Appearance Derivatives for Iso-normal Clustering of Scenes

Sanjeev J. Koppal and Srinivasa G. Narasimhan
Robotics Institute, Carnegie Mellon University, Pittsburgh, USA
Email: (koppal,srinivas)@ri.cmu.edu

Abstract

A new technique is proposed for scene analysis, called "appearance clustering". The key result of this approach is that the scene points can be clustered according to their surface normals, even when the geometry, material, and lighting are all unknown. This is achieved by analyzing an image sequence of a scene as it is illuminated by a smoothly-moving distant light source. In such a scenario, the brightness measurements at each pixel form a "continuous appearance profile". When the source path follows an unstructured trajectory (obtained, say, by smoothly hand-waving a light source), the locations of the *extrema* of the appearance profile provide a strong cue for the scene point's surface normal. Based on this observation a simple transformation of the appearance profiles and a distance metric are introduced that, together, can be used with any unsupervised clustering algorithm to obtain iso-normal clusters of a scene.

We support our algorithm empirically with comprehensive simulations of the Torrance-Sparrow and Oren-Nayar analytic BRDFs as well as experiments with 25 materials obtained from the MERL database of measured BRDFs. The method is also demonstrated on 45 examples from the CURET database, obtaining clusters on scenes with real textures such as artificial grass and ceramic tile, as well as anisotropic materials such as satin and velvet. Results of applying our algorithm to indoor and outdoor scenes containing a variety of complex geometry and materials are shown. As an example application, iso-normal clusters are used for lighting consistent texture transfer. Our algorithm is simple and does not require any complex lighting setup for data collection.

Digital Object Indentifier 10.1109/TPAMI.2008.148

1. Why Cluster Appearance?

Our world contains scenes of vastly varying appearances. These appearances depend on several different factors such as lighting, viewing geometry, material properties, and the 3D shapes of scenes. Extracting these properties from images (or image sequences) for scene analysis is an important inverse problem in vision. Unfortunately, these properties usually interact non-linearly, making their estimation from images difficult.

In order to make this problem tractable, several works have assumed prior knowledge of either lighting or scene reflectance or structure. Methods that assume known lighting include Woodham's classical photometric stereo ([1]) for lambertian scenes, as well as several extensions for non-lambertian low parameter BRDFs, such as the micro-facet model and the dichromatic model ([2],[3],[19],[4],[5],[6],[7],[8],[9],[10]). Two works that are of particular interest to us are by Healey ([19]), who segments a lambertian scene into regions that share local surface normal, and Goldman et. al., ([4]), who demonstrate that clustering of material properties enables the estimation of scene properties. Complementary to the above methods is the class of inverse rendering algorithms that estimate low parameter BRDFs and lighting ([11],[12]) using known 3D scene geometry. Ramamoorthi's thesis ([13]) provides a formal analysis of exactly when inverse rendering is possible for BRDFs and distant lighting that are represented using Spherical Harmonics. Finally, Hertzmann and Seitz ([14]) recover the geometry of objects by estimating combinations of spheres of few "basis materials" that best describe scene reflectance.

In this work, we present a novel approach for appearance analysis of static scenes consisting of a broad range of BRDFs, *without requiring any knowledge about scene geometry, material properties, lighting, or example calibration objects*[1]. Our approach involves dividing a complex scene into geometrically consistent clusters (scene points that have the same or very similar surface normals) irrespective of their material properties and lighting. Such a cue is useful since the number of unknowns related to scene geometry is reduced within each *iso-normal* cluster, leading to a more robust estimation of scene properties. The camera observing the scene is assumed to be orthographic. As the source moves, observations at each scene point over time result in a *continuous appearance profile* (see Fig. 1). The smoothness (continuity) of the

[1]In previous work, this was achieved only for simple BRDF models such as lambertian or Torrance-Sparrow ([15],[16],[17],[18]).

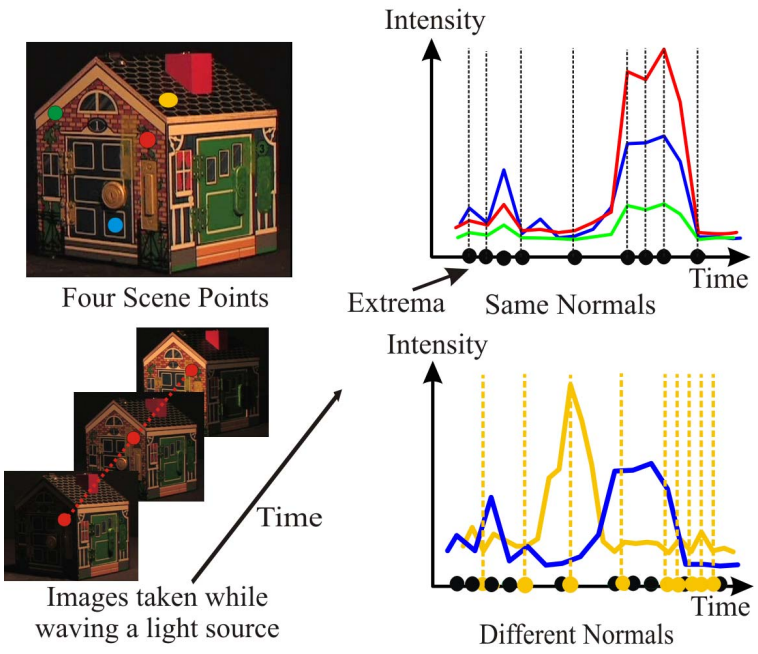

Fig. 1. **Appearance Profile and Extrema:** Here we show an image sequence obtained by illuminating a static scene with a moving, distant source. The appearance profile of a scene point is the observed intensities at a single pixel over time. The appearance profiles show several extrema (peaks and valleys) as illustrated on the right. More often than not, scene points with the same surface normal exhibit extrema at the same time instances. Similarly, most extrema locations do not match for scene points with different normals. This makes extrema locations excellent features for our clustering algorithm.

appearance profile is a strong notion that has not yet been fully exploited in computer vision[2]. In particular, we will show that the information contained in the derivatives of these profiles (specifically, when the derivatives are zero and extrema occur) is related to the surface normal of a scene point. Intuitively, if cast shadows are ignored, scene points with the same surface normal should 'light up' and 'go dark' at the same time. More precisely, this means that the foreshortening at a scene point creates brightness maxima and minima in the appearance profile. The idea of using foreshortening extrema for clustering is related to the notion of *orientation consistency* first proposed by Hertzmann and Seitz ([14]) and used to compute the surface normal at a scene point by comparing it with an "example" object with known shape and BRDF. In contrast, this method computes orientation consistencies *between* scene points of unknown

[2]Exceptions include space-time stereo [20] and the work of Hayakawa ([21]) that uses an arbitrary moving light source to alleviate the ambiguity in photometric stereo for Lambertian objects.

normals and BRDFs, without requiring any example object. The trade-off here is that a longer sequence of images is required than if example objects were available.

Our idea of exploiting brightness extrema as a feature for iso-normal clustering is supported by empirical evidence from both simulations and real data. In the first of these experiments surface normals were sampled uniformly on the hemisphere of directions and appearance profiles were created with the Torrance-Sparrow and Oren-Nayar analytic models for BRDFs ([3],[5]). We observed that extrema in iso-normal profiles that are *shared* (occur simultaneously) were invariant to material variations in the models, such as changes in albedo or surface roughness. A similar result was obtained for rendered scenes, using a well-known ray-tracing method ([32]), again using the Torrance-Sparrow and Oren-Nayar models. In addition, iso-normal profiles created from real measurements of 25 BRDFs in the MERL database ([34]) are shown to share their extrema. Finally, experiments were performed with real textures made of anisotropic materials, such as velvet and satin. Even in such scenarios, those profiles that were generated from scene points with the same surface normal share extrema locations in time.

