Multi-view shape reconstruction on CAVE System

Junyoung Park[*]
Computer Science Dept.
Tokyo Inst.of Technology

Hamid Laga[†]
Global Edge Institute
Tokyo Inst.of Technology.

Masayuki Nakajima[‡]
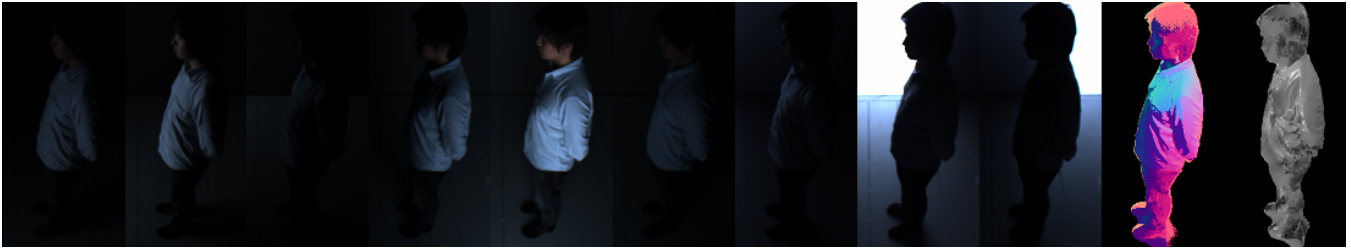Computer Science Dept.
Tokyo Inst.of Technology

Figure 1: *Our CAVE System acquires a multi-view image set of a target under 9 different illumination. From a image set, we reconstruct a surface shape of the target.*

1 Introduction

Nowadays, many applications, ranging from medicine to video games, require accurate 3D geometry and motion of human actors. Existing commercial motion capture systems record only the status of the markers that are attached to the human body. These provide a sparse representation of the true shape and therefore post-processing steps are necessary to recover the full-body movement. In this paper, we propose a new shape acquisition technique for 3D surface reconstruction. The capture system uses eight synchronized video cameras and 3 light projectors to record a subject. The light projectors are used in a CAVE-like setup to provide a illumination pattern. The proposed 3D reconstruction algorithm recovers accurate 3D geometry by integrating the depth and normal maps into a single linear system which can be solved efficiently. Depth maps are calculated using a space-time multiview stereo matching on the acquired set of images after projecting on them some illumination patterns. Surface normal map is estimated using an example-based photometric stereo technique[Hertzmann and Seitz 2005] that uses illumination patterns. We show also that an efficient GPU implementation of the stereo matching and photometric stereo algorithms results in a speedup of order 20 compared to CPU implementation.

2 Hardware Setup and Illumination Design

Hardware Setup   We use CAVE system as our measurement stage. Our CAVE system consists of 3 rear projection screens, and 3 DLP Projectors. Each of the projector projects light to each rear projection screen. Every screen is fully syncronized and controllable during camera capture process. To avoid the temporal changes of projected light color, we removed rotary color filter from projectors in advance. Removing rotary color filter also gives us more strong light intensity. We captured a shape of subject using fully syncronized 8 firewire cameras(PointGrey Flea2, which can capture 30Hz with 1024×768 resolution). 8 cameras are placed on around the CAVE screen. Figure[2] shows our setup.

There are some benefits to use CAVE system as a measurement stage. At first, we can get good diffused surface light from every projection screen. It reduce the occlusion problem from directional light. And the large surface light guarantees good signal to noise
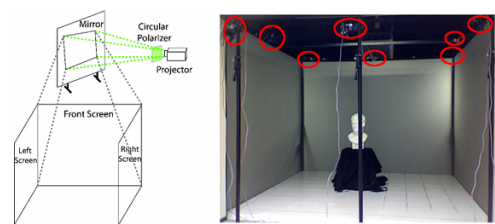
---

[*]e-mail: junyoung@img.cs.titech.ac.jp

[†]e-mail: hamid@img.cs.titech.ac.jp

[‡]e-mail: nakajima@img.cs.titech.ac.jp

Figure 2: Left *: Our CAVE setup with a projector which projects light to front screen. The left and the right screens have same setup as the front screen.* **Right** *: Our CAVE System as a measurement stage. Eight cameras are placed on around the CAVE screen.*

ratio than a bright point light source even the dynamic range of camera is narrow. And because we use a projector as a light source, every surface light can be fullly controlable in temporally and spatially. Therefore we can make fully controlable light patterns in the measurement stage.

Prior to the measurement, We calibrated our cameras in a geometric way. External and internal camera parameters are calculated using the multi-camera calibration technique. In our hardware setup, one pixel in the camera project approximately 1.5mm around the subject area.

Time-multiplexed illumination   We employed a binary light pattern to measure the shape of target. We divided each screen in 3 vertical areas, and illuminated one after another. Because we synchronized our screens and cameras, we know the light condition on each captured image. Figure[3] shows our illumination patterns.
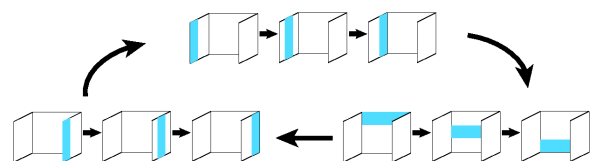

Figure 3: *Our Illumination Patterns on 3 CAVE Screen. Each screen illuminated in 3 different area by turns, and 3 screen illuminated one by one.*

3 Single View Surface Reconstruction

3.1 Depth Estimation

We ready a set of images on our time-multiplexed illumination pattern. The images are rectified from camera calibration. To estimate depth map from a image set, we extended a stereo matching technique of [Goesele et al. 2006] to a spacetime stereo matching technique which considers not only the spatial parameters but also the temporal parameters. Every depth map value has a confidence parameter which represents the accuracy of the depth value.

