

Object Modeling using Tomography and Photography

David T. Gering
MIT Artificial Intelligence Laboratory
Cambridge, MA 02139
gering@ai.mit.edu

William M. Wells III
Harvard Medical School and Brigham and
Women's Hospital, and
MIT Artificial Intelligence Laboratory
sw@ai.mit.edu

Abstract

This paper explores techniques for constructing a 3D computer model of an object from the real world by applying tomographic methods to a sequence of photographic images. While some existing methods can better handle occlusion and concavities, the techniques proposed here have the advantageous capability of generating very high-resolution models with attractive speed and simplicity. The application of these methods is presently limited to an appropriate class of mostly convex objects with Lambertian surfaces. The results are volume rendered or surface rendered to produce an interactive display of the object with near life-like realism.

1 Introduction

This paper addresses the problem of reconstructing an object from its images and rendering it from novel viewpoints. Virtualized reality, introduced by Kanade, Narayanan, and Rander [4,7], has applications from animated movies, to interactive education, to exchanging information on the World Wide Web. Creating virtual representations of real-world scenes departs from the conventional stereo correspondence problem in that information must be gathered and integrated from all sides of the scene.

This paper describes how tomography can be used to create a model of a 3D object and render it with near life-like realism, as demonstrated in Figure 1. Tomography refers to the process of forming cross-sectional images of an object by illuminating it from many different directions. Radon [8] first solved the problem of reconstructing an object from its projections in 1917. The field of medical imaging was revolutionized when Hounsfield invented the x-ray computed tomographic (CT) scanner for which he shared the Nobel prize in 1972. The output of a CT scan is a map of x-ray attenuation coefficients which offers doctors a view of internal organs. The new application considered here differs from

diagnostic medicine in that the object being imaged is opaque.



Figure 1: Example of a color, 3D model created using the methods described here.

We propose applying tomographic techniques to conventional photographic and video imagery. There are a variety of tomographic reconstruction methods. This paper applies a type of simple backprojection and suggests that more methods from the history of tomography are relevant to forming reconstructions from photographic images. For example, it may become necessary to have less restricted camera trajectories, which has already been studied by Horn [3].

2 Related work

Early work by Bolles, Baker, and Marimont [1] proposed constructing an epipolar-plane image which is simpler to analyze than a large set of images. Mellor, Teller, and Lozano-Perez [6] extended their approach to work with arbitrary camera positions and some forms of occlusion. Contour-based methods were applied by Szeliski [13] and Seales and Faugeras [11]. One approach explored here is similar to these in that it employs silhouettes such as Szeliski does, yet it avoids the complexity of using octrees. Our techniques share similarities with the recent approaches that have turned away from stereo methods to use first principles. Collins [2] digitized scene space into voxels that are traversed and

colored. Seitz and Dyer [10] furthered this approach with a voxel coloring algorithm to handle occlusion. Szeliski and Golland [12] also handle occlusion by way of an ordering technique as they simultaneously find disparities, true colors and opacities. Some approaches explored in this work differ from the other volumetric approaches in that there is no thresholding or segmentation until *after* the volume has been reconstructed.

The tomographic methods explored here have difficulty handling occlusion and concavities. However, for an appropriate class of mostly convex objects, these techniques are attractive for their speed and simplicity. The setup and calibration are fast and easy, and the equipment is simple, inexpensive, and commonplace. The algorithm scales linearly in the number of views and is very computationally inexpensive when compared with the other techniques. Best of all, the resulting images look pleasing and have very high resolution. The image on the previous page was reconstructed using 4.6 million voxels in 8 minutes on a Pentium II, 266MHz machine. For comparison, Seitz and Dyer state that their voxel coloring algorithm reconstructs 53,000 voxels in 95 seconds. If we could assume linear scaling, computing 4.6 million voxels would take 134 minutes.

3 Method

3.1 Data acquisition

In medical imaging, the patient lies on a table while the x-ray source and detector spin about. We took the easier approach of revolving the object of interest on a milling machine rotary table in view of a fixed camera and light source.