Once a profile's intensity maxima and minima are detected, a technique is needed to compare profiles to obtain iso-normal clusters. This is addressed by introducing a transformation to the appearance profile that linearly interpolates between extrema locations. Those profiles which share all their extrema locations become identical after the transformation. The transformed profiles can be used with *any supervised or unsupervised clustering technique* to obtain robust geometrically consistent scene clusters. In this paper, the simple clustering technique of k-means is used for all our results. We expect that more sophisticated methods would produce even better iso-normal groupings. Our algorithm uses the dot-product distance metric employed previously ([28]) to match profiles with shared extrema. Finally, our method requires that the user just hand-wave a light source and therefore image acquisition is not difficult. This is in contrast to methods that require complex illumination setups ([23]).

Clustering results are shown for scenes with textures from the CURET database and demonstrate that our method is able to cluster a variety of textures, such as artificial grass and straw, as well as anisotropic materials such as satin, fur, velvet, and metal paper. Our method produces valid iso-normal sub-clusters for scenes with complex geometry containing curved surfaces and shadows. We demonstrate clustering both on indoor scenes with everyday objects, such as tables and chairs, and a difficult outdoor scene from the WILD database. An application of our method in graphics is demonstrated by transferring profiles between scene points in iso-normal clusters. This allows texture transfer that is consistent across all the lighting changes in the given image sequence.

In summary, this paper presents a simple technique for creating iso-normal clusters for complex scenes. Such a strong geometric cue can be used to robustly estimate scene properties such as materials, lighting and geometry. Therefore, we believe our method has broad significance for vision and graphics.

2. Appearance Profiles and their Extrema

Consider images of a static scene illuminated by a smoothly-moving, distant light source. An appearance profile is a vector of intensities measured at a pixel over time, as illustrated in Figure 1. Direct clustering of these profiles fails to produce iso-normal groupings, even after normalizing for scale and offset in brightness. This is because the intensities are non-linear functions of geometric and material properties of a scene point. To create iso-normal clusters, it is critical to obtain a feature from the appearance profile that is invariant (or at least insensitive) to material properties. In this paper, such a feature is obtained by exploiting the continuity (smoothness) of an appearance profile, which yields information about the derivatives of the BRDF with respect to time. Our key idea is to detect *brightness extrema* (peaks and troughs), where the first order time derivatives of the appearance profile are zero. Extrema are said to be *shared* between two appearance profiles if they occur at the same time instance in both profiles. In this section, strong empirical evidence consisting of simulations and real data support shared extrema as a feature for iso-normal clustering.

An important factor that determines where extrema occur in a profile is the path of the light source. Consider a distant point light source that is being waved by a user. The trajectory of the light source is *unstructured* and it contrasts with the light source paths created by gantry setups

used in many previous works (such as [23]). While a structured path may have a regular shape, such as a spiral (see Figure 2), an unstructured path created by 'hand-waving' the light source may not have any standard, recognizable shape. Since the light source changes its position smoothly, but randomly, at every time-step the intensities at every scene point are generated stochastically. The empirical evidence in this section will show that iso-normal profiles produced by such unstructured paths share many extrema.

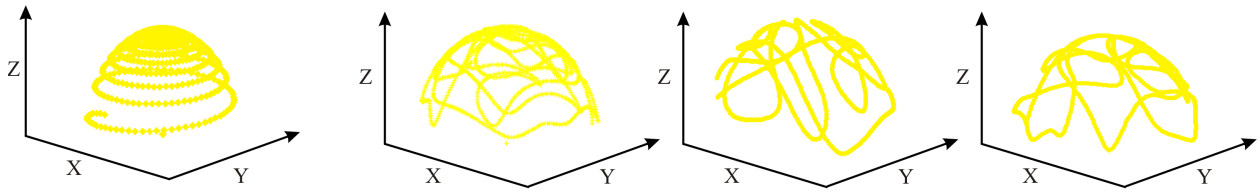

Fig. 2. **Structured versus Unstructured Paths for the Light Source.** Engineered setups such as a gantry may produce paths such as the spiral shape shown on the left. In contrast, in our method the user hand-waves the light source. This results in unstructured paths, such as the three shown on the right.

2.1 Shared Extrema in Iso-normal Profiles

Intuitively, shared brightness extrema are important since scene points that have the same normal should 'light up' and 'go dark' at the same time. We investigate the extrema present in profiles generated by BRDF simulations, rendered scenes, measured BRDFs and real textures. These experiments provide strong empirical evidence that shared extrema locations can be exploited to obtain iso-normal clusters.

BRDF Simulations: In Figure 3 profiles are generated for 50 unique surface normals that were sampled uniformly from the hemisphere of directions. The simulations consist of four BRDF models: Lambertian, Oren-Nayar, Torrance-Sparrow and Oren-Nayar + Torrance-Sparrow. The user creates a smooth, unstructured path for the light source on the hemisphere of directions. The material properties were varied (roughness σ from 0 to 1 and albedo ρ from 0 to 1) for each of these models producing over 20,000 profiles for each normal. The top row of Figure 3 shows some of these profiles for a particular surface normal which demonstrate significant variation

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

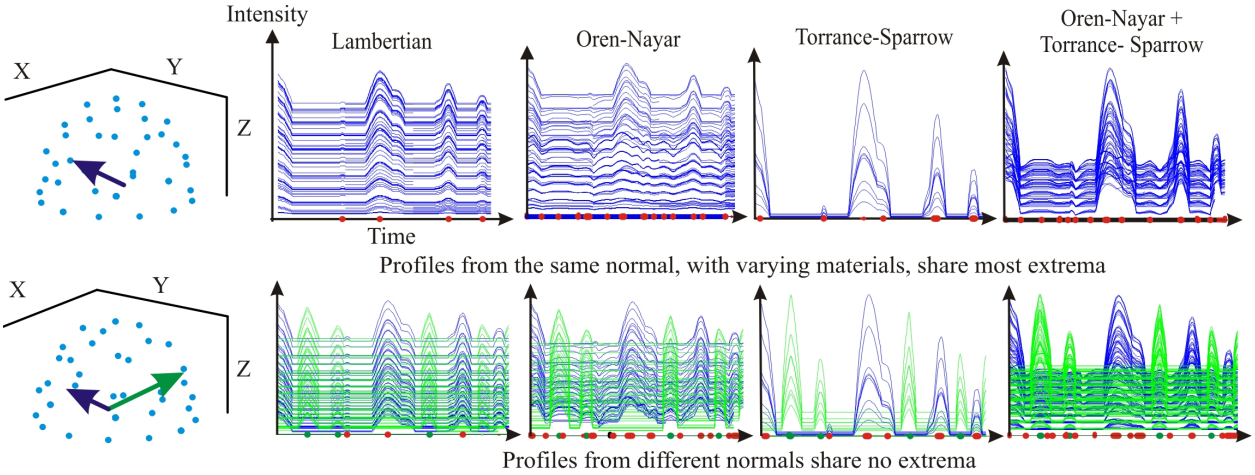

Fig. 3. **Simulations showing the link between Extrema and Surface Normal:** Appearance profiles are simulated for four BRDFs over a range of 20,000 material properties (only a few are shown for clarity). Profiles are only shown for two normals, although we simulated profiles for 50 (marked by blue dots on the hemisphere, on the left). In the graphs above, the extrema location of a profile is marked on the time axis by a colored dot. Note that profiles from the same local normal (top row) share most of the extrema locations, whereas profiles from different normals (bottom row) do not.

