Partial Scene Reconstruction using Time of Flight Imaging

Yuchen Zhang and Hongkai Xiong

Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai, China

ABSTRACT

This paper is devoted to generating the coordinates of partial 3D points in scene reconstruction via time of flight (ToF) images. Assuming the camera does not move, only the coordinates of the points in images are accessible. The exposure time is two trillionths of a second and the synthetic visualization shows that the light moves at half a trillion frames per second. In global light transport, direct components signify that the light is emitted from a light point and reflected from a scene point only once. Considering that the camera and source light point are supposed to be two focuses of an ellipsoid and have a constant distance at a time, we take into account both the constraints: (1) the distance is the sum of distances which light travels between the two focuses and the scene point; and (2) the focus of the camera, the scene point and the corresponding image point are in a line. It is worth mentioning that calibration is necessary to obtain the coordinates of the light point. The calibration can be done in the next two steps: (1) choose a scene that contains some pairs of points in the same depth, of which positions are known; and (2) take the positions into the last two constraints and get the coordinates of the light point. After calculating the coordinates of scene points, MeshLab is used to build the partial scene model. The proposed approach is favorable to estimate the exact distance between two scene points.

Keywords: global light transport, direct components, time of flight, partial scene reconstruction, ellipsoid focus

1. INTRODUCTION

Scene reconstruction is one of the research focus of computer vision and widely applied in various fields. It aims to reconstruct watertight 3D models from calibrated photographs of a realistic object. Although many algorithms have been developed for this problem, efforts still have to be made to achieve a 3D modeling with both high efficiency and high quality. The previous scene reconstruction methods can be classified into four categories: 3D volumetric approaches[1], surface evolution techniques[2], feature extraction and expansion algorithm[3], and depth map based methods[4]. The first approach relates to 3D volumetric techniques that use the visual hull of the scene, a constraint on the topology of the scene, to infer occlusions. In the second approach, we utilize silhouette constraints to shrink to the structure. Feature based algorithms, the third approach, extract feature points to produce point clouds. The last approach called depth map based methods uses binocular stereo to form a depth map, which is exploited to calculate the point clouds.

In traditional reconstruction methods, multi-view images are utilized to reconstruct the scene. The images are captured by multiple cameras or a camera in different views. In order to produce point clouds between two adjacent images, they are required to have enough overlap. Depth map based methods are the most used in 3D reconstruction, and it is a huge challenge to produce a precise and continuous depth map. Based on feature extraction, wrong matches may occur and sometimes point clouds are not enough. The camera calibration is also a big problem. Using cameras stable in a cycle, the calibration can be completed. Much time is used to measure the cameras' coordinates. If the images are captured by a moving camera, errors may happen in the later calibration. It is a huge challenge in scene reconstruction to find a technique that is applied to capture the images easily and calibrate the camera preciously.

Femto Photography consists of femtosecond laser illumination, picosecond-accurate detectors and mathematical reconstruction techniques. The lastest technique can record global light transport situation, which is composed of direct and indirect components. The light source is a Titanium Sapphire laser that emits pulses

Further author information:
Yuchen Zhang: E-mail: slyzzhangyc@163.com
Hongkai Xiong: E-mail: xionghongkai@sjtu.edu.cn

Optoelectronic Imaging and Multimedia Technology III, edited by Qionghai Dai, Tsutomu Shimura,
Proc. of SPIE Vol. 9273, 927315 · © 2014 SPIE · CCC code: 0277-786X/14/$18
doi: 10.1117/12.2071937

at regular intervals every 13 nanoseconds. These pulses illuminate the scene, and also trigger the picosecond accurate streak tube which captures the light returned from the scene. The light transport is composed of constituent direct, subsurface scattering, and interreflection components[5]. The direct component means that the light transfers from the source directly. By calculating the propagation time, the sum distance of the scene point from source and the scene point to camera center can be obtained. In 3D computer graphics, global illumination is composed of a group of algorithms that are meant to add more realistic lighting to 3D scenes. Global light transport is utilized by structured light methods when decomposing multi-bounce light transport into individual bounces[6,7]. Geometric information is extracted from second-bounce light transport[8]. A method that separates high-frequency direct transport from low-frequency global transport requires as little as a single photo under structured lighting[9]. The global separation techniques are useful in many applications for precisely capturing scene depth information through global illumination[10–12]. Geometry acquisition that accounts for both interreflections and subsurface scattering light transport effects within a scene can be extracted from designed structured illumination patterns[13]. Such algorithms take into account not only the light directly from a light source, which is called direct illumination, but also the light rays reflected by other surfaces in the scene, which is called indirect illumination. In [1], a separation method based on ToF imaging is proposed. Usually, ToF images are exploited in analysing the composition of global illumination, and there is little combination of ToF images and scene reconstruction. The images used in traditional reconstruction algorithms are captured in different views, while ToF images are captured at different time. Since ToF images contain information of global illumination, they can be conveniently used for partial scene reconstruction.

