Abstract—This paper explores the new image processing domain involving images from depth cameras. While depth cameras are 2D arrays of data, the information they store is three dimensional in nature. Given a good depth image, the question now becomes what can we do with this added information. This paper explores the specific application of image relighting. Signal processing techniques of Gaussian and Bilateral Filtering are shown to be necessary to reduce noise in the surface normal estimation, critical for model-based lighting. The results include a Dual Bilateral Filter that uses information from both the depth and color image to estimate the surface normals.

I. INTRODUCTION

The purpose of this paper is to investigate image processing on depth images. Depth images are two dimensional arrays of pixels that correspond to the distance between an object and the camera. Figure 1 shows a typical depth image of a person and a chair. These images offer new opportunities and open new doors for novel applications which depend on the knowledge of a 3D environment.

II. DEPTH CAMERAS

Depth is the missing component of 2D images. In general, images are produced through the projection of a 3D scene onto a 2D plane. Given a point on the image, a ray can be traced from the center of the camera through that point to an object in the scene. Each point defines such a ray. The direction of the ray is fixed by the camera center and the image plane. The only question is the magnitude of this ray. One method to recover this depth information comes from the human visual system. Humans have the ability to perceive depth through the differences, known as disparity, between the separate images captured by each eye. In computer vision, two cameras are placed a set distance apart to mimic the eyes. Then, by determining a dense correspondence map, the disparity of each point is calculated. In both computer and human vision, this process is not extremely accurate. In areas with little texture, it can be difficult to detect correspondence points and therefore difficult to estimate the changes in depth. The human visual system compensates for this by incorporating vast amounts of prior knowledge of objects. This prior knowledge includes known physical characteristics of objects, known constrained movement patterns, and known reflectance parameters. Sadly, this sort of extensive prior knowledge is very difficult to realize in a computer system. Therefore, stereo vision is constrained to very rough depth maps.

For this reason, depth cameras, as seen in Figure 2 are extremely powerful. Instead of using a passive method to collect the data, these devices often utilize infrared illumination to actively probe a scene for its depth. The two main methods for determining the depth are by time of flight and structured light.

A. Time of Flight

Time of flight infrared distance calculation is not necessarily new technology. Laser measuring devices have been used for years to measure a distance along a line. The issue is that these devices only calculate the distance for one point at a time. Advancing technology has now opened up new doors which allow individual pixels to perform this time of flight distance calculation, thus providing a 2D grid of distance measurements. One of the disadvantages of this process is that pixels which calculate the time of flight distance are larger than traditional CCD elements. Thus the amount of pixels that can be fit onto an imaging device is smaller, which in turn means a smaller resolution size. One advantage is that each pixel is calculating its distance independently of others; thus typically more robust and accurate.
B. Structured Light

The other method for determining the depth is through structured light. As previously stated, if correspondence points can be established on an image disparity measurements can approximate the object depths. Structured light produces these correspondences by actively illuminating a scene with a lighting pattern that has some nice regularity. This pattern is infrared and therefore can not be seen by a human observer. It provides strong depth cues due to the warping of the lighting pattern over surfaces. An advantage of structured light is that it uses traditional CCD technology to determine the warped pattern. This technology is cheap and small and will likely lead to very compact and affordable depth cameras. A disadvantage of the structured light is that it is not as accurate as time of flight. The structure must first be estimated and then the distance estimated from the estimate of the structure.

III. APPLICATIONS OF DEPTH IMAGES

Depth image technology is relatively new. Companies have recently developed prototypes that are coming close to being commercially sold, yet an important question still remains mostly unanswered. This question is: Given a depth image, what can we do with it? Unlike color images, which contain a lot of visual information, depth images are smoothly varying and do not easily convey the depth information. Image Relighting using Depth is the topic of this paper, but the follow exhibit two other current areas of research involving depth images.

A. Image Segmentation

A classically difficult problem has been to automatically extract a foreground object out of a background scene from a single image. Current methods rely heavily on multiple images and large amounts of movement from the foreground object. It is assumed that the foreground is closer to the camera and moving in order to constrain the segmentation problem. With the depth camera and the assumption that the foreground is closer to the camera, a simple mask can be created by thresholding the depth image. This mask determines that every distance beyond a certain value is background and should be excluded. There are still issues that must be addressed in order to have clean smooth edges and to prevent unwanted holes in the mask, but this problem is much more realizable, particularly for real-time applications.

B. Depth Based Image Rendering

Depth Based Image Rendering [1], DBIR, is a technique that utilizes depth information to perform image based rendering. Image Based Rendering is a technique which generates a new image directly from a set of images. In contrast, Model Based Rendering uses a set of images to first generate a 3D model of an object from which it can then generate novel views. Disparity (or depth) plays an important role in image based rendering due the projection ambiguity inherent in cameras. Images from depth cameras overcome this ambiguity, and therefore are well suited for image to image warping. The current technique maps depth from a depth image to one or more color camera views, which then fill in the missing depth through bilateral filtering with the color image. After processing, each color image now has a depth image counter part. Now color information can be propagated to a novel view point for novel image rendering. The advantage of DBIR is that it requires very few cameras and does not rely on any kind of correspondence matching.