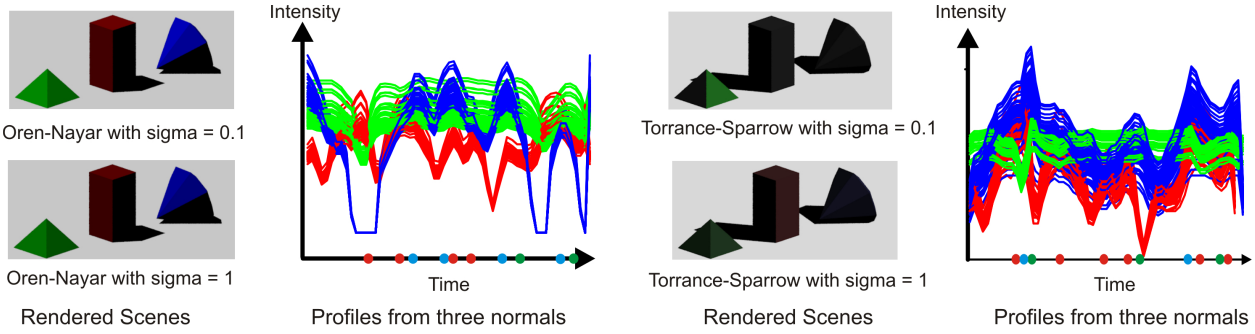

Fig. 4. **Profiles from iso-planar regions of rendered scenes show the link between Extrema and Surface Normal:** Three piecewise planar shapes are rendered using both the Oren-Nayar general lambertian model, as well as the Torrance-Sparrow off-specular model under smoothly moving distance light source. Varying the roughness parameters σ in both models produced 2,000 profiles, of which we show a few here. Profiles from three normals in the rendered images are shown, and iso-normal profiles have the same color. These experiments indicate that iso-normal profiles share the same extrema. The common maxima for each of the three normals is marked using the same color as the profiles.

in the profile shape due to changing material properties. Despite these differences extrema locations were found that were common to over 95% of a normal's profiles and therefore insensitive to the changes in albedo and roughness. Furthermore, these shared extrema locations were unique to a particular normal. In the bottom row of Figure 3, profiles from two different normals are shown, with their shared extrema locations marked on the x-axis.

Rendered Scene: A scene was rendered using a commonly used ray-tracing tool ([32]) generating profiles under conditions similar to real scenes, with effects such as cast shadows and interreflections. In Figure 4 such a scene is shown consisting of three piecewise planar objects: a pyramid, a box, and a diamond. This was rendered using the Oren-Nayar and Torrance-Sparrow models whose material properties were varied (roughness σ from 0 to 1 and albedo ρ from 0 to 1) to create 100 instances of the scene, two of which are shown in Figure 4. A light source moving in an unstructured path was simulated, producing 40 renderings of each such scene instance. There are nine unique normals in our scene and, unlike the previous scenario, each normal was associated with at least 2,000 profiles. On the left of Figure 4, four images of the rendered scene are shown with different material properties. When the objects are rendered with the Oren-Nayar model, increased roughness makes the objects appear flatter and darker. Similarly sharp highlights in the Torrance-Sparrow model, such as on the right facet of the green pyramid, become blurred as roughness increases. On the right of Figure 4 profiles from three different normals are plotted and, for clarity, only the maxima locations are marked on the x-axis. Even though all the profiles vary significantly with change in material properties, the iso-normal profiles share the same extrema.

Measured BRDFs: The two previous experiments dealt with profiles generated from artificial scenes. Next, 25 real materials were selected from a BRDF database measured by Matusik et al ([34]). To create appearance profiles from these measured BRDFs, the light source path was simulated by the user, similar to the previous cases. For each material, the hemisphere of directions was sampled uniformly to create 25 unique normals. Each normal was used to generate profiles whose material properties corresponded to the measured BRDFs. For each of these normals, over 90% of the profiles shared their extrema, and in Figure 5 we show 5 such profiles from two such different normals. These experiments demonstrate that the extrema feature has significance beyond the commonly used BRDF models.

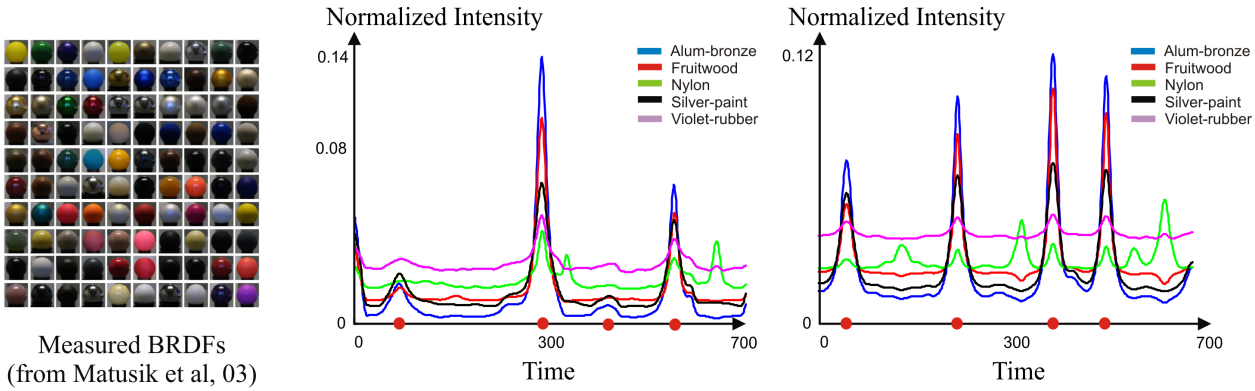

Fig. 5. **Profiles from real, measured BRDFs:** Matusik et al ([34]) measured a large number of real world materials by using spheres, as shown on the left in a figure reproduced from their paper. Profiles were created from these measurements by simulating the path of a hand-waved light source and these are shown for two different surface normals. In each graph, the profiles for five materials share extrema locations (given by red dots) and these locations are different for the two normals.

Fig. 6. **Real profiles from different materials show the link between Extrema and Surface Normal:** Six experiments are conducted by placing textures at different orientations with respect to the camera. Note that the appearance of the different patches vary greatly, even though they are all flat on the board. Some profiles are shown from each of the textures in a color-coded fashion. Even though these materials are anisotropic (satin and velvet), their profiles still share extrema, which are marked by black dots on the x-axis.

Real Textures: In addition to the above simulations, real experiments were conducted with four anisotropic textures: silk, crepe satin, mink velvet, and royal velvet. These materials cannot be modeled by the Oren-Nayar and Torrance-Sparrow BRDFs. The textures were attached on a planar board so that they all have the same surface normal, as shown in Figure 6. The profiles from these textures were measured for six different orientations of the board, and some of these are shown in the figure. These materials have complex appearance effects; for example, the maroon mink velvet exhibits strong vertical specularities which change in position and width as the orientation of the board is changed. Even though the profiles are drastically different from each other, they still share some extrema because they have the same surface normal.

In conclusion, we have presented four sets of experiments providing strong empirical evidence that shared extrema can be used as a material-invariant cue for a scene point's local surface normal. In the next section, an algorithm is presented that exploits this feature to produce iso-normal clusters.

3. Algorithm to Create Iso-normal Clusters

Our algorithm is extremely simple to implement and is summarized in Table 1. It has four steps, **a)** collect images of a scene by simply waving a source smoothly and randomly while detecting brightness extrema, **b)** transform the appearance profiles (Figure 7) and **c)** use a common similarity metric to **d)** cluster the scene. Any number of sophisticated learning techniques (such unsupervised, semi-supervised or supervised methods) can be used for the clustering part of our algorithm. In this section, we will discuss the transformation and metric which are powerful enough to allow the relatively simple k-means algorithm to produce accurate results. A useful heuristic for deciding the number of clusters to input to the algorithm is also provided.

Table 1

Step 1 (Input):

While acquiring frames by randomly waving a light,
Detect intensity extrema at each pixel and
store their occurrences in time.
(No need to store whole image sequence)
end

Step 2 (Transformation):

Construct a feature vector from each scene point's profile by piece wise linear interpolation of its extrema stored in Step 1 (Figure 7).

Step 3 (Metric):

Compute distance metric between (normalized) feature vectors $\vec{a}$ and $\vec{b}$ using dot-product: Distance = 1 - $\vec{a}^t\vec{b}$.