A pullnix color CCD camera was affixed with a FUJI 75 mm lens (long enough to assume orthographic projection) positioned 93 inches from the object. The object was a plastic representation of Tweety Bird in front of a black background. The scene was irradiated with a spotlight reflected off a white panel directly behind the camera. 180 color pictures of size 160x180 were taken over 360 degrees of rotation. Samples are shown in Figure 2.

3.2 The radon transform

The Radon transform, $P(a,d)$, of a function, $f(x,y)$, is computed by forming a projection of the function for each of many views, a . For a given projection, $P(a)$, the value at location d , $P(a,d)$, is determined by the line integral through $f(x,y)$ along a ray. This ray makes an angle, a , with the x-axis, and the ray stands a distance, d , from a parallel ray through the origin. Hence, the origin projects

onto $d=0$. This relationship can be written using a delta function as:

$$P(a,d) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y) \delta(x \cos a + y \sin a - d) dx dy$$

Figure 3 demonstrates this graphically.

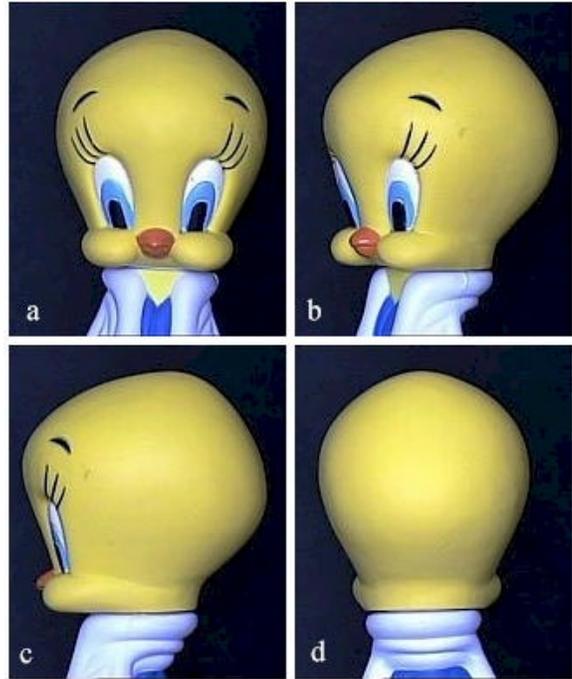


Figure 2: Original photographic views, (a) through (d), are taken at 0, 50, 90, and 180 degrees, respectively.

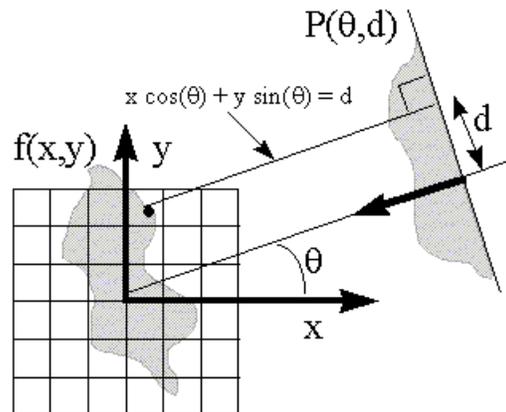


Figure 3: Computation of the radon transform.

3.3 The modified radon transform

A sequence of photographic images can be thought of as a modified Radon transform that differs in being discrete and operating on opaque data. Each image is taken at a different view angle, a . Each row of an image is a projection through an axial slice of the object of interest. Projections are vertically stacked to form a 3D modified Radon transform. Each pixel of each row has a value determined by a photographic detector horizontally located a distance, d , from the center of the lens.

The line integrals in the Radon transform become something else in the modified version. For translucent objects, the detectors *see* the result of compositing the color and opacity of volume elements, or voxels, in the order they are encountered along the ray from the object to the camera. For opaque objects, such as our Tweety Bird, each detector receives the scene radiance of a surface patch on the object.

