

A Sensor for Simultaneously Capturing Texture and Shape by Projecting Structured Infrared Light

Kiyotaka Akasaka, Ryusuke Sagawa, Yasushi Yagi
Institute of Scientific and Industrial Research, Osaka University,
8-1 Mihogaoka, Ibaraki-shi, Osaka, 567-0047, JAPAN
{akasaka,sagawa,yagi}@am.sanken.osaka-u.ac.jp

Abstract

Simultaneous capture of the texture and shape of a moving object in real time is expected to be applicable to various fields including virtual reality and object recognition. Two difficulties must be overcome to develop a sensor able to achieve this feature: fast capturing of shape and the simultaneous capture of texture and shape. One-shot capturing methods based on projecting colored structured lights have already been proposed to obtain shape at a high frame rate. However, since these methods used visible lights, it is impossible to capture texture and shape simultaneously. In this paper, we propose a method that uses projected infrared structured light. Since the proposed method uses visible light for texture and infrared light for shape, simultaneous capturing can be achieved. In addition, a system was developed that maps texture on to the captured shape without occlusion by placing the cameras for visible and infrared lights coaxially.

1 Introduction

If the texture and shape of a moving object is captured simultaneously, the data can be applied to various fields, such as virtual reality (VR), computer vision, gesture recognition, and robotics. For example, if the texture and shape of a person in motion is obtained, the model of the person can be easily placed in a virtual world by using 3D computer graphics. To achieve the goal of simultaneously capturing texture and shape in real-time, it is necessary that a sensor synchronously captures the texture and shape of an object at a high frame rate.

Techniques to acquire the shape of an object is categorized roughly into two types, active and passive methods. Stereo vision [5] and visual hulls [9] are often used in passive methods. Since these methods measure the shape of an object using synchronized multiple cameras, it is suit-

able for use in capturing texture and shape simultaneously. Moreover, it is possible to capture data at a high frame rate because all the input data is obtained at the camera frame rate. However, the accuracy of the measured shape is lower compared to active methods; it is sufficient to use a 3D model for computer graphics. With use of a visual hull, a large system is required since many cameras are necessary to observe an object from various viewpoints.

For active methods, laser range finders and structured-light methods can be employed. A laser range finder is often used as an accurate method to measure shape. Laser range finders measure the distance between the sensor and an object based on triangulation or time-of-flight. A time-of-flight sensor that changes the direction of a laser beam by using a mirror cannot capture shapes at a high frame rate since each change takes several seconds or minutes. Although fast range finders [1, 2] have recently been developed, their resolution and accuracy are not sufficient due to the trade-off between time and accuracy.

Structured-light methods cast light patterns on an object by using a projector and the resulting patterns are observed by a camera. This technique locates the corresponding coordinates between the projector and the structured-light patterns observed by the camera. The distance from the camera to the object is computed with the triangulation based on the correspondences. Some fast methods [8, 4, 11, 3] are proposed based on structured-light methods. These procedures measure shape at a high frame rate. Especially [11, 3] which are suitable for real-time acquisition because they can capture shape with a one-shot pattern. However, these methods have difficulty in obtaining texture simultaneously as visible light patterns are projected to measure shape. As two solutions to this problem, time-sharing methods [10, 7] and fringe image based methods [12, 6] have been proposed. First methods switch between capturing texture and shape at a high frame rate. Second methods project few different color patterns (fringe patterns) and restore color texture from captured images. However, misalignments between texture and shape may appear since the acquisition is not simulta-

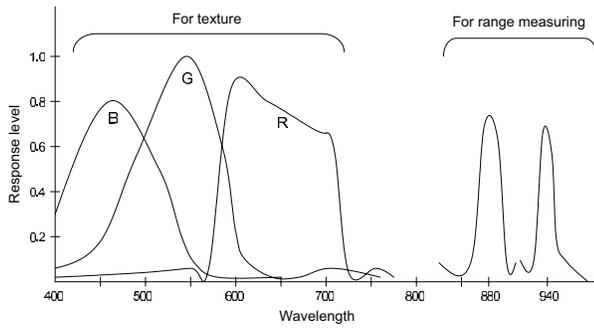


Figure 1. Distribution of the wavelengths for capturing texture and shape.