Step 4 (Output):

Cluster the normals based on the metric in Step 3.

Table 1. **Algorithm:** Our method is simple to implement. The input to the algorithm is a sequence of scene images that are collected by hand-waving a light source. At each pixel, only the locations of all the brightness maxima and minima are stored. We then linearly interpolate these extrema locations as shown in Figure 7. Therefore, each pixel location is associated with a transformed profile. These profiles are then grouped using the dot-product dissimilarity metric with any clustering algorithm, such as k-means or hierarchical clustering.

Transformation applied to profiles: From the discussion in the previous sections, it may appear obvious that extrema locations in a profile should be directly used for clustering. However, in real-world scenarios the extrema location is sensitive to noise. In addition, it is not clear how to compare profiles with different numbers of extrema. These issues are solved by a transformation of the appearance profiles that involves linearly interpolating between extrema locations. The new, transformed profile consists entirely of line segments, as shown in Figure 7. The slope of each line segment is the sign of the profile's derivative in that segment: it is either $+1$ or -1. Since profiles sharing all their extrema have identical sign for their first derivatives, after transformation such profiles will become identical. There is no difficulty in comparing profiles with different numbers of extrema since all the transformed profiles have the same length. In addition, the transformed profile can be recreated just from extrema locations and the whole profile need not be stored in memory. Moreover, detecting an extrema requires deciding if the profile brightness values are increasing or decreasing in a small window of time. This means

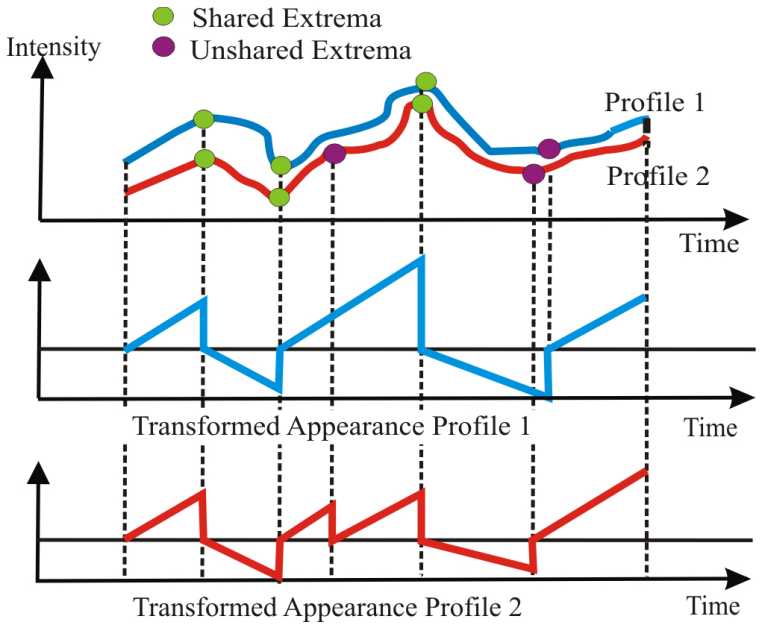

Fig. 7. **Transformation Applied to Appearance Profiles:** This illustration shows the effect of transformation on two hypothetical appearance profiles. Consider the 'segments' between extrema. The slope of transformed profile is the sign of the first derivative of a segment. Therefore two segments that have positive first derivative (monotonically increasing) get the same positive slope of 1. Note that in segments where there are no unshared extrema, the transformed values are identical.

the whole sequence of images need not be stored in memory and, in theory, our algorithm can run in real time by processing only a sliding window of fixed size and storing only the extrema locations for each profile.

Distance Metric between Profiles: To cluster the transformed profiles a distance metric must be specified. We use the "dot-product" metric which has been shown to accurately match extrema locations of profiles ([28]). Mathematically, if A and B are the transformed appearance profiles of two scene points, the "dot-product" metric is simply $1 - a^T\ b$, where a and b are the unit vectors obtained by normalizing the profiles A and B. We also analyzed and compared clustering accuracy of the dot-product metric with the Euclidean metric using two common unsupervised clustering methods, k-means and hierarchical clustering (Figure 8), for a simple, table-top piecewise planar scene. The factor of improvement achieved with varying numbers of clusters and numbers of extrema used in clustering is plotted in the figure. In all cases, our metric shows significant improvement.

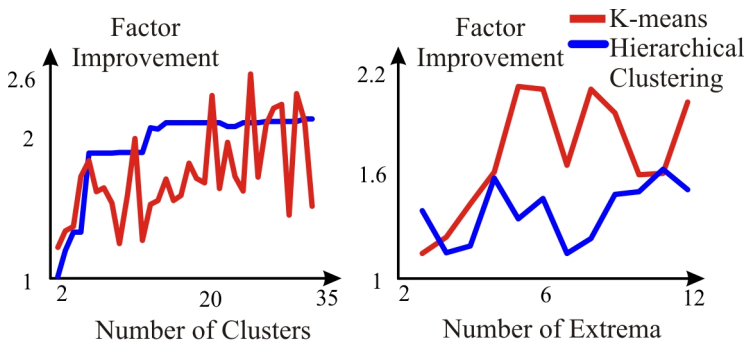

Fig. 8. **K-means versus Hierarchical Clustering.** The clustering results for a simple piecewise planar scene whose ground truth can be obtained easily are analyzed. The factor of improvement is the ratio of the clustering accuracy using our metric to the clustering accuracy using the Euclidean metric. The left hand side plot shows the factor of improvement variation with increase in number of clusters. As expected, this graph plateaus due to over clustering. The right hand side plot shows the factor of improvement obtained for different number of extrema used in the appearance profiles. Note that the k-means graph is jagged because initialization is non-deterministic. In both cases, our metric performs significantly better than the Euclidean metric.

Finally, one of the parameters that needs to be decided by our algorithm is the number of clusters, k. Calculating the number of clusters automatically is an unsolved problem in machine learning. It may be possible to use domain specific knowledge about appearance profiles to calculate the right k for our clustering algorithm, but that would require some knowledge of scene properties. Instead we advocate a well-known and simple method to decide k called hierarchical agglomerative clustering, which involves merging clusters. First, over-cluster the scene with a large k. In successive steps, clusters are merged if their distance is less than some user-defined threshold. However, there are many cases, such as smoothly moving cast shadows and curved surfaces, for which there are an infinite number of valid clusters in the scene. Clustering simply gives a piecewise approximation, and the best choice of a k for the algorithm is hard to obtain in such cases. In these scenarios we suggest the user input a reasonably large value of k. Although this over-clusters the scene, it makes sure that the sub-clusters produced are consistently iso-normal.

4. Experiments with Real Scenes

We will now demonstrate our algorithm using a wide range of real indoor and outdoor scenes with complex scene structure and material properties. Our setup consists of a Canon XL2 digital video camera observing a static scene as shown in Figure 9. As discussed before, number of clusters is decided using a simple merging technique, For example, in Figure 10 we cluster a painted house model where the number of clusters, k, was automatically selected when the distance between clusters became greater than a user-defined threshold (which was 0.5 in this experiment).

Fig. 9. Our acquisition setup with a Canon XL2 video camera, a 60 watt light attached to a wand. In real experiments the camera and light source are further away to satisfy orthographic assumptions.