IV. WHY RELIGHT?

Before we explore how a depth image can be used to relight an image, first let us look into why a user would want to relight an image. As we can see from Figure 3, not all pictures that are taken have the best lighting environment. And it turns out that illumination is an important factor in how humans perceive an environment. Illumination contains important cues of structure as well as important psychological cues. Consider the classic technique of holding a flash light under one’s chin in order to tell a scary story. This has the ability to cause an unsettling feeling because it is unnatural to see light coming the bottom of the human face. Also, consider the hundreds of thousands of dollars spent in modern film studios. The objective is to capture perfectly lit scenes at the moment of capture. The average user will not have access to these kinds of resources and not all scenes can have such carefully controlled lighting setups. One might desire to capture the scene as is, and simply modify the light of the scene after the moment it was recorded. Thus, we have the problem of image relighting.

V. LIGHTING THEORY

In order to see how a depth image can help to relight an image, we first examine the simplest known lighting model. It was developed in the 18th century by the mathematician Johann Heinrich Lambert. He realized that certain surfaces reflect light equally in all directions, independent of the viewing angle. These surfaces are called diffuse, in that they
diffuse light in all directions. For diffuse surfaces, the amount of light that is reflected is proportional to the solid angle of light that hits the surface. This means that light sources which are perpendicular to the surface cause the brightest reflectance, while light sources approaching parallel contribute less and less light. This can be described for one source in the equation

\[ i = \rho \max((n, s), 0) \]  

(1)

Here \( \rho \) is a constant called the albedo, which represents the amount of light a surface absorbs, \( n \) is the normal to the surface, and \( s \) is a unit vector pointing in the direction of the light source. The inner product between the surface normal and the light source vector captures the phenomenon that Lambert observed. Thus, from a depth image, surface normals can be calculated. The spherical harmonic representation is

\[
I_{u,v} = \rho_{u,v} \int_{S^2} L(\omega) \max((n_{u,v}, \omega), 0) \; d\omega
\]  

(2)

where \((u, v)\) describes the position in an image, \(L(\omega)\) is the continuous lighting function, and \(S^2\) is the unit sphere. Recent advances in understanding reflectance have also shown that Lambert’s equation can be expressed in another form using spherical harmonics. This representation is useful because it can approximate arbitrary diffuse lighting conditions very well using only a few lighting coefficients. The spherical harmonic representation is

\[
I_{u,v} = \rho_{u,v} \sum_{i=1}^{\infty} h_i(n_{u,v}) l_i \approx \rho_{u,v} \sum_{i=1}^{9} h_i(n_{u,v}) l_i
\]  

(3)

where \(h_i\) are the spherical harmonics which are functions of the surface normals, and \(l_i\) are the spherical harmonic lighting coefficients. In both versions of Lambert’s equation, we can see that if the surface normals and albedos are known, then a novel lighting function can be applied to relight an image. In fact, it has been shown that if we know just the surface normals and albedos for the given image [2], then we can use the spherical harmonic representation to estimate both the lighting and the albedos for the given image [2]. Thus, in order to relight, one needs the surface normals.

\[
\frac{\partial f}{\partial x} = \left[ \frac{\partial x}{\partial x}, \frac{\partial y}{\partial y}, \frac{\partial F(x,y)}{\partial x} \right] = \left[ 1, 0, \frac{\partial F}{\partial x} \right]
\]

\[
\frac{\partial f}{\partial y} = \left[ 0, 1, \frac{\partial F}{\partial y} \right]
\]

Given two vectors on the surface of curve, a third vector perpendicular to both can be calculated using the cross product. Thus, the surface normal is

\[
\frac{\partial f}{\partial x} \times \frac{\partial f}{\partial y} = \begin{vmatrix} i & j & k \\ \frac{\partial F}{\partial x} & \frac{\partial F}{\partial y} & 1 \end{vmatrix} = \left[ \frac{\partial F}{\partial x}, \frac{\partial F}{\partial y}, -1 \right]
\]

Here, it is seen that the surface normal can be described by its gradients in the x and y directions. In discrete image processing, gradient are often approximated by the differencing between two points. Such as

\[
F_x = F(n, m + 1) - F(n, m)
\]  

(4)

Thus, from a depth image, surface normals can be calculated. Ignoring the albedos for the moment, the reflectance of a scene can now be rendered. Figure 4 shows a color image and its depth map. Figure 5 shows a reflectance image using the estimated surface normals and a single source placed close to the scene.