Tomographic methods would ideally be applied under ambient lighting conditions. Patches of a Lambertian surface have a hemispherical radiation pattern with brightness proportional to the cosine of the angle between the view ray and the surface normal. This effect is cancelled by foreshortening so that different views receive identical brightness information from a given portion of the surface. However, one goal of our work is to model objects as simply and inexpensively as possible. Given this constraint, ambient lighting conditions are not always possible, and we used diffuse lighting. In a diffuse lighting model, the power irradiating a surface patch falls off with the cosine of the angle between the incident light ray and the surface normal. (Since the light source was positioned directly behind the camera in or case, we can refer to the view angle and the angle of incidence interchangeably.) Each non-occluded surface patch makes a brightness contribution to the modified Radon transform in the half of all views which see its front side.

Diffuse lighting also introduces artifacts into the final reconstruction. Consider the fact that we implement tomography by spinning the object about one axis. Surface normals have an axial component and a vertical component. Patches that are more vertically oriented will contribute less to the modified Radon transform, and this will appear as an artifact in a reconstruction because more vertically oriented patches will be darker.

For shiny surfaces, incident light reflects about the surface normal at an angle equal to the angle of incidence, just as seen in a mirror. When this angle of reflection coincides with the viewing angle, a bright spot appears as specular reflection. Therefore, if a patch is shiny, then specular highlights will appear in the modified Radon transform. This leads to artifacts in the final reconstruction, because the highlights will be fixed rather

than varying with the viewing angle of the observer who is exploring the reconstructed object.

Putting these two, small artifacts aside for the remainder of the discussion, call the axial component of a patch's surface normal, a_n , the angle of incident light, a_i , and the patch's brightness, I . If the patch is small enough to be seen by only one detector at a time, then the patch's total contribution to the modified Radon transform is:

$$\sum_{a_i=0}^{360} \text{MAX}(I * \cos(a_i - a_n), 0)$$

We will later refer to $a_i - a_n$ as *phi*.

3.4 Backprojection

Object reconstruction becomes a task of inverting the modified Radon transform. Refer again to Figure 3 to observe how we can recover the function, $f(x,y)$. Each $P(a)$ can be projected back onto the x-y grid, thus earning the name *backprojection*. The pixel with a black dot (located at $x=1, y=3$) can be recovered by computing the value of d from the angle, a , and the coordinates, x and y , from the ray equation:

$$x \cos(a) + y \sin(a) = d$$

Then, d is used to index into $P(a)$ to obtain a value, $P(a,d)$, to assign to the pixel. This process is repeated for each view, and the values are summed to accumulate one aggregate pixel value.

3.5 Filtered Backprojection

Suppose $f(x,y)$ is an impulse of density at the origin. Then, each image will contribute a ray of brightness that emanates from the center of the image and intersects the center of the reconstruction. The composite reconstruction will appear much like a bunch of bright spokes intersecting at the center. In the limit of infinitely many views, the reconstruction will approach a $1/r$ distribution, where r is the distance from the origin.

If we view the image collection process and (unfiltered) backprojection together as a linear system, we see that the impulse response is $1/r$, and because the components of the system are linear (and shift-invariant), the response of the system to an arbitrary density will be a convolution of that density by $1/r$.

Ideally, we would prefer the impulse response of our imaging and reconstruction to be close to a delta function. To obtain this, we could post-filter the reconstruction with a filter that has been designed to be an implementation of the inverse of $1/r$ — this would be a *de-blurring* or *de-convolution* approach to the problem. It turns out that in this situation, some efficiency may be gained by using a suitable filter in the images before they are backprojected

— this is the standard *filtered backprojection* (FBP) approach.

It is useful to observe that the standard filters used in the first step of FBP are essentially high-pass filters. Therefore, they may be also viewed as edge-enhancing filters, and then FBP may be summarized as *edge-enhance, then backproject*.