Our algorithm was first tested on piecewise planar scenes consisting of real textures from the CURET database ([29]). The CURET patches are arranged in a scene and light source

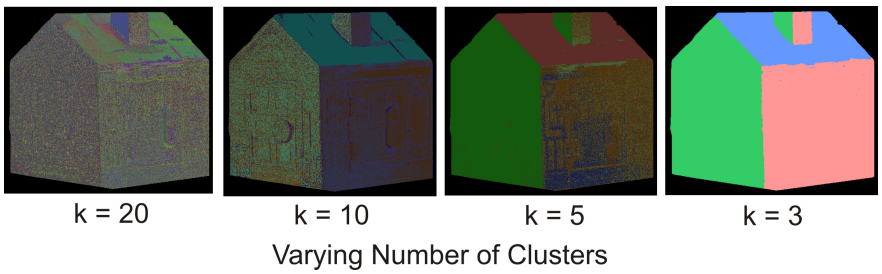

Fig. 10. **Choosing the number of clusters:** Here the number of clusters, k, are varied for a particular scene. In each case the clusters created are still iso-normal. Merging the clusters at each step allows us to create bigger iso-normal clusters. The merging is stopped when the distance between cluster centers is larger than a user-defined threshold.

is waved, creating 200 video frames. Note the boxed regions at the top of Figure 11. These textures are artificial grass and real straw, whose appearances are quite complex; for example, artificial grass has many specularities and is also rough. The second row shows steel wool and real wool, which are 3D textures with uneven height from the patch surface. Our algorithm clustered all these textures properly, even though their material properties were very different. In Figure 12, our algorithm clusters anisotropic materials, such as velvet, satin, shiny paper and fur, implying that our method works even in some cases that are not described well by our illumination model. Results are also shown for non-planar objects which contain an infinite number of normals. In these cases, our method creates a piecewise planar approximation of the continuous curved surface. For example, in the cylinder in the figure, the clusters are thin vertical stripes corresponding to regions of similar surface normal.

In Figure 13 we show more complex planar scenes, containing occlusion, cast shadowing and inter-reflection. In these regions, our method may over-cluster the scene, but note that the smaller clusters are still geometrically consistent. Clustering was also demonstrated on some everyday, indoor scenes such as the chair and table shown in Figure 14. Even though these are non-lambertian scenes with materials such as wood, plastic, metal and smooth tile, our algorithm creates meaningful clusters. In Figure 15, clustering results obtained for outdoor images of a scene collected from the WILD database ([30]) are shown. Note that this scene does not satisfy many of our assumptions. For example, it is illuminated by the sun and sky instead of a randomly-moving point light source. There is also significant depth in the scene (see [30]), violating the orthographic assumption. A good result is still obtained because the diverse and random illumination due to weather (sunny, cloudy, fog, mist) creates appearance profiles with enough intensity variation to produce valid clusters.

We believe iso-normal clusters will enable a variety of applications in vision and graphics. One such application is transferring texture in videos. The challenge here is to transfer appearance that is consistent under varying illumination. Profiles within a cluster share the same intensity extrema and therefore the corresponding scene points 'light up' and 'go dark' together. Transferring profiles within a cluster creates new pixels whose brightness varies *consistently*, as shown in Figure 16. The complex appearance effects of the materials are preserved through the length of this video sequence.

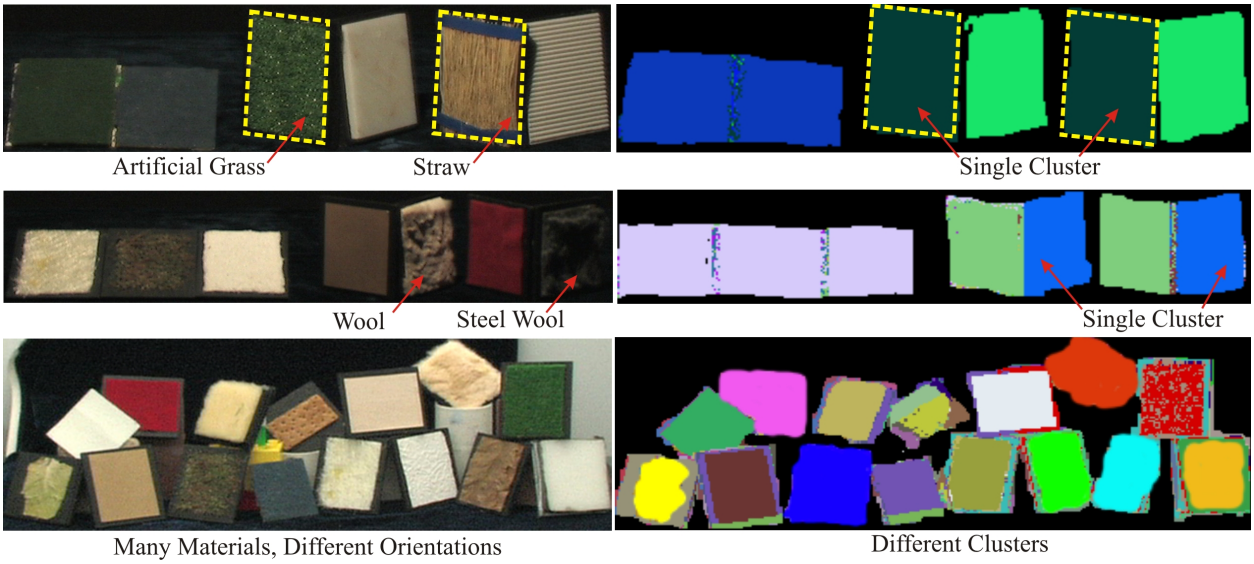

Fig. 11. **Results obtained when our algorithm is used to cluster materials in the CURET Database.** Image sequences of real CURET textures were obtained by waving a light source (We did not use the still images distributed by Columbia University). Notice the top row containing materials such as artificial grass and straw and the middle row with examples of real wool and steel wool. Despite significant appearance differences, these samples cluster together accurately because they share the same surface normal. Please see video at [33] for better visualization.

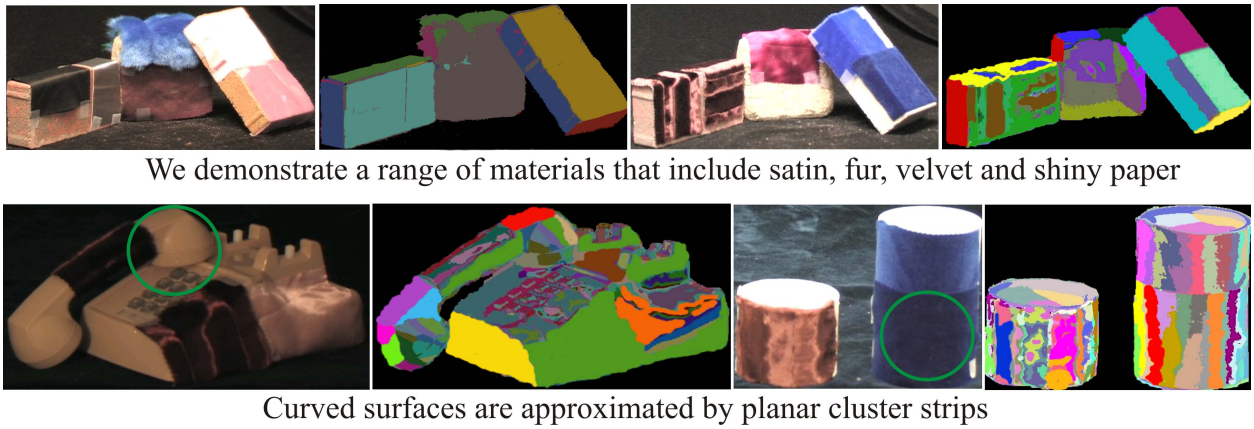

Fig. 12. **Clustering curved surfaces with complex (possibly anisotropic) materials** When anisotropic BRDFs are present in the scenes, our method still produces meaningful clusters. Furthermore, for curved surfaces, our method produces a piecewise planar approximation. Please see video at [33] for better visualization.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

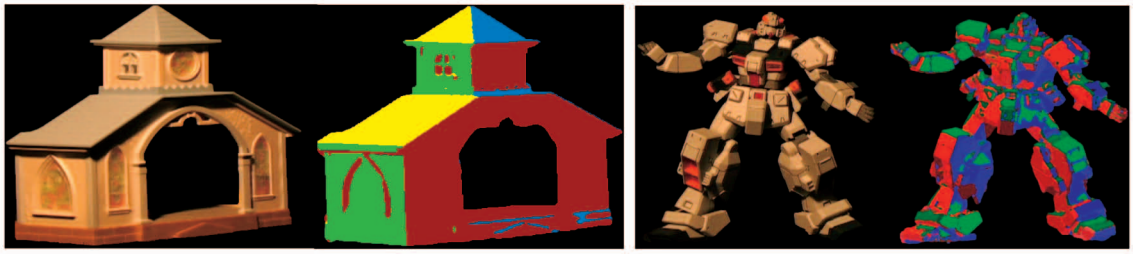

Fig. 13. **Clustering surfaces with cast shadows.** When complex effects such as cast shadows and inter-reflections are present in the scenes, our method works for simpler scenes such as on the left. For more complex scenes, such as on the right, our method may fail to group all pixels in the scene that have the same normal. Instead, the algorithm simply over clusters the scene into smaller iso-normal clusters.