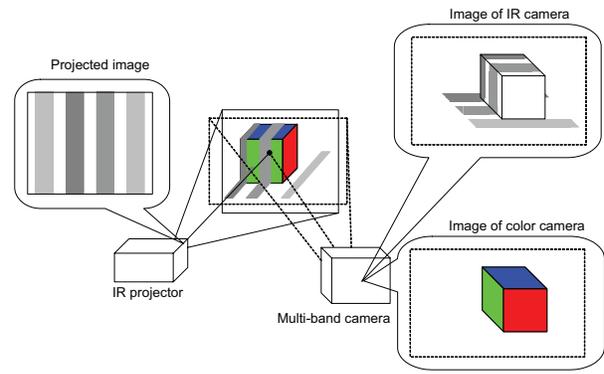


Figure 2. Overview of the sensor for the simultaneous capturing of texture and shape.

neous in both methods.

As previously stated, the structured-light method cannot capture texture and shape simultaneously as visible light, used for capturing shape, is also required for capturing texture. Therefore, we propose a system that projects structured infrared light (IR) instead of visible light. Since the proposed system captures shape by using IR light, texture and shape can be simultaneously captured because the texture of an object is not affected by the projected light. Moreover, the proposed system can obtain texture without occlusion since the cameras for visible and IR lights are placed coaxially.

We describe a system to capture texture and shape simultaneously by using structured IR light in Section 2. Next, an experimental system is tested to obtain texture and shape in Section 3. The paper is summarized in Section 4.

2 A Sensor for Simultaneous Capturing of Texture and Shape

The sensor proposed in this paper has been designed to separate the wavelength used for capturing shape from the wavelength used for capturing texture. Since visible light is used for texture, a camera can acquire a texture image independent of an actively projected pattern by using IR light. Namely, the proposed system measures three wavelengths of visible light, red, green, and blue for texture, and two wavelengths of IR light as shown in Figure 1. Since the cameras used in our experimental system acquires lights up to the wavelength of 1000nm, we used 880nm and 940nm lights as the central wavelengths of the two IR lights.

An overview of the proposed system, which consists of an IR projector and a multi-band camera, is presented first. Next is a description of the projector and camera. Lastly, we explain the process by which the structured light from the projector and the shape of an object can be obtained with

one shot of the pattern.

2.1 Overview of the Proposed Sensor

The proposed sensor for the simultaneous capture of texture and shape consists of an IR projector and a multi-band camera as shown in Figure 2. The IR projector casts two different patterns by using two near-IR lights of 880nm and 940nm. A multi-band camera used for object observation is equipped with a color CCD to capture RGB visible lights and two monochrome CCDs to capture near-IR lights. Since the monochrome cameras are equipped with band-pass filters of either 880nm or 940nm, the camera observes five wavelengths. For capturing texture, an external light source is used for illumination. In addition, it is necessary to eliminate IR light from the external light to avoid disturbance when capturing the shape.

2.2 Infrared Projector

The IR projector of the proposed system has an internal structure shown in Figure 3. The emitted lights of 880nm and 940nm from the light sources become structured lights by passing through the masks. The lights are projected outside after being combined by a prism. This projection system is similar to a 3-LCD projector.

Infrared LEDs were used as the light source. Since the wavelength range of an LED is narrow, the crosstalk between different wavelengths can be avoided without a band-pass filter. Pattern masks are placed in front of the light sources. Since the proposed sensor captures shape in one shot, it is not necessary to change patterns during acquisition. Thus, the patterns are printed on transparent films.

After the lights passing through the masks are combined by a prism, the structured light is projected with focusing by

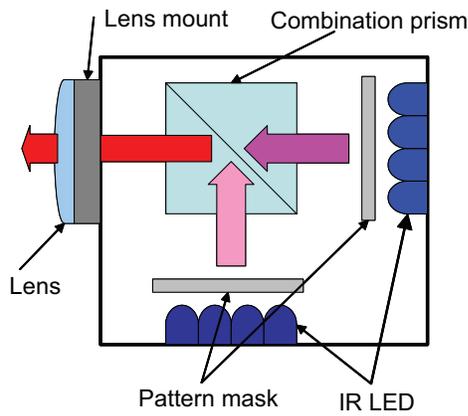


Figure 3. The internal structure of the IR projector of the proposed system.