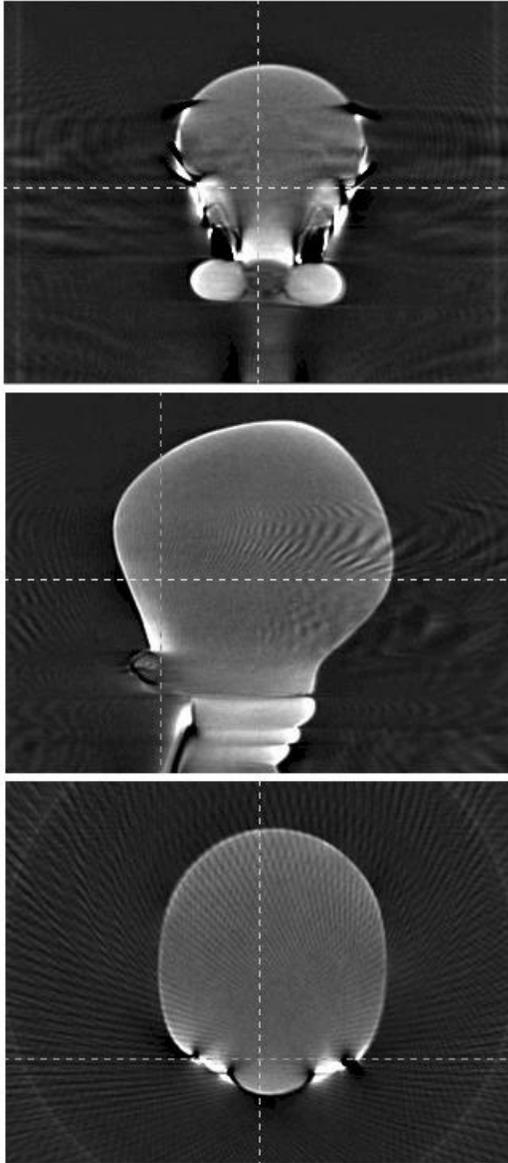


Figure 4: Cross sections of reconstructed density using filtered backprojection.

We employed a 1D filtering kernel prior to backprojection. The kernel is due to Ramachandran et al. [9], uses linear interpolation, and has the form:

$$k(0) = 1/4$$

$$k(i) = \begin{cases} \frac{-1}{(pi)^2}, & \text{odd } i \\ 0, & \text{even } i \end{cases}$$

The results of FBP reconstruction are shown in Figure 4 (the contrast of the images has been adjusted for viewing). It is a 3D volume containing a representation of the object as an intensity density. In this example, the object appears as a bright region embedded in a dark volume. In the reconstruction, the bright region is shaped like the original 3D object. The surface of the object appears as a prominently bright shell, and the brightness of the shell is modulated by the brightness of the original object. For example, in the coronal cross section, which has been chosen to intersect the reconstruction in the region of the face, the bright and dark areas of the eyes are clearly visible.

3.6 A new filter for backprojection

Figure 2a-d shows samples of the original photographs. There is a modified Radon transform image for each axial slice of Tweety. Figure 5a shows the Radon transform for one slice through Tweety's eyes. Each row of the image is a projection for a view, or $P(a)$, and each column is a d .

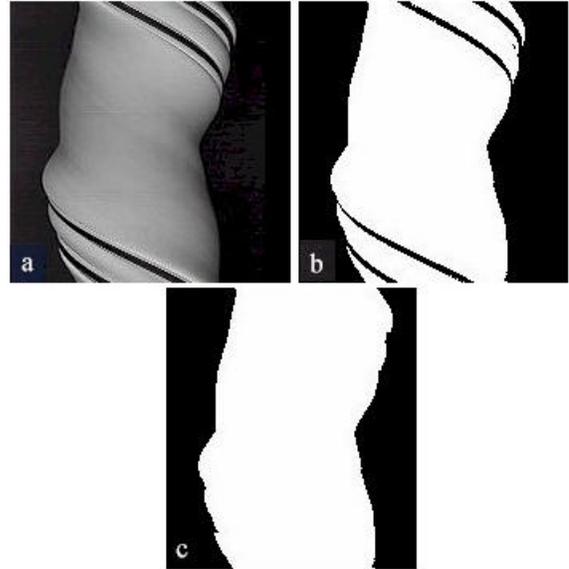


Figure 5: Radon transform (a). Image (b) is derived from applying a threshold to (a) just above the background noise floor. Image (c) is derived from filling the gaps along each row of (b).