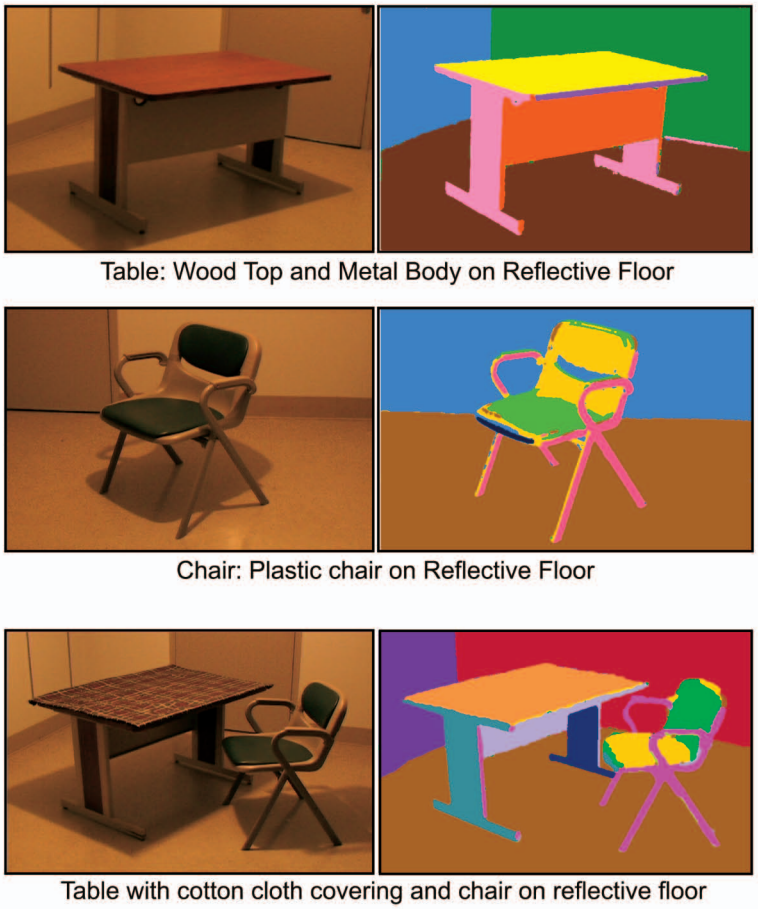

Fig. 14. **Clustering Indoor Scenes.** The following are three indoor scenes containing non-lambertian objects, such as the metal table with a wood top, a metal door, reflective floor tile, plastic chairs and a texture cotton cloth placed on the table. In spite of all of these our clustering algorithm does well. Note that some of the clusters have been merged by the user for presentation purposes only. Please see video at [33] for better visualization.

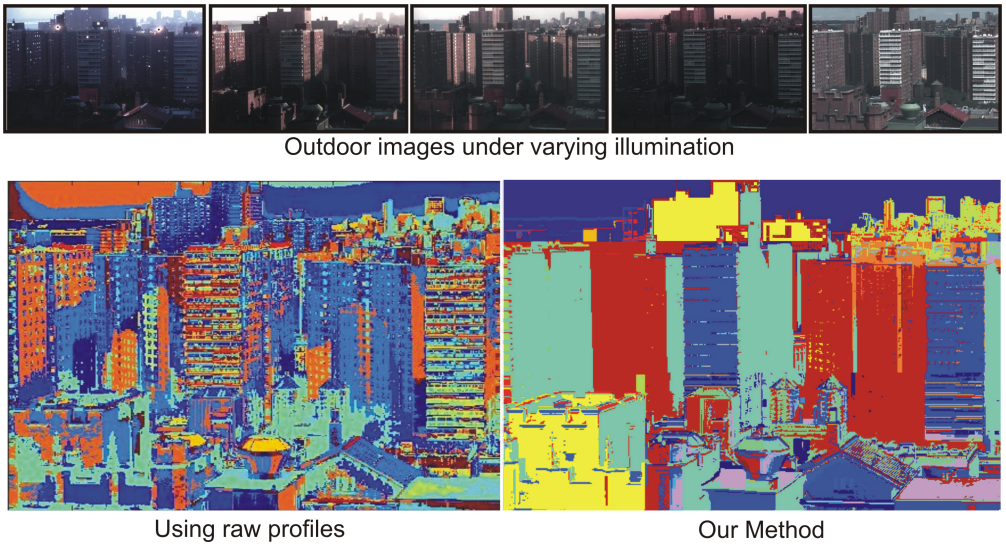

Fig. 15. **Clustering WILD Database:** Note the complex appearance effects that occur in this data set. Our transformation of the appearance profile and the dot-product distance metric does significantly better than using Euclidean distance metric on raw profiles. In both cases, k-means was used to cluster appearance profiles. Note: some sub-clusters were merged for better viewing only. Please see video at [33] for the variation in appearances in the input image sequence.

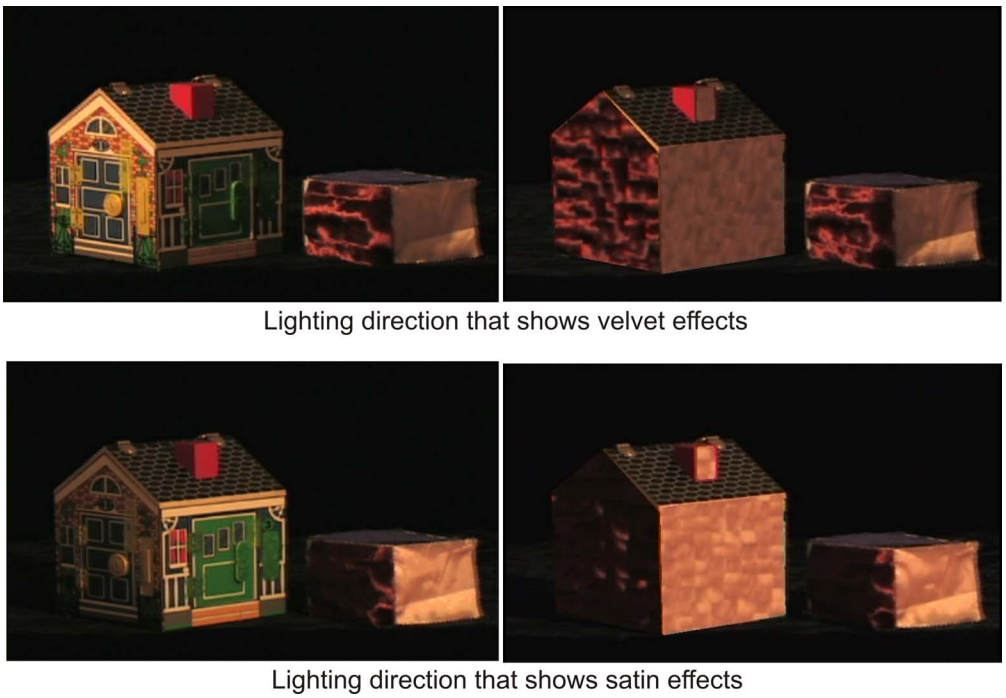

Fig. 16. **Texture transfer** of complex materials (such as velvet and satin) between similar surface normals in a scene. A patch of the original scene is chosen by the user and a simple repetitive texture synthesis method is used to transfer this patch onto other areas of the scene with the same surface normal. Note the consistency in geometry and lighting in the transferred regions. Please see video at [33] for many more lighting variations.