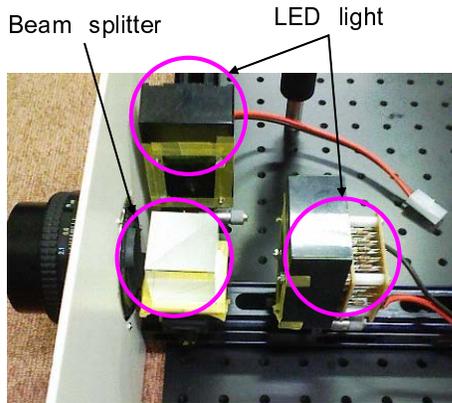


Figure 4. An IR projector experimental system.

an objective lens. To focus the patterns on an object surface, it is necessary to focus on the masks. Our experimental system focuses the light only by the objective lens. Because the prism is placed in the optical path between the lens and mask, a lens with a long back focus is appropriate. Therefore, we used a lens of a medium format camera. Figure 4 shows an IR projector experimental system.

2.3 Multi-band Camera

The multi-band camera process used to capture texture of visible and IR structured light simultaneously is explained in this section. If cameras for texture and shape are placed in different positions, areas of a 3D model cannot be texture-mapped due to occlusion from the texture camera. Therefore, the cameras for visible and infrared lights are placed

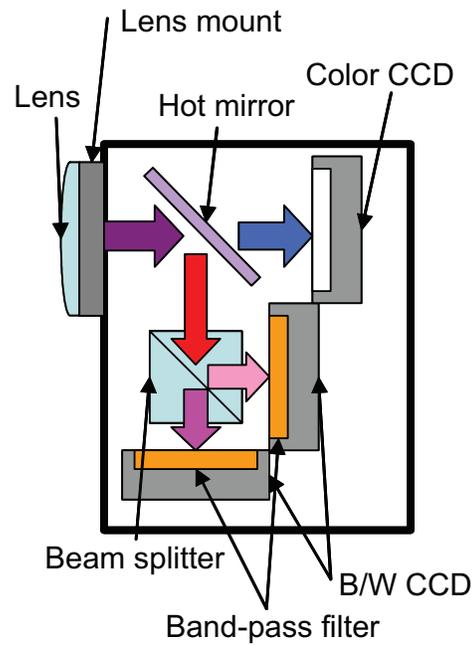


Figure 5. The internal structure of the multi-band camera in the proposed system.

coaxially in the proposed system.

Figure 5 shows the internal structure of the multi-band camera. The incoming light passing through the objective lens is first split into visible and IR lights by a hot mirror which reflects IR and transmits visible light. The transmitted visible light is focused on a color CCD and a texture image is captured. Meanwhile, the reflected IR light is split into two directions by a beam splitter. After passing through either the 880nm or 940nm band-pass filter, each IR light is focused on a monochrome CCD. The patterns projected on an object are then captured by the two CCDs. The optical path lengths from the objective lens to the three CCDs are set to be equivalent. Thus, multi-band imaging can be obtained by the system. A lens of a medium format camera was used to achieve a long back focus due to reasons similar to the IR projector. Figure 4 shows an experimental multi-band camera system.

2.4 Capturing Shape by One-shot Infrared Structured Light

If the projection parameters of a projector and a camera are known, a 3D point can be computed by finding the correspondence between a point from the projector and a point from the camera. Color structured light patterns based on de Bruijn sequence [11, 3] are used as a method to find the correspondence from a single image. A q -ary de Bruijn se-

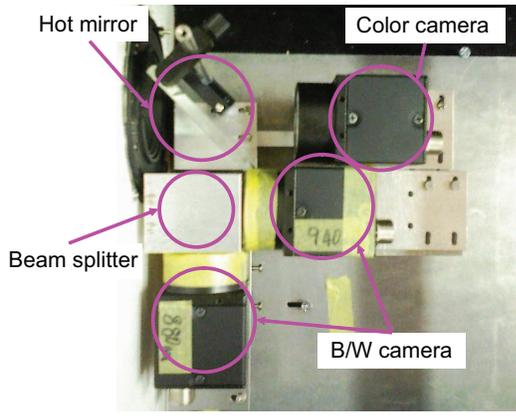


Figure 6. A multi-band camera experimental system.

quence of order n is a sequence of length q^n consisting of an alphabet of size q in which every possible subsequence of length n is present exactly once. If a projected pattern is encoded by two or more symbols distinguished in a camera image, the correspondence between an element in the projected pattern and the observed pattern is uniquely determined by matching subsequences of length n in a de Bruijn pattern.