Figure 6a shows the unfiltered backprojection of the first row of 5a. Figure 6b adds the backprojection of the

middle row. Observe how the yellow color of the back of Tweety's head has overwritten the blue and black color of his eyes. The eye color is still apparent, yet it has lost its original brilliant saturation by being averaged with *intruder colors* that do not belong at that location on the x-y grid.

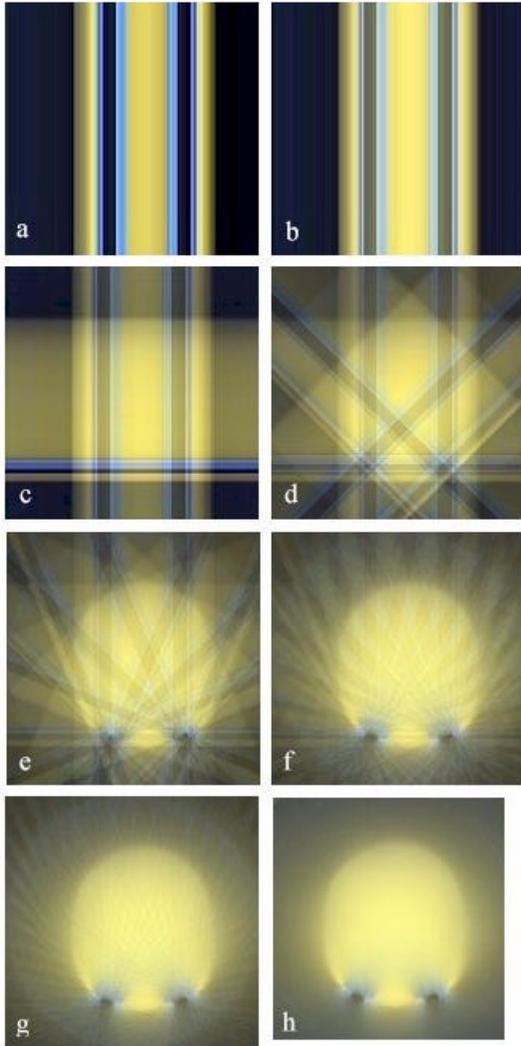


Figure 6: Images (a) through (h), are tomographic reconstructions using 1, 2, 4, 8, 16, 32, 64, and 180 equally spaced views, respectively.

The process continues unfolding in Figure 6c, which is a tomographic reconstruction from 4 equally spaced views. Observe that the blue and black color of surface patches in the eyes are still seen from the side views, but the patches experience foreshortening. Figure 6 demonstrates that the shape of the object emerges in as few as 16 views, but shape and color accuracy continue improving as more views are added.

During backprojection, a pixel corresponding to a surface patch on Tweety, receives back its contribution to the Radon transform from the views which saw it as expressed earlier in equation form. However, the other half of the views also contribute *intruder colors* to the pixel. The pixel's color becomes averaged with the average color seen by that pixel's detector in the other half of the views. The end result (Figure 6h) is a reconstruction of both the object's shape and texture situated in a *tomographic fog*.

A correct inverse modified Radon transform will assign each pixel the exact value that it contributed to the transform. We can dissipate the fog and block intruder colors by stopping the projections from views that do not see the patch. However, an accurate inverse modified Radon transform is not desired! When the absolute value of the difference between the angle of incident light and the axial component of the surface normal, $|\phi|$, is large (near 90 degrees), the patch is seen to be darker by the detector. Therefore, the patch's contribution to the modified Radon transform is darker than the patch itself, and this will result in a 3D model with muted colors.

