

REAL-TIME LIGHT FALL-OFF STEREO

Miao Liao, Liang Wang, Ruigang Yang

University of Kentucky
{mliao3,lwangd,ryang}@cs.uky.edu

Minglun Gong

Memorial University of Newfoundland
gongml@gmail.com

ABSTRACT

We present a *real-time* depth recovery system using Light Fall-off Stereo (LFS). Our system contains two co-axial point light sources (LEDs) synchronized with a video camera. The video camera captures the scene under these two LEDs in complementary states (e.g., one on, one off). Based on the inverse square law for light intensity, the depth can be directly solved using the pixel ratio from two consecutive frames. We demonstrate the effectiveness of our approach with a number of real world scenes. Quantitative evaluation shows that our system compares favorably to other commercial real-time 3D range sensors, particularly in textured areas. We believe our system offers a low-cost high-resolution alternative for depth sensing under controlled lighting.

1. INTRODUCTION

Many applications, such as robot navigation and augmented reality, require real-time range information in a dynamic environment. In this paper we developed a novel system that uses the inverse square law for light intensity to estimate depth information. Based on the formulation in [1] our system uses a single camera to capture a scene under two different lighting conditions: one illuminated by a near point light source and the other by a far one. Per-pixel depth is solved based on the pixel intensity ratio and the distance between the two lights, without the need for matching pixels.

The main contribution of this paper is a novel depth range system that can generate a VGA (640×480) resolution depth map at 30Hz. Quantitative accuracy evaluation shows that our system compares favorably to other commercial 3D range sensors, particularly in textured areas. In addition, our system is made of commodity off-the-shelf components, offering an inexpensive solution to real-time, high-resolution, video-rate range sensing.

1.1. Related work

Recovering 3D shapes from images is one of the fundamental tasks in computer vision. While there is a plethora of techniques to achieve this, we will focus on the methods that are capable of generating real-time depth maps with live input.

The most common way of computing depth map is to use stereovision. Recently, several stereo methods have been developed to exploit the processing power of modern graphics hardware [2, 3, 4, 5]. Although tremendous progress has been made in stereovision, the fundamental correspondences problem remains difficult in real-world applications.

The correspondence problem can be greatly simplified with active illumination. Many real-time structured light scanners (e.g. [6, 7, 8]) can obtain high quality results. These systems typically require multiple frames, which limit the object motion, and have difficulty with high-frequency textures.

New range sensors have also been developed using shuttered light-pulse (SLP) technologies [9]. 3DV Systems, Ltd. and Canesta, Inc. [10, 11] have both developed SLP technologies. However They are either very expensive (e.g. over fifty thousand US dollars for a 3DV system) or have limited resolutions (e.g., 64×64 for a Canesta sensor).

Our system builds on the algorithms described in [1] which use the inverse-square law to recover 3D shape information. Compared to previously developed techniques, our approach only requires two images and the use of commodity off-the-shelf components provides an inexpensive way to produce high-resolution depth maps. More importantly, experiments show that our system provides better depth maps that are independent of scene texture.

2. METHODS

2.1. Light Fall-off Stereo

It is well known that the intensity of light emitted from a source of constant intrinsic luminosity falls off as the square of the distance from the object. Under this *inverse square* law, the observed intensity of a surface point p can be formulated as:

$$I_p = \frac{L(\theta)}{r_p^2} \rho(\theta, \phi), \quad (1)$$

where $L(\theta)$ is the light radiance along incident direction θ . r_p is the distance between the light source and p . $\rho(\theta, \phi)$ is the BRDF (Bidirectional Reflectance Distribution Function) of surface point p and ϕ is the viewing direction.

Now if the light source is moved away from point p along the direction θ by amount Δr . The observed intensity of surface point p under the new setting becomes:

$$I'_p = \frac{L(\theta)}{(r_p + \Delta r)^2} \rho(\theta, \phi) \quad (2)$$

Computing the ratio of the above two equations makes the ratio between I_p and I'_p related only to the depth:

$$\frac{I_p}{I'_p} = \frac{(r_p + \Delta r)^2}{r_p^2} \implies r_p = \frac{\Delta r}{\sqrt{I_p/I'_p} - 1} \quad (3)$$

One critical requirement for the above formulation is that the incident light direction remains the same during the illuminator's movement. An occluder is introduced in [1] to make this possible. As illustrated in Figure 1, an opaque board with a small aperture in the center is placed at the near lighting position. The plane on which the occluder lies is referred as near lighting plane. The point light source is first placed at the position of aperture, S , and illuminates the entire scene. Then the light source is moved onto a second lighting plane, which is parallel to the occluder and referred as far lighting plane. The light translates on the second plane and illuminates part of the scene each time. Consider an arbitrary point on the surface (e.g., q), it was illuminated first by a point light source at S , then by a point light source at T' that goes through S . Therefore its incident light direction remains unchanged and Equation 3 can be applied to estimate the range of q .