In [11, 3], RGB color stripes are projected and symbols are assigned to the color changes on the edges of the stripes. However, this assignment may cause detection of a false edge due to the misalignment of the projected patterns between the different wavelengths since the states of three wavelengths simultaneously change. As the projected patterns are aligned manually in the experimental system developed in this paper, the misalignment error is larger than will be caused by a commercial projector. Therefore, assignment which may cause misdetection should be avoided.

The proposed system in this paper uses a single wavelength to detect the position of a symbol. The other wavelength is used only for determining which symbol is detected. Thus detection errors of the symbol position from using multiple wavelengths can be avoided. In the study, we used 880nm light for detecting the position and 940nm light for determining the symbol. The symbols $s = \{s_0, s_1, s_2\}$ for each pixel were designed by using these two wavelengths as follows:

$$\begin{aligned} s_0 &= \neg L_{880} \\ s_1 &= L_{880} \wedge L_{940} \\ s_2 &= L_{880} \wedge \neg L_{940}, \end{aligned} \quad (1)$$

where L_{880} and L_{940} represent the light of the respective wavelength projected on the pixel. The 880nm light is projected as thin lines instead of wide stripes. A symbol is

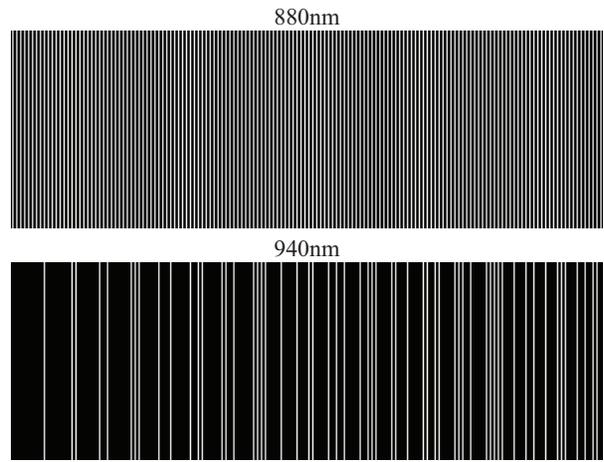


Figure 7. Pattern masks for 880nm and 940nm lights used in the experimental system.

assigned only at the peak of the light intensity where L_{880} is true. Furthermore, dilation for the 940nm light is applied after detecting the lines to absorb the misalignment between 880nm and 940nm lights. A de Bruijn sequence is encoded only by using s_1 and s_2 . No symbol is assigned for a pixel categorized as s_0 . The projected pattern of the experimental system is shown in Figure 7. The pattern is generated with $n = 8$ and $q = 2$. The number of symbols along a scan line is 150.

To find the positions of symbols, the proposed method detects the lines by filtering. First, the image derivative of the input image is computed by a Sobel filter for the x-derivative. The lines are detected by a Canny filter applied to the x-derivative. During this operation, we intervene to detect only lines in which the second-order x-derivative is negative, since the local maximum of intensity is important but the local minimum should be ignored. To detect the position to subpixel accuracy, we compute the position of a local maximum by fitting a quadratic function to the intensity of the input image around the pixels of a detected line.

The detected symbols are matched to the de Bruijn sequence that is projected by the IR projector by using Dynamic Programming matching (DP matching) [11]. Once the correspondence is determined, the distance from the camera to an object surface is computed by triangulation. A 3D mesh model is constructed by connecting neighboring points in the sequence.

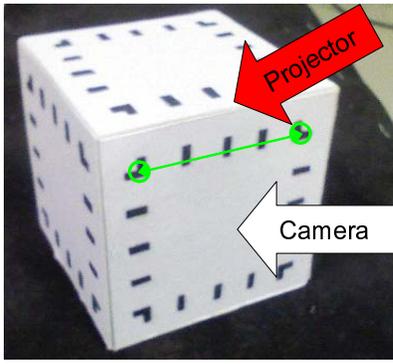


Figure 8. An experiment measuring a known object.