We can correct for both problems simultaneously by weighting each backprojection by $\cos(\phi)$. However, the surface normals are not initially known. This suggests a two-pass algorithm where the first pass uses standard filtered backprojection to recover the object. The surface normals are then computed from the reconstruction, and then filtered backprojection is re-implemented with a different filter for each ϕ .

3.7 Rendering the reconstruction

The end goal of this work is a realistic graphics rendering of the object. Filtered backprojection of all three color components of the data provides a 3D volume of object texture. The data can be visualized through volume rendering. Consider the 2D output image to be a window into the scene. For each pixel in the window, a ray is cast into the scene, and a color value can be assigned to the pixel using one of a variety of methods. Maximum Intensity Projection (MIP) [14] colors the pixel with the maximum intensity found along the ray. Composite rendering composites colors that the ray passes through according to the opacity values of those voxels, and the order in which the ray encounters them.

Either method encounters trouble with Tweety Bird. The black eyes are nearly indistinguishable from the dark surroundings, so MIP colors the eyes with the yellow color from the back of Tweety's head. Composite rendering requires opacity information.

Figure 7 shows a MIP of the reconstructed intensities from a viewpoint that was chosen to be below the plane of the views of the input data. In this preliminary experiment, the front/back surface issue was handled by

manually suppressing the rear half of the intensity data. The purpose of this experiment is to demonstrate that useful surface brightness information is present in the reconstruction, yet the filtering has introduced highlight and shadow artifacts most pronounced around the eyebrows, which are high-frequency texture. Consequently, the next renderings were generated from reconstructions that result from unfiltered backprojection.

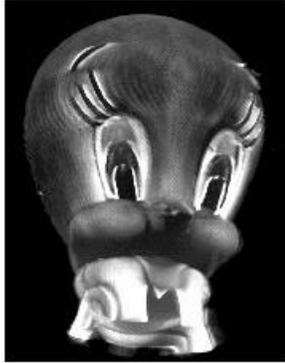


Figure 7: MIP of filtered, backprojected reconstruction from novel viewpoint.

3.8 Recovering opacity

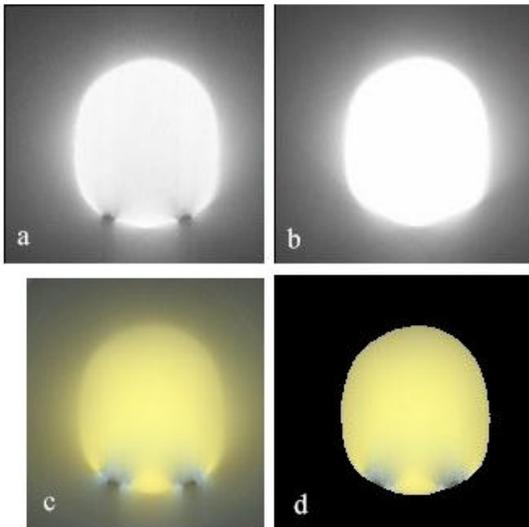


Figure 8: Image (a) is an opacity map derived from backprojecting (5.b). Image (b) is derived from (5.c) to assign non-zero opacity to the eyeballs. Image (c) is one slice of the reconstruction, while image (d) is the same slice rendered using the opacity information in (b).

To aid the rendering process, a second backprojection (but with only a single color component: gray) was

performed to create an opacity map to complement the reconstructed texture map. The Radon transforms were automatically thresholded just above the noise floor of the background, as shown in Figure 5.c.

A reconstruction made from this filtered transform is shown in Figure 8.a, where the eyeballs do not exist. Using the assumption of a mostly convex object, the threshold operation was modified to fill in gaps. The resulting Radon transform shown in Figure 5.d was used to reconstruct the opacity map of Figure 8.b.