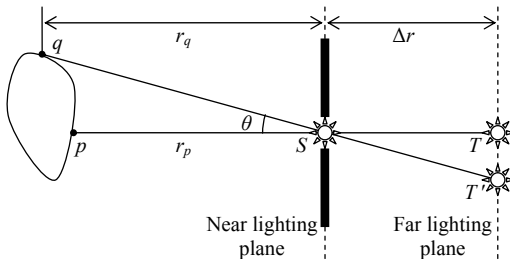


Fig. 1. The setup for recovering the depth map of a scene.

It is further argued in [1] that when the scale of the objects is much smaller than the distance of the light, the variation in incident lighting direction can be ignored. That is, the occluder can be removed and one can approximate the illumination effect obtained at position T' using the one obtained at position T . Our real-time LFS system adopts the same approximation. From the image pair captured under lighting positions S and T the per-pixel depth value can be recovered.

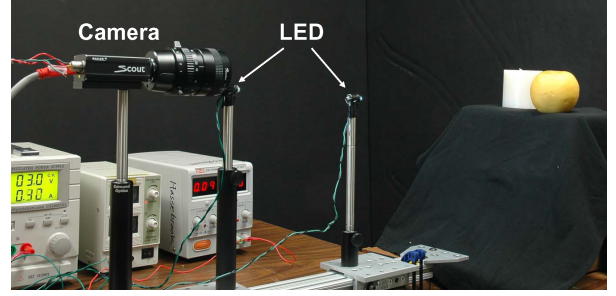


Fig. 2. Our experimental setup consists of two LED light sources and a video camera on a linear translation stage.

3. PROTOTYPE SYSTEM IMPLEMENTATION

3.1. Experimental setup

Our experimental setup consists of two 3W LED light sources and a video camera on a linear translation stage. The LEDs and video cameras are co-axial. The LEDs occlude a small part of the scene in the camera image, for which we mask out. The camera can capture 640×480 gray-scale images at 60Hz with progressive scan. The camera responds linearly to light intensity and 8-bit images are used throughout.

The two LEDs are synchronized with the camera's shutter. The camera generates a TTL signal when it opens its shutter. With each shutter pulse, the LEDs toggle their on/off state. LEDs can be switched on or off in the order of 100 nanoseconds. The LEDs heat up when powered. During the "start-up" process, the device's temperature rapidly increases, and the LED's forward current decreases until it reaches a steady state, at which point we start the capture. We have verified that the light output is very stable once the LED reaches its steady state.

While an LED is an excellent point light source, its spatial distribution of radiance is not uniform. As a result, we must radiometrically calibrate the two LEDs. The procedure is straightforward. Before running the system, we turn on the two lights in turn to illuminate a piece of white paper covering the entire field of view of the camera. The calibration object is captured by the camera under the two lighting conditions. From the two images I_n and I_f , we measure the ratio (R) of the corresponding pixel values, that is

$$R(u, v) = L_n(\theta)/L_f(\theta) = (d_n/d_f)^2 * I_n(u, v)/I_f(u, v), \quad (4)$$

where u, v are the pixel coordinates, d_n, d_f are distance between calibration object and the near and far light respectively. Referring to equation 1, under our assumption that the incident lighting direction change is small enough, intensity variation cannot be explained by the inverse square law is attributed to the light radiance function.

3.2. Run-time algorithm

The run-time system consists of two parallel threads, one is for image capture and the other is for depth computation and display.

The capture thread waits for camera images transferred via the IEEE1394 bus and alternatively stores them into the near and the far image buffers.

In the depth computation thread, image(I_f) from far light is first corrected by calibration ratio R ,

$$I_c(u, v) = I_f(u, v) * R(u, v), \quad (5)$$

where I_c is corrected image. Then we plug $I_p = I_n(u, v)$ and $I'_p = I_c(u, v)$ into equation 3 to compute the depth of pixel (u, v).