3 Experiments

We first estimate the accuracy of the experimental system by measuring a known object. A cube of known size shown in Figure 8 is captured and the obtained 3D model and the actual object are compared. The errors in the distance between two points and the angle of normal directions of the two planes are used as criteria. For example, the distance between the two green circle shown in Figure 8 is 41.50mm, the distance in the measure model is 42.05mm; the error is 0.55mm. The angle between the two normal vectors is 89.64 degrees. Thus, the experimental system captured two planes perpendicular to each other accurately. The time to generate a model from an image was 60 msec using a Pentium4 3GHz processor. Since the method can obtain a shape in one shot, it can generate 3D models at a high frame rate.

Next, we measured an object of an unknown shape. The IR projector, multi-band camera, and a target object as shown in Figure 9 were used. Figure 10 shows an example of the images captured by the multi-band camera. The top photo is the image acquired by the color CCD. The IR pattern is not captured by the CCD for visible light. Therefore, the texture is simultaneously captured with the shape. The middle and bottom photos are the images captured by the monochrome CCDs with 880nm and 940nm band-pass filters, respectively. Since the wavelengths are completely separated by the band-pass filters, crosstalk between different wavelengths does not occur. The lines in the 880nm image are placed at regular intervals in contrast to the lines in the 940nm image which are placed only for the position of the symbol s_1 . Figure 11 shows the result of the peak detection described in Section 2.4 applied to these two images. The green and red lines correspond to s_1 and s_2 , respectively.

Figure 12 and Figure 13 show the generated model by

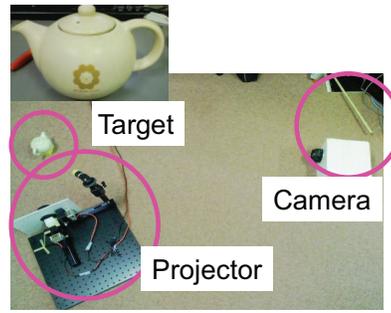


Figure 9. Experimental overview

matching the detected symbols and the projected de Bruijn sequence by DP matching. Figure 12 is a model without texture and Figure 13 is a model after texture mapping. Since the proposed system places the cameras for texture and shape coaxially, it can map the texture for all captured meshes without occlusion.

4 Conclusion

This paper proposed a sensor for the simultaneous capture of texture and shape by projecting structured infrared light. If we simultaneously capture texture and shape of a moving object in real time, the data can be applied to various fields including virtual reality and object recognition. To develop the sensor to achieve simultaneous capturing of texture and shape, there are two problems to overcome: the fast capture of shape and the simultaneous capturing of texture and shape. The first issue is resolved by using a one-shot pattern by applying a method of colored structured lights. The second difficulty is answered by projecting infrared structured light instead of visible light. Since the proposed method uses visible light for texture and infrared light for shape, simultaneous capturing can be achieved. In addition, we developed a system that maps texture on the captured shape without occlusion by placing the cameras for visible and infrared lights coaxially.

There are plans to increase the number of wavelengths in the future when capturing shape to improve the accuracy and robustness. Since the number of symbols in a de Bruijn sequence will increase, intricate shapes can be obtained by reducing the order of the de Bruijn sequence. Furthermore, robust encoding against noise can be adopted by using redundant symbols.

References

- [1] Canesta, Inc. CanestaVision EP Development Kit. <http://www.canesta.com/devkit.htm>.