The color and opacity information form a 4-component volume that can be used successfully for rendering. Figure 8.c and 8.d show a rendering of the unfiltered, one-pass reconstructed slice with and without using opacity information to segment the image.

4 Results

In an effort to perform object modeling in the most computationally inexpensive way, renderings were made from the unfiltered, single-pass backprojection with the help of the opacity map.

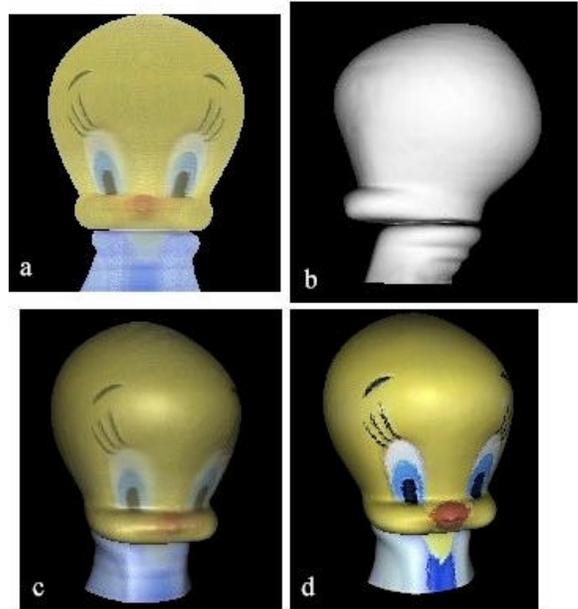


Figure 9: Image (a) is a volume rendering using intensity projection of texture from (8.c) and the opacity information from (8.b). Image (b) is a surface model of the opacity map in (8.b). Image (c) is the surface model of (b) with the texture from the backprojected reconstruction (8.c). Image (d) is similar, except the texture is taken from the original photographic images (2).

4.1 Volume rendering

Figure 9.a shows an intensity projection made from casting rays through the opacity information until a voxel is encountered with a opacity greater than 90% of full. Then the color values were taken from that voxel.

The flat result is a nature of intensity projection. More beautiful results can be obtained by computing the surface normals of the opacity map and applying a lighting model to shade the object. This technique was performed with the surface rendering described below.

4.2 Surface rendering

An alternative to visualizing our reconstructed volumes through ray-casting is to construct a surface model as a collection of triangles. The Marching Cubes algorithm [5] was run on the opacity map to extract the surface. Marching Cubes can be thought of as creating a polygon that intersects each voxel where the volume's scalar data crosses a threshold, such as opacity being 90% of full value. Figure 9.b shows a 3D rendering of the Tweety's surface. Figure 9.c colored each triangle vertex using the color components from the unfiltered backprojection, and the surface normals from the opacity map. Figure 10 shows a close-up of the many triangles that comprise Tweety's surface.

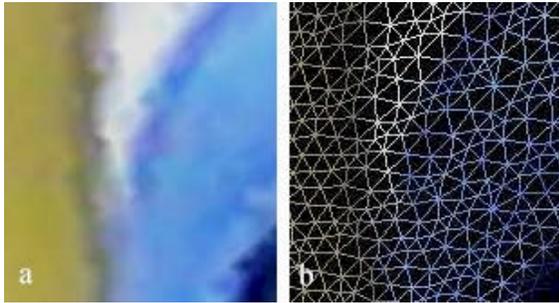


Figure 10: Image (a) is a close-up view of the eye of (6.d). Image (b) is a wire frame representation.

4.3 Projected texture

Figure 9.d is the most life-like representation of Tweety. The surface model described above was colored by computing the axial component of each surface normal. The angle of this vector was used to select the original photographic image with the closest view angle. (Equivalently, a row was selected from the modified Radon transform for this axial slice). Then the detector, d , which observed the vertex in this view was computed from the ray equation described earlier. The vertex's color could then be taken from $P(a,d)$.