5. Discussion

The clustering algorithm based on brightness extrema that is presented in this paper produces results for many real-world surfaces. However analytically proving that our method will work for scenes with arbitrary materials and geometry is difficult. In the discussion below, we will both describe some of our algorithm's limitations, as well as suggest certain heuristics supported by empirical evidence.

Number of Extrema: We have shown empirical evidence linking shared extrema in profiles and surface normal. This raises the question as to what would be the minimum number of extrema needed to properly cluster a scene, and the related number of frames required in the input video. This is difficult to calculate because the number of extrema needed depends both on the geometry of the scene and its material properties. For example, in the first six rows of Figure 17, planes consisting of very dissimilar materials are placed at different orientations. Cross-clustering between the two planes becomes less likely if the materials are different and larger angles between the planes allow easier disambiguation. In contrast, in the last four rows of Figure 17, more extrema are needed since the materials are identical. These conflicting factors of material and geometry properties make it difficult to say exactly how many extrema will be needed, especially since both these factors are unknown to our algorithm. Although we do not address this issue here, we propose a heuristic that takes advantage of the fact that the user can interactively create profiles by controlling the light source trajectory. Consider a situation where the user is aware of the approximate range of normals in the scene. The light source could be moved in a way that crosses directly over these surface normals (such that foreshortening is maximum) at different times, creating extrema that are picked up by the clustering algorithm. In a similar way, waving the light source over the normal of a region with difficult material properties could result in needing fewer frames to create iso-normal clusters.

Orthographic Projection: Another assumption that we have not relaxed is that of orthographic projection. If the scene has significant depth, then this is violated and therefore surfaces that have the same local normal, but are at great distances from each other, may cluster separately. However, in a manner similar to the case of cast shadows, the solution is to overcluster the scene. The important point to note here is that each of the separate clusters created are still iso-normal.

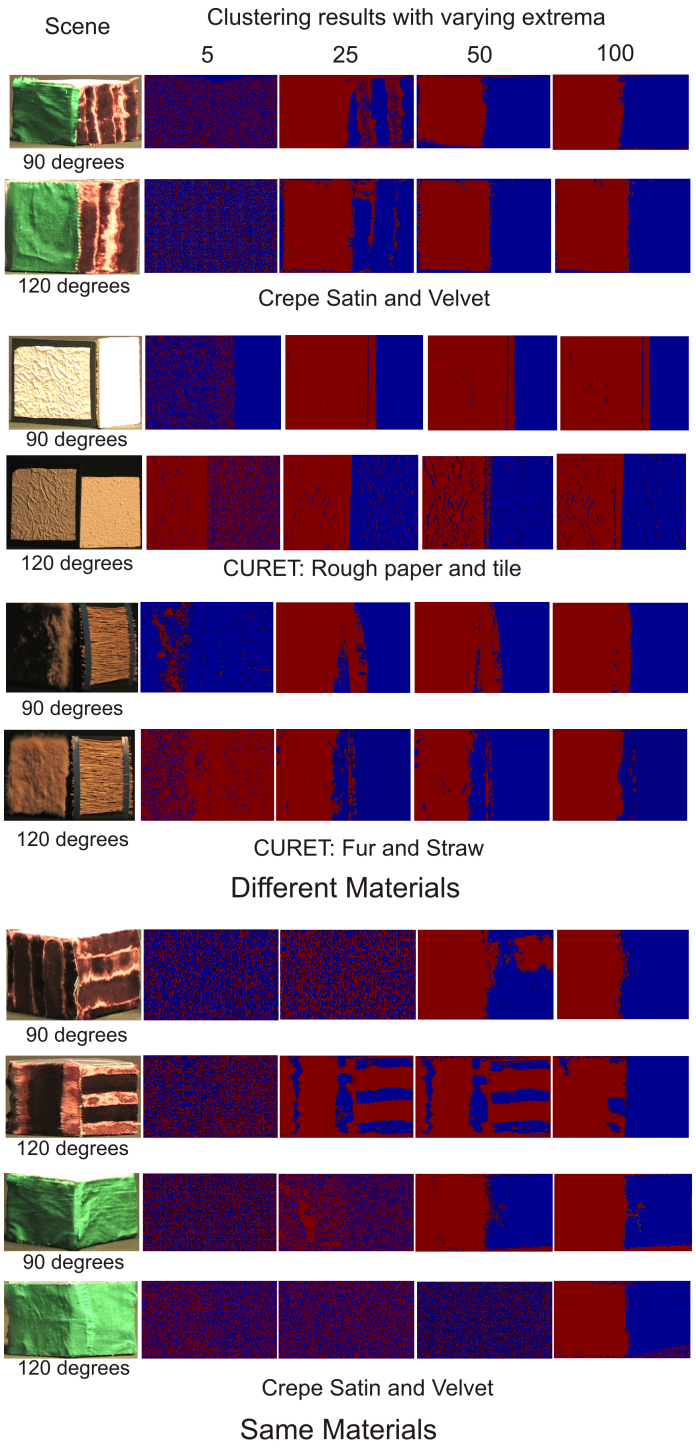

Fig. 17. **Material vs. Geometry Variation:** We analyze how the clustering result changes as the number of shared extrema are increased, in a scene with two planes. In all cases, increasing the shared extrema creates better clustering. In the first six rows, drastically different materials allow easy disambiguation of the two planes. Fewer extrema are needed to disambiguate the two planes, irrespective of the angle between the planes. In contrast, in the last four rows, more extrema are needed as the angles between the plane increases since the materials are identical.

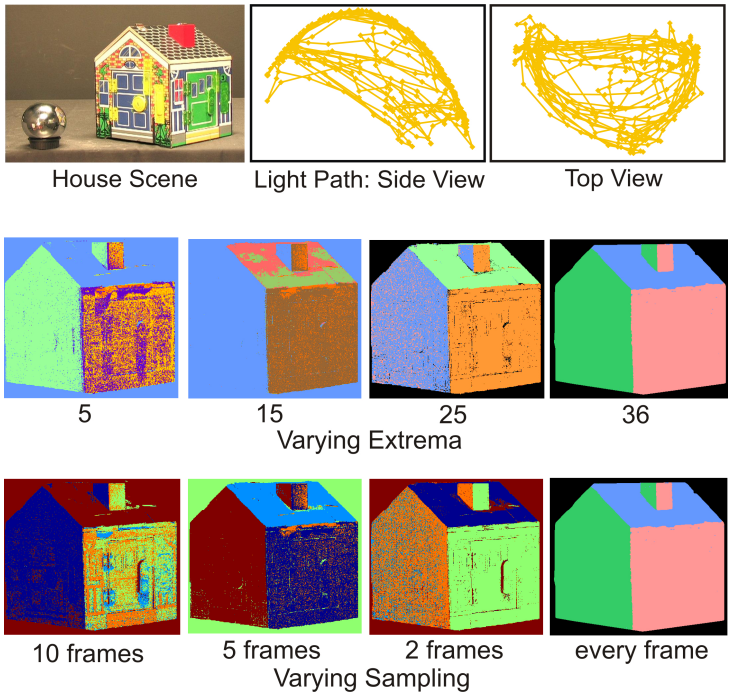

Fig. 18. **Varying some aspects of our algorithm:** Here we show the variation in the quality of results as we change different parameters. The top row illustrates the scene, as well as two views of the light source path on the unit sphere. In the second row, the length of the video is changed to include progressively increasing numbers of extrema. As the numbers of extrema increase the resulting clusters become more iso-normally consistent. In the last row, we show how the clustering results change as the sampling of the profiles changes. Note that iso-normal clusters are produced even while sampling every fifth frame. This shows that even if we do not have the exact extrema locations in the profile, the rough estimate given by sub-sampling is good enough to create iso-normal clusters.