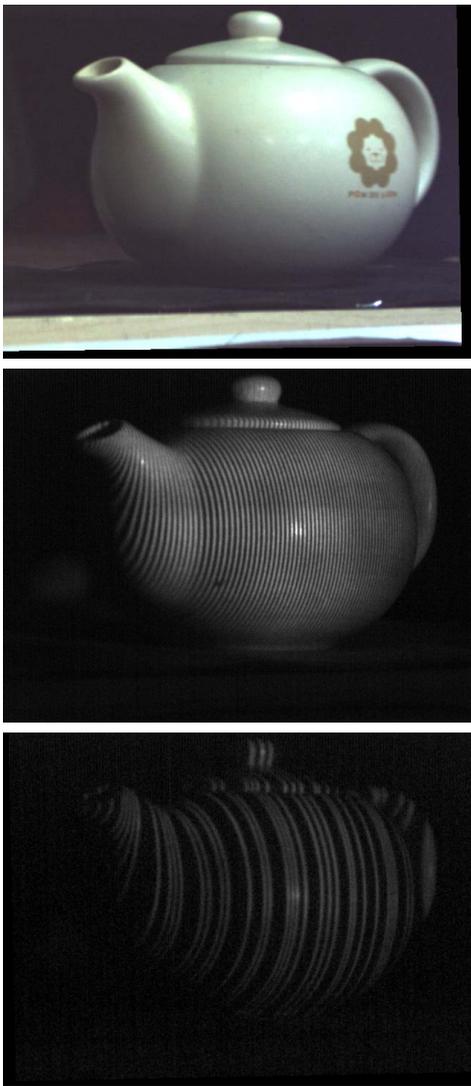


Figure 10. Images captured by the multi-band camera: a texture image (top), an image with a 880nm filter (middle), and an image with a 940nm filter (bottom).

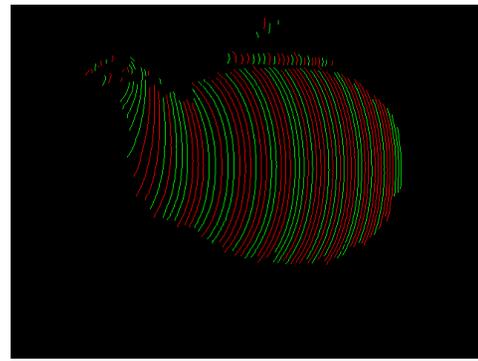


Figure 11. Detected lines

[2] CSEM, Inc. SwissRanger SR-3000. <http://www.swissranger.ch/index.php>.

[3] C. Je, S. Lee, and R. Park. High-contrast color stripe pattern for rapid structured-light range imaging. In *Proc. Eighth European Conference Computer Vision*, pages 95–107, 2004.

[4] T.P. Koninckx and L. Van Gool. Real-time range acquisition by adaptive structured light. *IEEE Transaction Pattern Analysis Machine Intelligence*, 28(3):432–445, 2006.

[5] P.J Narayanan, Peter Rander, and Takeo Kanade. Constructing virtual worlds using dense stereo. In *Proceedings of the Sixth IEEE International Conference on ComputerVision (ICCV'98)*, pages 3–10, January 1998.

[6] J. Pan, P. S. Huang, and F.-P. Chiang. Color-coded binary fringe projection technique for 3-D shape measurement. *Optical Engineering*, 44:3606–+, February 2005.

[7] R. Raskar, G. Welch, M. Cutts, A. Lake, L. Stesin, and Henry Fuchs. The office of the future : A unified approach to image-based modeling and spatially immersive displays. In *Proc. ACM SIGGRAPH*, 1998.

[8] S. Rusinkiewicz, O. Hall-Holt, and M. Levoy. Real-time 3d model acquisition. In *Proc. SIGGRAPH*, pages 438–446, 2002.

[9] T.Matsuyama, X. Wu, T. Takai, and T. Wada. Real-time dynamic 3d object shape reconstruction and high-fidelity texture mapping for 3d video. *IEEE Trans. on Circuits and Systems for Video Technology*, CSVT-14(3):357–369, 3 2004.

[10] M. Waschbusch, S. Wurmlin, D. Cotting, F. Sadlo, and M. Gross. Scalable 3d video of dynamic scenes. In *The Visual Computer (Proceedings of Pacific Graphics 2005)*. Springer, 2005.

[11] L. Zhang, B. Curless, and S. Seitz. Rapid shape acquisition using color structured light and multi-pass dynamic programming. In *Proc. First International Symposium 3D Data Processing Visualization and Transmission*, pages 24–36, 2002.

[12] Z. Zhang, D. Zhang, X. Peng, and X. Hu. Color texture extraction from fringe image based on full-field projection. *Optical Engineering*, 42:1935–1939, July 2003.