This process is effectively performing backprojection with a delta function that selects the view that saw a

surface patch most directly. Thus, unconcluded patches are recovered with their fully saturated brightness.

$$\sum_{a_i=0}^{360} \delta(a_i - a_n) (I * \cos(a_i - a_n))$$

4.4 Future work

Figure 9.d appears a little rough due to inaccuracies in the computation of the surface normals. Errors could be averaged out by modifying the above equation to weight a number of views that observed the patch most directly. Future work can similarly use the opacity map to assign weights to eliminate occluded views. The occlusion in some types of objects could be handled by performing tomography about two different axis of rotation. Furthermore, another class of tomographic techniques [15] offers algebraic reconstruction techniques (ART) as an iterative alternative to backprojection.

5 Summary

This paper presents an analysis of how tomography could offer a simple, computationally inexpensive solution to recovering an object's shape and texture from multiple views. These techniques apply best to objects that are mostly convex, opaque, and with Lambertian surfaces. Under these circumstances, the algorithms produce nice-looking, very high-resolution results in a comparatively short time.

References

- [1] R. C. Bolles and H. H. Baker and D. H. Marimont. Epipolar-Plane Image Analysis: An Approach to Determining Structure from Motion. *International Journal of Computer Vision*. 1(7):7-55, 1987.
- [2] R. Collins. A space-sweep approach to true multi-image matching. *Proc. Computer vision and Pattern Recognition Conf.*, 358-363, 1996.
- [3] B. K. P. Horn. Density Reconstruction using Arbitrary Ray-Sampling Schemes. *Proceedings of the IEEE*, 66(5):551-562, May, 1978.
- [4] T. Kanade, P. J. Narayanan, P. W. Rander. Virtualized Reality: Concepts and Early Results. *IEEE Workshop on the Representation of Visual Scenes*. Boston, June 24, 1995.
- [5] W. E. Lorensen and H. E. Cline. Marching Cubes: A High Resolution 3D Surface Construction Algorithm. *Computer Graphics* 21(3):163-169, July 1987.
- [6] J. P. Mellor, S. Teller, and Thomas Lozano-Perez. Dense Depth Maps from Epipolar Images. Technical Report 1593. MIT Artificial Intelligence Laboratory, 1996.
- [7] P. J. Narayanan and P. W. Rander and T. Kanade. Constructing Virtual Worlds Using Dense Stereo.

Proceedings of the 6th International Conference on Computer Vision. Bombay, 3-10, 1998.

- [8] J. Radon. Über die Bestimmung von Funktionen durch ihre Integralwerte angs gewisser Mannigfaltigkeiten. *Berichte Sachsische Akademie der Wissenschaften*, Leipzig, Math. — Phys. Kl. 69:262-267, 1917.
- [9] G. N. Ramachandran and A. V. Lakshminarayanan. Three Dimensional Reconstruction from Radiographs and Electron Micrographs: Applications of Convolutions Instead of Fourier Transforms. *Proc. Natl. Acad. Sci. USA*, 68:2236-2240, 1971.
- [10] S. M. Seitz and C. M. Dyer. Photorealistic scene reconstruction by space coloring. *IEEE Computer society Conference on Computer Vision and Pattern Recognition (CVPR '97)*, San Juan, Puerto Rico, June 1997.
- [11] W. Seales and O. Faugeras. Building three-dimensional object models from image sequences. *CVGIP: Image Understanding*, 3(61):308-324, 1995.
- [12] R. Szeliski and P. Golland. Stereo Matching with Transparency and Matting. *Proceedings of the 6th International Conference on Computer Vision.* Bombay, 1998.
- [13] R. Szeliski. Rapid octree construction from image sequences. *CVGIP: Image Understanding*. 1(58):23-32, 1993.
- [14] U. Tiede and K.H. Hohne et al. Investigation of Medical 3D-Rendering Algorithms. *IEEE Computer Graphics Applications*, pages 41-51, March 1990.
- [15] S. Webb. *The Physics of Medical Imaging.* Institute of Physics Publishing. London, 1988.