VI. SURFACE NORMALS

We can derive the surface normals of a depth image as such, consider a function

\[
f(x, y) = [x, y, F(x, y)]
\]

which describes a curve in space defined over the x-y plane. Then vectors tangent to the surface of \(f\) in the x and y directions can be calculated by

\[
\frac{\partial f}{\partial x} = \left[ \frac{\partial x}{\partial x} \frac{\partial y}{\partial x} \frac{\partial F(x,y)}{\partial x} \right] = \left[ 1, 0, \frac{\partial F}{\partial x} \right]
\]

\[
\frac{\partial f}{\partial y} = \left[ 0, 1, \frac{\partial F}{\partial y} \right]
\]

(a) Color Image

(b) Depth Image

Fig. 4. Color and Depth Images using for experimentation
VII. Filtering Gradients

Figure 5 reveals that this naive surface normal estimation is very poor. We can see that the depth map is mostly smooth with discontinuities at edges, whereas the rendering is very textured. This is an issue of noise in the depth map, particularly a condition of quantization. Since this depth map was saved as an image, some of the areas are quantized to flat regions; thus, where there should be some kind of curve, there is flatness distorting the gradients and the surface normals significantly.

From ECE 547, we learned that one way to reduce the noise is to filter an image. Therefore, the first attempt to clean the surface normals was a gaussian blurring of the depth image. This image can be seen in Figure 6. This rendering is much smoother than the unfiltered attempt, but there are still issues, particularly at edge boundaries, where the blurring process has smoothed edges that previously represented distinct objects.

Next, we examine what effect filtering the depth image has on the gradients. Define a 1D function $f(n)$, and define a gradient of that function as

$$ f_g(n) = f(n + 1) - f(n) $$

The gradient of a filtered function is a weighted sum of the gradients of unfiltered function. The problem with gaussian filtering is that it averages gradients across boundary edges. This problem is common of all types of convolution filtering; therefore, I looked to the bilateral filter [3]. The bilateral filter employs filtering in both the spatial domain and the color domain. The gaussian filter only considers how close two pixels are in space when determining the averaging weight. The bilateral filter still considers the spatial distance, but it also takes the difference between two pixel’s color information. The bilateral filter can be expressed as

$$ \hat{I}_u = \frac{1}{W_u} \sum_{v \in N_u} G_{\sigma^2}(|x_u - x_v|)G_{\sigma^2}(|I_u - I_v|)I_v $$

$$ W_u = \sum_{v \in N_u} G_{\sigma^2}(|x_u - x_v|)G_{\sigma^2}(|I_u - I_v|) $$

Here $\hat{I}_u$ is the filtered value at point $u$, $G_{\sigma^2}$ denotes a gaussian function of variance $\sigma$, and $N_u$ is a neighborhood around point $u$. $W_u$ is a normalization factor so the filter sums to one. The results of bilateral filtering can be seen in Figure 8 where the filtering is performed using the intensity information from the depth image itself, and Figure 9 using the color information from the color image. The intuition behind using the color image for filtering comes from [1], which observed that the color and depth images often have similar object boundary edges.

The final attempt to filter the gradient is a combination of the two bilateral filters from the last step. Thus, it is a dual bilateral filter which incorporates information from the edges
Fig. 7. Comparison between the reflectance images with various filters. (a) No filter, (b) Gaussian filter, (c) Bilateral filter using depth edges, (d) Bilateral filter using color edges, (e) Dual Bilateral filter

Fig. 9. Reflectance using surface normals after Bilateral Filter with color edge information

Fig. 10. Reflectance using surface normals after Dual Bilateral Filter

\[
\hat{I}_u = \frac{1}{W_u} \sum_{v \in N_u} G_{\sigma^2_v} (|x_u - x_v|) G_{\sigma^2_c} (|I_{c,u} - I_{c,v}|) \cdot G_{\sigma^2_d} (|I_{d,u} - I_{d,v}|) I_v
\]

Results for this filter can be seen in Figure 10. A comparison of the various filtering results can be seen in Figure 7. From here we see that the dual bilateral filter produces the best results. One disadvantage of the using the color information for filtering is that some of the texture in the color image leaks unwanted into the final image.

Fig. 11. Image Relight using Dual Bilateral Filter for surface normal estimation and Spherical Harmonic Least Squares albedo estimation

VIII. FUTURE WORK

While this work has improved the robustness of the surface normals for noisy depth maps, there is still much work to be done with the image relighting problem. Figure 11 shows a relit image using the dual bilateral filter for surface normal estimation and spherical harmonic least squares estimation for the albedo map. The following lists the main areas still in need of improvement.

• Identify a way to utilize the strong edge content in the color images without the unwanted texture information
• Reduce the boldness of the attached shadows. Typically, light will reflect off close by surfaces and thus soften shadows.
• Correct the lack of cast shadows and specular highlights.
• Remove the cast shadows and highlights from the previous image.

IX. CONCLUSION

In conclusion, the application of using depth images to relight images relies heavily upon accurate surface normal estimation. This research has shown that the bilateral filter can be employed successfully to filter noisy depth images in order to generate smooth surface normals while still persevering the edge information critical to the depth map.
REFERENCES