Sampling Rate: We also need to address the issue of sampling in our appearance profiles. We have used the continuity of the smoothly moving light source to capture extrema locations. However, in reality, incident intensity is measured through discrete frames of the camera and, therefore, we can only obtain a sampled version of the actual appearance profile. What is the minimum sampling rate of the appearance profile such that clustering still gets valid iso-normal clusters? Consider a scene illuminated by a light source waved by a user, as in Figure 18. In the last row the sampling of the profile is slowly increased, and the results of clustering at each step is shown. Although at very low sampling rate the clustering breaks down, it remains robust for all other sampling rates. This is not surprising once we recall that our intuition for using brightness extrema as an iso-normal feature was that foreshortening makes scene points

'light up' or 'go dark'. If the path of our light source is smooth, then the foreshortening term, $\vec{n}.\vec{s}(t)$, is low frequency and, therefore, many samples are not required to capture its effect in the appearance profile. To calculate the exact minimum sampling frequency, the path of the light source as well as the scene geometry are needed, both of which are unknown to our algorithm. Since the foreshortening component has low frequency, we suggest collecting data using a video camera with frame rate above 30 fps. In practice, we have seen that this is sufficient to enable proper sampling of the profiles.

6. Conclusions

Recovering all aspects of a physics-based appearance model is difficult when scene properties such as lighting, geometry and materials are unknown. Unsupervised machine learning methods are very useful in vision when the sole input are images of the scene. We believe our method is novel because it brings together the effectiveness of an unsupervised learning algorithm with the physical meaning of an appearance model. The key insight was to use the continuity of the light source to extract information about the scene geometry using both a clustering algorithm and an appearance model. In summary, we have described how the derivatives of appearance (encoded as extrema locations) are related to scene geometry. We demonstrated an algorithm to exploit these extrema to create iso-normal clusters of a scene and to use these clusters for effective scene analysis. Our algorithm has no prior information about geometry, material or light sources. Although there are significant areas that require future work, we believe our method holds promise for several applications in vision and graphics.

7. Acknowledgements

This research was supported by NSF Awards #CCF-0541230 and #CCF-0541307, and an ONR Award #N00014-05-1-0188. The authors thank Alexei Efros and Mohit Gupta for technical discussions, Janice Brochetti for proof-reading and Shree Nayar for providing the CURET textures ([29]).

References

[1] R. J. Woodham, "Photometric stereo," *MIT AI Memo*, 1978.

[2] G. J. Klinker, S. A. Shafer, and T. Kanade, "A physical approach to color image understanding," *IJCV*, 1990.

[3] M. Oren and S. K. Nayar, "Generalization of the lambertian model and implications for machine vision," *IJCV*, 1995.

[4] D. Goldman, B. Curless, A. Hertzmann, and S. Seitz, "Shape and spatially-varying brdfs from photometric stereo," *ICCV*, 2005.

[5] K. E. Torrance and E. M. Sparrow, "Theory for off-specular reflection from roughened surfacesw," *JOSA*, 1967.

[6] E. Coleman and R. Jain, "Obtaining 3-dimensional shape of textured and specular surfaces using four-source photometry," *Intl Conf. Color in Graphics and Image Processing*, 1982.

[7] A. Shashua, "On photometric issues in 3d visual recognition from a single 2d image," *IJCV*, 1997.

[8] S. Mallick, T. Zickler, D. Kriegman, and P. Belhumeur, "Beyond lambert: Reconstructing specular surfaces using color," *ICCV*, 2005.

[9] H. D. Tagare and R. J. P. deFigueiredo, "A theory of photometric stereo for a class of diffuse non-lambertian surfaces," *IEEE Transactions on PAMI*, 1991.

[10] S. Mallick, T. Zickler, D. Kriegman, and P. Belhumeur, "Beyond lambert: Reconstructing surfaces with arbitrary brdfs," *ICCV*, 2001.

[11] Y. Sato, M. D. Wheeler, and K. Ikeuchi, "Object shape and reflectance modeling from observation," *SIGGRAPH*, 1997.

[12] S. Marschner, S. Westin, E. Lafortune, K. Torrance, and D. Greenberg, "Image-based brdf measurement including human skin," *Eurographics Workshop on Rendering*, 1999.

[13] R. Ramamoorthi and P. Hanrahan, "A signal-processing framework for inverse rendering," *SIGGRAPH*, 2001.

[14] A. Hertzmann and S. Seitz, "Example-based photometric stereo: Shape Reconstruction with general, varying BRDFs," *PAMI*, 2005.

[15] S. K. Nayar, K. Ikeuchi, and T. Kanade, "Surface reflection: Physical and Geometrical perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1991.

[16] S. K. Nayar, K. Ikeuchi, and T. Kanade, "Determining shape and reflectance of hybrid surfaces by photometric sampling," *IEEE Transactions on Robotics and Automation*, 1990.

[17] A. S. Georghiades, "Recovering 3-d shape and reflectance from a small number of photographs," *Eurographics Workshop on Rendering*, 2003.

[18] R. Basri and D. W. Jacobs, "Photometric Stereo with General, Unknown Lighting', *IJCV*, 2006.

[19] G. Healey and L. Z. Wang, "Segmenting surface shape using colored illumination," *SCIA*, 1997.

[20] L. Zhang, B. Curless, and S. Seitz, "Spacetime stereo: Shape recovery for dynamic scenes," *In Proc. CVPR*, 2003.

[21] H. Hayakawa, "Photometric stereo under a light-source with arbitrary motion," *JOSA*, 1994.

[22] S. G. Narasimhan, V. Ramesh, and S. K. Nayar, "A class of photometric invariants: Separating material from shape and illumination," *ICCV*, 2003.

[23] P. Debevec, T. Hawkins, C. Tchou, H. Duiker, W. Sarokin, and M. Sagar, "Acquiring the reflectance field of a human face," *In Proc. SIGGRAPH*, 2000.

[24] J. DeYoung and A. Fournier, "Properties of tabulated bidirectional reflectance distribution," *Graphics Interface*, 1997.

[25] A. Fournier, "Separating reflection functions for linear radiosity," *Eurographics Workshop on Rendering*, 1995.

[26] S. Rusinkiewicz, "A new change of variables for efficient brdf representation," *Eurographics Workshop on Rendering*, 1998.

[27] J. Kautz and M. D. McCool, "Interactive rendering with arbitrary brdfs using separable approximations," *Eurographics Workshop on Rendering*, 1999.

[28] G. Salton, "Automatic text processing: The transformation, analysis, and retrieval of information by computer," 1989.

[29] K. J. Dana, B. V. Ginneken, S. K. Nayar, and J. J. Koenderink, "Reflectance and texture of real world surfaces," *CVPR*, 1997.

[30] S. G. Narasimhan, C. Wang, and S. K. Nayar, "All the images of an outdoor scene," *ECCV*, 2002.

[31] H. P. A. Lensch, J. Kautz, M. Goesele, W. Heidrick and H. Seidel "Image-Based Reconstruction of Spatial Appearance and Geometric Detail," *TOG*, 2003.

[32] M. Pharr and G. Humphreys, "Physically Based Rendering: From Theory to Implementation," *Elsevier*, 2004.

[33] S. J. Koppal and S. Narasimhan, "Appearance Clustering Web Page," *http://www.cs.cmu.edu/∼koppal/clustering.html*, 2006.

[34] W. Matusik, H. Pfister, M. Brand and L. McMillan, "A Data-Driven Reflectance Model," *TOG*, 2003.