometry in a surface texture consists of a finite number of geometric configurations called *geometric textons*. Furthermore, we assume that these geometric textons can be estimated using image observations. We are given N images of a surface texture corresponding to N different combinations of L, V. This image set is the sampled BTF. Standard stereo is not used in order to avoid the task of point correspondences. Also, we don't use photometric stereo because the local surface normal is not sufficient for cast shadows. Instead we estimate the local geometric structure and absolute height using a finite library of geometric primitives. This approach is motivated by the surface structure comprising a fine-scale geometric height variation that has limited range and exhibits some degree of spatial invariance. That is, local structure repeats and we expect the local structure to be well characterized by a finite set of predefined primitives. Thus our method is fundamentally different from [19], which explicitly recovers dense shape using shadow graphs, shape-from-shadow and shape-from-shading methods.

The basic schematics of our approach are illustrated in Figures 3, 4 and 5. We begin by learning a library of geometric textons. A key property of the geometric textons is that they contain information about absolute height in addition to local surface normals in a small neighborhood (e.g.

Figure 3: Example surfaces with known geometry are used to learn repeating geometric configurations which are called *geometric textons*.

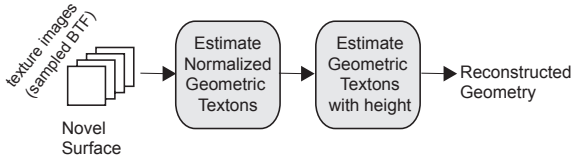

Figure 4: Estimation of the local geometric texton. The sampled BTF of a novel surface is given. Then, the normalized geometric texton is estimated. Finally, the family of geometric textons with the same surface normal configuration but different heights is then considered and the closest match is used as the estimated geometric texton. To reconstruct geometry overlapping geometric textons are averaged.

7×7). While the overall surface structure can be obtained by the surface normals, absolute height is required to recover cast shadows and occlusions. The next step is to label observations of a novel sample by using the sparsely sampled BTF to estimate the local geometric primitive at each pixel. The estimation of the local primitive is done in a two stage approach. First, primitives with surface normal information but no absolute height information are estimated. We use the term normalized geometric textons when referring to these geometric primitives. Once the normalized geometric texton has been identified, the absolute height is obtained by a comparison to the sampled BTF. After the map of labeled surface points is established, we reconstruct an estimate of the local fine-scale geometry. While this reconstruction is coarse, the key geometric structures which contribute to cast shadows and occlusions are recovered.

2.2 Library of Geometric Textons

We assume that a finite library geometric textons can be used to represent the local fine scale geometry. For example these primitives correspond to height edges, ridges, etc. We define this library of geometric textons using a set of training samples with known geometry and k-means clustering. As in appearance-based texton methods such as [9][1, 2], we use training images to obtain the library, however here we cluster on geometry information instead of intensity information. The training surfaces are a set of surfaces that exhibit sufficiently varied surface structure so that a library can be created with these examples. The training surfaces are not the surfaces of interest. Unlike the training images,

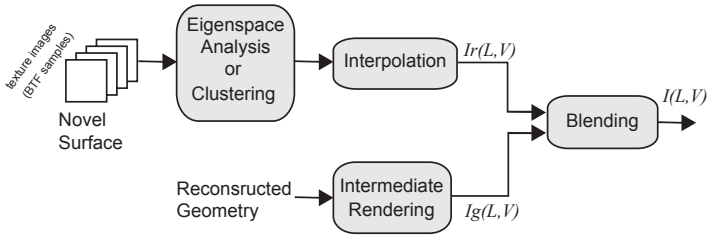

Figure 5: The sampled BTF is used to obtain a reflectance based estimate of the I_r. The reconstructed geometry is used to obtain a different estimate I_g. When L, V is not in the set of sampled imaging parameters and when the BTF sampling is sparse, the cast shadow information in I_g will be much better than I_r. A blended result $I(L, V)$ combines I_g and I_r.

the novel surface shown as the input in Figure 4 has no prior geometric information. In fact, for our results we use synthetic rough surfaces for training, generated by simulating two dimensional gaussian random fields that has a gaussian covariance function. By varying the roughness and the effective correlation length, various input surfaces can be achieved.