Before the computation, we exclude those pixels that will potentially give bad results. Those pixels include saturated ones in either the near image or the far image. Saturated pixels (especially highlight areas) are usually not real measurements of light intensity; they are likely to result in inaccurate depth estimates. The pixels with intensities below a certain threshold are also excluded, because these pixels are either background or reside in shadow areas. Furthermore, low intensity values are more sensitive to noise. For those bad pixels, we simply set them to black in the depth map. This is why black holes are occasionally present in the depth map, likely the results of surface highlights and shadows. Finally, the depth map is smoothed by a mean filter.

Since graphics cards are excellent for parallel image processing, the entire depth computation pipeline is implemented on the graphics processing unit (GPU). In this case, the captured images are directly transferred to two textures on the graphics board.

4. EXPERIMENT AND RESULTS

In our experimental setup, the typical distance between near and far lights is 85mm. The valid working volume is determined by the dynamic range of the camera. With 8-bit images, it is approximately 365mm-1000mm (distance to the near light) with a depth resolution of about 4mm.

4.1. Quantitative evaluation

We first evaluate the quality performance of our system by comparing it with two other commercially available live range sensors. The first one is Canesta range sensor which is able to generate low resolution (64×64) range maps at video frame rate. The other is the Z-mini from 3DV Systems, Ltd. that can provides high resolution (maximum 640×480) live range maps.

As shown in figure 3 the target object in the first experiment (first row) is a piece of white paper glued onto a planar surface. This paper can be regarded as a perfect Lambertian

reflector with constant surface albedo. In the second experiment (second row) the white paper is replaced by a piece of paper containing rich textures. The object is carefully placed so visually the principle axis of these sensors are perpendicular to the plane and go through the plane's geometric center.

One thing worth noting is that these three sensors have different fields of view, resolutions and their recovered range maps are not within the same coordinate system. These limitations make a metric comparison with ground truth difficult. To warrant a fair evaluation, we first normalize their output depth values to a uniform space. It is done by calculating a scale factor so that the sample mean of the depth values is normalized to 0.5. Afterwards we apply a plane fitting algorithm to each sample data and compute their mean square deviation. Clearly smaller variance implies better reconstruction quality.

The recovered 3D shapes and error rates of these sensors are presented in Figure 3 and Table 1 respectively. Our raw range map is processed with a 5×5 smoothing filter to reduce high frequency noise resulting primarily from the CCD camera. In general, when the sample target is textureless, all three sensors yield satisfactory results (the Canesta sensor returns a single-colored depth map). However, when the target contains non-uniform surface albedo, our system outperforms the other two. It is not surprising given the fact that our depth values are recovered from the ratio of two images, effectively cancel out the surface albedo's influence. Conversely, the reconstruction model adopted by SLP sensors suffers from bias as a function of object intensity [12].

	Canesta	Z-mini	LFS
newspaper	0.0741	0.0260	0.0101
white paper	0	0.0166	0.0021

Table 1. Numerical errors of depth recovered by different range sensors.

4.2. Live system

Figure 4 shows some live images from our system. Shadow areas are automatically detected and masked out during the depth map computation process. The scene contains objects with different shapes and reflectance properties. The resulted depth map is fairly accurate despite the slightly non-Lambertian surface reflectance.

Our camera captures at 60fps and the off-line depth computation can achieve 60fps on a Geforce 8800 graphics card from NVIDIA. But given that two images are required to generate one depth map, our system's overall speed performance is 30fps.

5. CONCLUSION

In this paper we presented a novel system that can generate real-time depth maps. Our system, based on the formulation