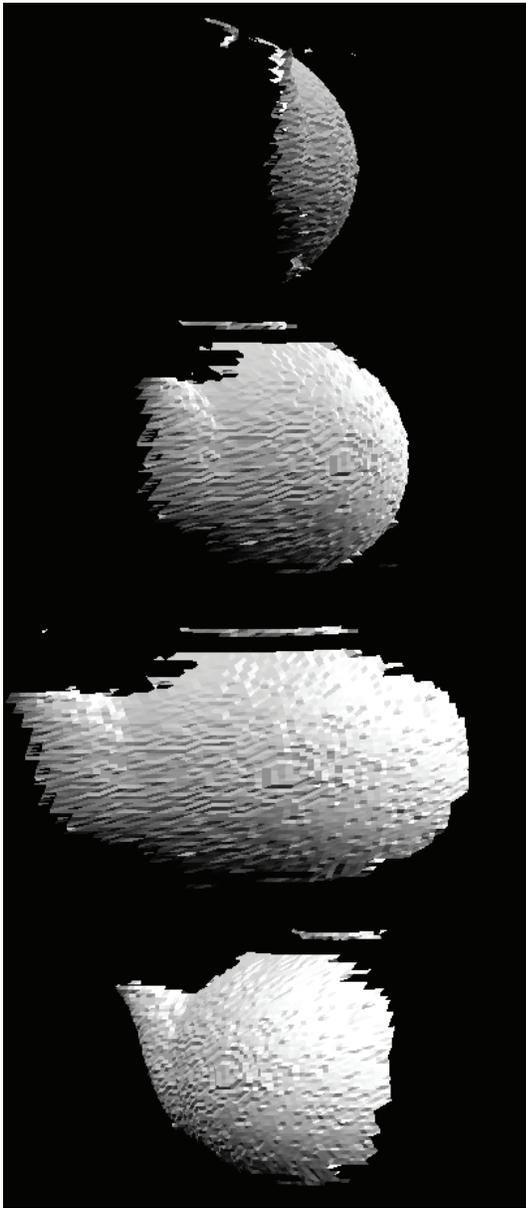


Figure 12. Generated 3D mesh sample model without texture.

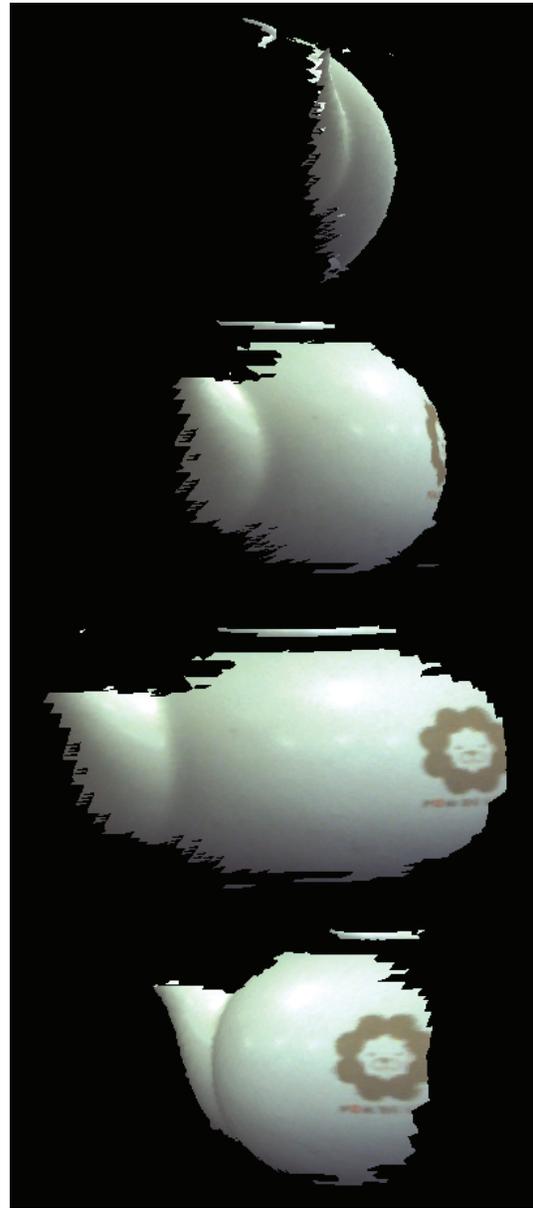


Figure 13. Generated 3D mesh sample model with texture.