For clustering geometry, we consider a 7×7 patch of surface height values, where the center value is subtracted so that each patch as zero height in the center. In addition we consider the surface normal in the center of the patch. Therefore there is a 52 dimensional vector g at each surface point is given by

$$g = [h(1) \quad h(2) \quad ... \quad h(49) \quad n_x \quad n_y \quad n_z], \qquad (1)$$

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

2.3 Estimating Geometric Texton Labels

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

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

We assume no prior geometry, e.g. from stereo, laser scanning, or other methods. The hypothesis is that the local reflectance distribution observed under several viewing and illumination directions is sufficient to infer the local geometric primitive or texton.

Computationally the estimation of the local geometric texton relies on comparing the observed BTF with the prediction obtained by rendering the geometric texton. In order to render the geometric texton, we must choose a shading model and for simplicity Lambertian shading is chosen. However, any shading model, or multiple shading models,

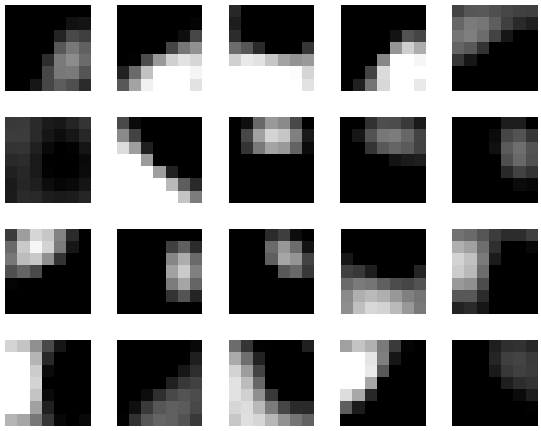

Figure 6: An illustration of the normalized geometric textons with $k = 20$.

could be employed. Also, the surface need not strictly adhere to the shading model, as long as the best match to the observed BTF is the correct geometric texton.

Let k denote the number of geometric textons. For each of the k geometric textons and for each of the N imaging parameters L_i, V_i, we render a 7×7 image. The center pixels of these images form a vector v of length N for each of the k geometric textons. At each pixel x, y, we have an N dimensional vector u from the sampled BTF $f_s(x, y)$. The correct geometric texton label l at that pixel is given by

$$l = \arg\max_j (v_j \cdot u) \quad j \in [1..k]. \tag{3}$$

At this point, the geometric textons considered have zero height at the center as described in Section 2.2. These textons are the normalized geometric textons. To get absolute height instead of relative height, we consider linear transformations of the height vector h encoded within the primitive g that best match f_s. To reconstruct the geometry we put the 7×7 region of height information in the correct position and average the overlapping regions.

2.4 Combining Reflectance and Geometry

Consider the goal of generating the full BTF $f(x, y, L, V)$ from the sampled BTF f_s. Using only reflectance information in the form of images, some type of interpolation is needed. For example, if eigenspace methods are used, then a linear combination of the basis images is required to give us $f(x, y, L, V)$. Alternatively, if 3D textons are used as in [9][10], then the appearance vectors associated with the 3D textons must be interpolated to get the correct texture image.

For an arbitrary L, V, let I_r denote the texture image obtained using only reflectance information f_s as shown in Figure 5. Let I_g be the image obtained by rendering the reconstructed geometry. Figure 5 illustrates that a combination of this reflectance and geometric information is used to obtain a prediction with accurate cast shadows. The basic idea is that I_r is a good prediction except where there are cast shadows. The location of the cast shadows can be predicted using I_g. Specifically, a binary mask M is created to indicate the location of the cast shadows. In practice, the mask is blurred to remove abrupt intensity changes. The two images I_r and I_g are blended to get the result I given by

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

This equation explicitly shows that the representation is a hybrid of reflectance and geometric methods and motivates the name hybrid texton method. Note that because I_g is primarily used for cast shadow information, the exact shading model used in the intermediate rendering step in Figure 5 is not critical.

3 Results

To illustrate how this representation can be used, we show result of predicting texture appearance. Given a sampled BTF f_s, we wish to predict the image using imaging parameters L, V, where L, V is not in the set of N imaging parameters L_i, V_i. For our results, the number of classes $k = 20$, and the number of BTF sample image $N = 9$. The height data from the k geometric textons is illustrated in Figure 6. Notice these local primitives correspond to basic structural elements.