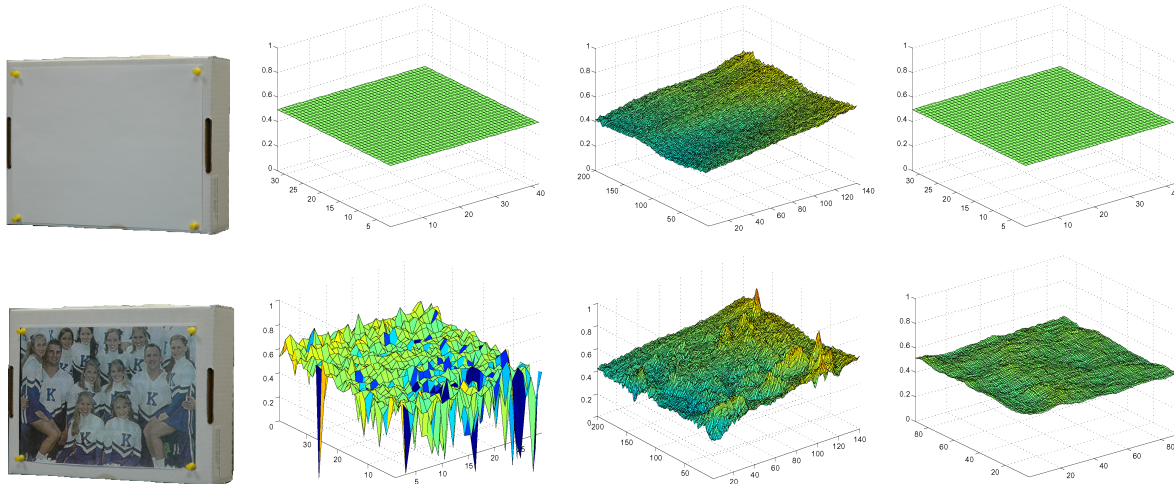


Fig. 3. Depth recovered by different sensors. From left to right: sample scenes, 3D plots of the recovered scene depth from Canesta, Z-mini and LFS. The mean depth is normalized to 0.5.

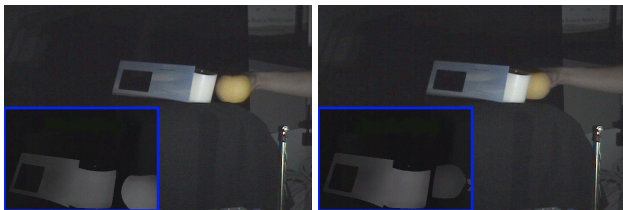


Fig. 4. Some snapshots of our real-time system results. The insets show the depth maps.

in [1], takes two images under different lighting conditions to estimate the range for each pixel, without the need for matching. Compared to commercial 3D range sensors, it is more robust to textured areas (when the object remains static or the object motion between two frames is smaller than one pixel). Our prototype is made from off-the-shelf, low cost components with a simple computation model. It can be used in low-cost embedded systems. We believe our system provides a viable alternative for 3D range sensing under controlled lighting.

6. REFERENCES

- [1] M. Liao, L. Wang, R. Yang, and M. Gong, “Light fall-off stereo,” *Proceedings of CVPR*, 2007.
- [2] R. Yang and M. Pollefeys, “Multi-resolution real-time stereo on commodity graphics hardware,” *Proceedings of CVPR*, pp. 211–218, 2003.
- [3] C. Zach, A. Klaus, and K. Karner, “Accurate dense stereo reconstruction using graphics hardware,” *Proceedings of EUROGRAPHICS*, pp. 227–234, 2003.
- [4] N. Cornells and L. Van Gool, “Real-time connectivity constrained depth map computation using programmable graphics hardware,” *Proceedings of CVPR*, pp. 1099–1104, 2005.
- [5] Minglun Gong and Yee-Hong Yang, “Near real-time reliable stereo matching using programmable graphics hardware,” *Proceedings of CVPR*, pp. 924–931, 2005.
- [6] Olaf Hall-Holt and Szymon Rusinkiewicz, “Stripe boundary codes for real-time structured-light range scanning of moving objects,” *Proceedings of ICCV*, 2001.
- [7] L. Zhang, B. Curless, and S. Seitz, “Rapid shape acquisition using color structured light and multi-pass dynamic programming,” *Proceedings of 3DPVT*, 2002.
- [8] Thomas P. Koninckx and Luc Van Gool, “Real-time range acquisition by adaptive structured light,” *PAMI*, vol. 28, pp. 432–445, 2006.
- [9] H. Gonzalez-Banos and J. Davis, “Computing depth under ambient illumination using multi-shuttered light,” *Proceedings of CVPR*, pp. 234–241, 2004.
- [10] G. Yahav and G. Iddan, “Optimal ranging camera,” *United States Patent; no. US 6,057,909*, May 2002.
- [11] C. Bamji, “Cmos-compatible 3-dim. image sensor ic,” *United States Patent, no. US 6,323,942 B1*, November 2001.
- [12] J. Davis and H. Gonzalez-Banos, “Enhanced shape recovery with shuttered pulses of light,” *Proceedings of IEEE International Workshop on Projector-Camera Systems*, 2003.