3.2 Normal Estimation

We use example-based photometric stereo approach[Hertzmann and Seitz 2005] to estimate surface normals. We use white diffuse sphere as a reference object and calculate the normal direction on every camera pixel. We capture the sphere on full illumination, and find a center of circle on image plane using hough transform. From the circle center rays on every camera, we can calculate a midposition of sphere center using least squares fitting, then reproject it to each image plane. The reprojected sphere center has a normal which is an opposite to camera ray direction. When we calculate a normal of a point in sphere image, we first find a grazing angle point which intersects between silhouette of sphere and a line which passes through the sphere center and target pixel. Then we can calculate the sphere normal on the grazing angle point, and interpolate between the grazing angle normal and sphere center normal using slerp(Figure.4).
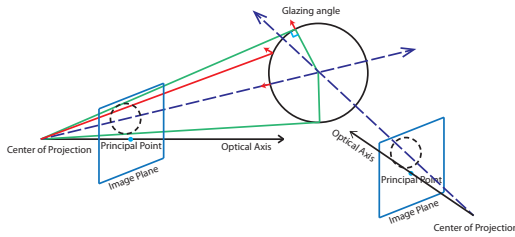

Figure 4: *Normal direction of sphere on perspective view.*

We capture a reference object on our illumination pattern to build a lookup table between the normal and a normalized observation vector(OV). The OV is a set of intensities observed at a pixel over the illumination pattern. To estimate a surface normal of a target object, we ready a normalized OV under same illumination pattern, search for the closest vector on the look up table, then retrieve the related normal.

3.3 Surface Reconstruction

In the previous section, we uses a multi-view stereo mathcing and photometric stereo to estimante the depth map and the normal map of the target object respectively. In general, multi-view stereo matching suffers to estimate a corresponding point on smooth surface. On the other hand, photometric stereo suffers to estimate a normal map on noisy textured surface. To reconstruct a surface of target, we have to consider to recover these weak points. We will transcribe the value of depth map and normal map on pixel (i, j) as $D(i, j)$ and $N(i, j)$ respectively.

We define a vector which starts from a 3D point on a pixel(i,j) and to other 3D point on the neighbor pixel(i+1, j). the vector is perpendicular to the normal of the pixel(i,j).

$$(\vec{r}_{i,j} d_{i,j} - \vec{r}_{i+1,j} d_{i+1,j}) \cdot \vec{N}_{i,j} \tag{1}$$

Similarly, in the vertical direction:

$$(\vec{r}_{i,j} d_{i,j} - \vec{r}_{i,j+1} d_{i,j+1}) \cdot \vec{N}_{i,j} \tag{2}$$

If we propagate these two constraints for all of the pixels, we get a twice of constraints than the number of unknown parameters $d_{i,j}$.

In addition, we define more constraints from depth map information. We select a high accuracy depth points on depth map, and insert the depth value as constraints.

$$d_{i,j} = D_{i,j} \tag{3}$$

Then we solve the overdetermined linear system using QR decomposition. The solution is a least squars of these constraints. Because our linear system designed to use depth information directly, this linear equation is not an ill posed problem, so we can directly compose the multi-view surface with a littel effort.

4 Conclusions

4.1 GPU Implementation

We implemented the depth / normal estimation methods on GPU using CUDA. Both of them are well adapted on parallel computing. Comparing with non-parallel calculation, the depth estimation methods are speeds up approximately 20 times faster, and normal estimation methods are 50 times faster than non-parallel calculation. We implemented the surface reconstruction method on not GPU but CPU because it uses largely sparse matrix calculation. We run our algorithm on a quad core PC(Intel Core i7 930, 2.8GHz) and a nVidia geForce GTX260 Graphics Card as a GPU. It takes around 5 minute to reconstrct one set of images.

4.2 Result and future work

Comparing to the state of the art surface reconstrction method of [Vlasic et al. 2009], Our surface reconstruction method is more simple and reliable because we does not handle the surface with any assumption.

In the future work, we will evaluate our method comparing with other surface reconstruction methods.

References

GOESELE, M., CURLESS, B., AND SEITZ, S. M. 2006. Multiview stereo revisited. In *CVPR '06: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society, Washington, DC, USA, 2402–2409.

HERTZMANN, A., AND SEITZ, S. M. 2005. Example-based photometric stereo: Shape reconstruction with general, varying brdfs. *IEEE Trans. Pattern Anal. Mach. Intell. 27*, 8, 1254–1264.

VLASIC, D., PEERS, P., BARAN, I., DEBEVEC, P., POPOVIĆ, J., RUSINKIEWICZ, S., AND MATUSIK, W. 2009. Dynamic shape capture using multi-view photometric stereo. In *SIGGRAPH Asia '09: ACM SIGGRAPH Asia 2009 papers*, ACM, New York, NY, USA, 1–11.