We use texture images in the CUReT database [5] from Sample28 (crumpled paper) and Sample11(rough plaster). An example of the 9 BTF images used in the estimation is shown in Figure 7. The prediction results are shown in Figure 8, 9 and 10. In each of these results the frontal view image with $L = 67.4$ is not included in the sampled BTF. It is shown as ground truth to evaluate the quality of the rendered result. In each result, the BTF images used for interpolation to get I_r correspond to $L = 22.5^\circ$ and $L = 45^\circ$. A simple interpolation is sufficient to match the intensities everywhere except the cast shadow region. Since no similar cast shadows exist in the 9 BTF images, the cast shadows cannot be predicted by any other linear combination of these images. The figures show I_r, I_g and the final result. By comparing the rightmost images (top and bottom row) it is clear that the final result predicts the overall appearance including cast shadows.

4 Implications for Vision

While we have motivated the problem with appearance prediction, the representation has significant future implications in other areas. Texture recognition and classification may be improved with this representation especially when the training images are obtained with different imaging parameters than the testing conditions. The fundamental prob-

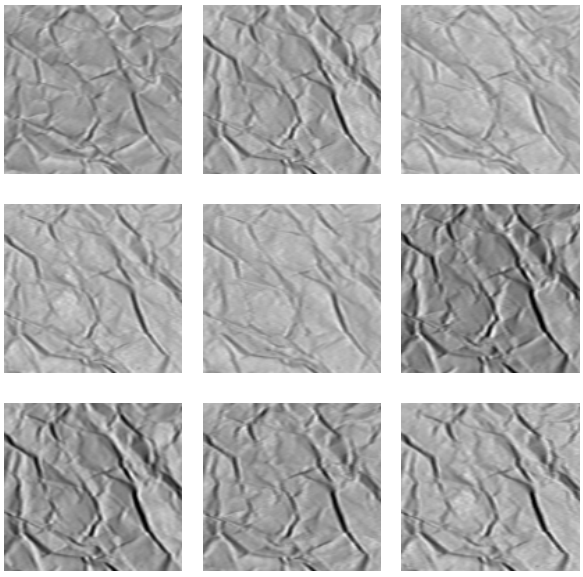

Figure 7: BTF samples for crumpled paper used in generating the result in Figure 8. For these results, $N = 9$ which means f_s consists of 9 texture images.

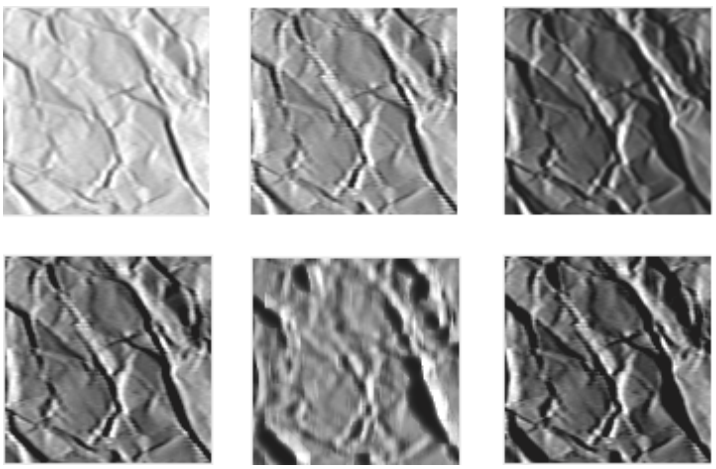

Figure 8: Predicted texture appearance using hybrid texton method using crumpled paper. Top row: texture image with $L = 22.5^\circ$ (left), $L = 45^\circ$ (center), and $L = 67.5^\circ$ (right). The viewing direction is frontal for each of these images. The sampled BTF f_s for this example did not contain the rightmost image with $L = 67.5^\circ$. The goal in this result is to predict this texture image. Bottom row: I_r (left), I_g (center), predicted image $I(L, V)$ with $L = 67.5$. Compare the rightmost image of both rows to compare ground truth (top row) with the predicted image (bottom row).

lem of point correspondences may be assisted because the underlying surface representation can predict the appearance of the corresponding region in the other image.

5 Summary and Conclusion

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

References

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

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

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

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

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

[6] J. Dong and M. Chantler. Comparison of five 3d surface texture synthesis methods. *3rd International Workshop on Texture Analysis and Synthesis (Texture 2003)*, pages 19–24, 2003.

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

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

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

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

[11] S. Magda and D. J. Kriegman. Texture synthesis on arbitrary meshes. *Proceedings of the Eurographics Symposium on Rendering*, pages 82–89, 2003.

[12] K. Nishino, Y. Sato, and K. Ikeuchi. Eigentexture method: Appearance compression and synthesis based on a 3d model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1257–1265, November 2001.

[13] M. Oren and S. K. Nayar. Generalization of the lambertian model and implications for machine vision. *International Journal of Computer Vision*, 14:227–251, 1995.

[14] S. C. Pont and J. J. Koenderink. Bidirectional texture contrast function. *Proceedings of the European Conference on Computer Vision*, pages 808–822, 2002.

[15] S. C. Pont and J. J. Koenderink. Brdf of specular surfaces with hemispherical pits. *Journal of the Optical Society of America A*, pages 2456–2466, 2002.

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

[17] B. van Ginneken, J. J. Koenderink, and K. J. Dana. Texture histograms as a function of irradiation and viewing direction. *International Journal of Computer Vision*, 31(2-3):169–184, 1999.

[18] M. Varma and A. Zisserman. Classifying images of materials. *Proceedings of the European Conference on Computer Vision*, pages 255–271, 2002.

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

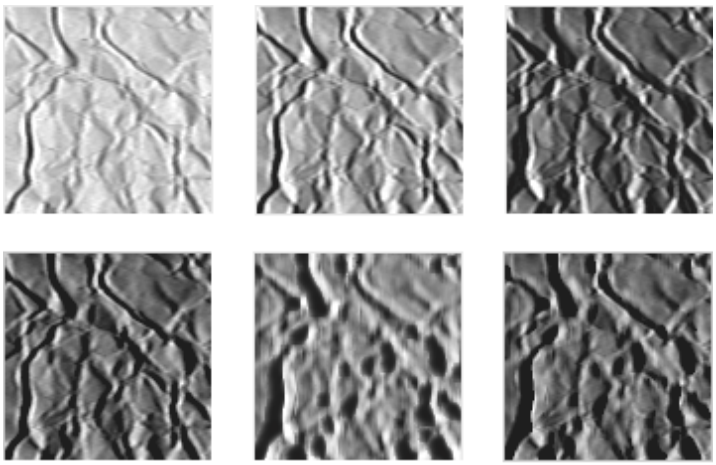

Figure 9: Predicted texture appearance using hybrid texton method using another example of crumpled paper. Top row: texture image with $L = 22.5^\circ$ (left), $L = 45^\circ$ (center), and $L = 67.5^\circ$ (right). The viewing direction is frontal for each of these images. The sampled BTF f_s for this example did not contain the rightmost image with $L = 67.5^\circ$. The goal in this result is to predict this texture image. Bottom row: I_r (left), I_g (center), predicted image $I(L, V)$ with $L = 67.5$. Compare the rightmost image of both rows to compare ground truth (top row) with the predicted image (bottom row).

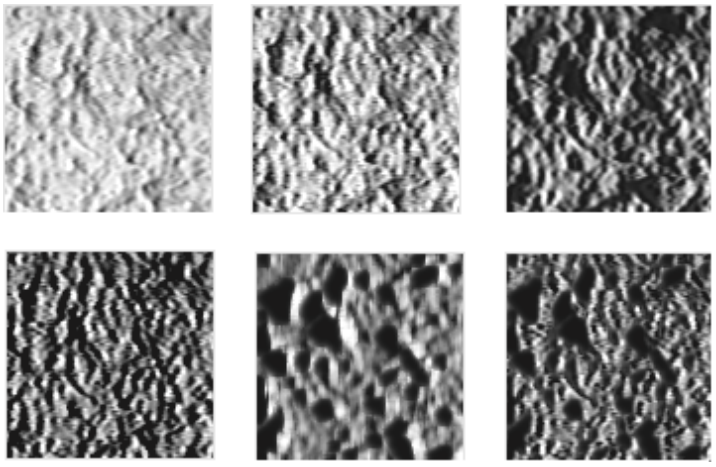

Figure 10: Predicted texture appearance using hybrid texton method using rough plaster. Top row: texture image with $L = 22.5^\circ$ (left), $L = 45^\circ$ (center), and $L = 67.5^\circ$ (right). The viewing direction is frontal for each of these images. The sampled BTF f_s for this example did not contain the rightmost image with $L = 67.5^\circ$. The goal in this result is to predict this texture image. Bottom row: I_r (left), I_g (center), predicted image $I(L, V)$ with $L = 67.5$. Compare the rightmost image of both rows to compare ground truth (top row) with the predicted image (bottom row).