This paper aims to reconstruct partial 3D model from ToF images of a scene. Unlike the previous reconstruction methods, the scene reconstruction does not require other view images. The direct component, marked as the white area in Fig.1a, is extracted from ToF images. The geometrical relationships lead to two constraints. Considering that the camera and virtual light source are supposed to be two focuses of an ellipsoid, the distance that light travels between the two focuses and the scene point is a constant. The focus of the camera, the scene point and the corresponding image point are supposed to be collinear. Using these two criteria, the coordinates of scene points can be yielded. These two constraints also help finish the calibration and gain the initial parameters.

2. CALCULATE THE COORDINATES OF SCENE POINTS

This section can be separated into three parts: (1) to extract the direct component of global illumination; (2) to calculate the coordinates of the camera and the virtual light source; (3) to obtain the coordinates of the scene points.

2.1 Scene reconstruction using direct component analysis

The ToF images are stored as x-y-t volumes. x and y correspond to pixel locations in a conventional image and t corresponds to time. An image at a given time t is produced by an x-y slice of the volume. The time profile P(t) returns the pixels' intensity as a function of time. The datasets[14] are from MIT Media Lab.

A ToF image can record multiple lights that propagate in the scene. The time profile P(t) within the ToF image, which is observed by the instrument, is a sum of time profiles. It consists of the direct time profile D(t), subsurface scattering profile S(t), and interreflection profile I(t). Each of them has different features and can be extracted by different methods. Interreflection is the reflection of light from a surface such that an incident ray is reflected at many angles rather than at just one angle as in the case of specular reflection. Subsurface scattering, also known as subsurface light transport, is a mechanism of light transport in which light penetrates the surface of a translucent object. It is scattered by interacting with the material, and exits the surface at a different point.

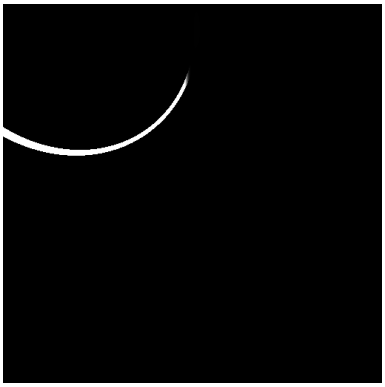
Among all the light paths that pass through a pixel at different time t, there is a direct light path traveling the shortest distance, which consists of only one bounce. The first impulse observed in the time profile P(t) is the direct component. The algorithm separates direct and global time profiles by localizing the direct component within the time profile P(t), and extracting the direct component D(t) by imposing smoothness constraints on the global component G(t). As some pixels within a ToF image may have no direct components, time profiles without a direct component are rejected by thresholding the first peak intensity. The algorithm refers to Algorithm 1.

Algorithm 1 Direct component extraction

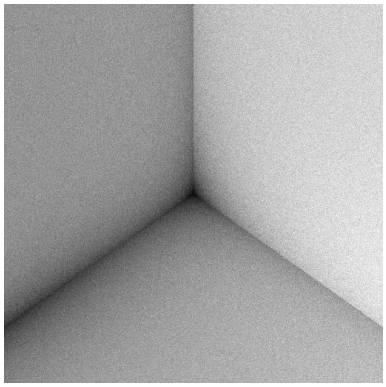
Require: the time profile P(t)

1. Compute the derivative of the time profile P'(t)
2. The start of the direct component is calculated by solving t_{start} = argmin$_t P'$(t)> α. Typically, the value of α is 0.05max$_t P'$(t)
3. The apex of the direct component is obtained by solving t_{middle}= argmin$_t P'$(t)<β subject to t>t_{start}. Typical the value of β is 0.0001
4. The end of the direct component is computed by $t_{end} = t_{start} + 2(t_{middle} - t_{start})$.
5. Time profile values between t_{start} and t_{end} are smoothly interpolated to generate the global component G(T).
6. Direct component D(t)= P(t)-G(t).

Binary images denote the direct component which is extracted from ToF images. In each time t_0, the area with direct components is white shown in Fig.1a. Fig.1b shows the structure of scene.



(a) The binary image showing the direct component



(b) The structure of the scene

Figure 1. The binary image and the structure of scene.

2.2 Initial parameters calculation

It should be mentioned that the acquisition of scene points' coordinates requires measuring the time elapsed between emitting a pulse of light onto a point in a scene, and the backscattered light returning to the sensor. The laser illuminates an unknown point in the scene and produces a virtual light source. The scene is illuminated by this virtual light source and sufficient information provided by the resulting direct component is utilized to recover the initial parameters of the device.

The laser intersects a (unknown) point L=(L_x,L_y,L_z) in world space at a (unknown) time t_0 in the scene, with the camera centered at point C=(0,0,0). If the camera receives the first bounce of light at time t at camera pixel (x,y), the direct reflection point P=(P_x,P_y,P_z) must satisfy the following formula:

$$||P - L||_2 + ||P - C||_2 = k(t - t_0) \tag{1}$$

where k denotes a constant that accounts for the light velocity in world coordinate system, t_0 is the moment that the light is emitted from the virtual light source, t is the moment that the light is received by the camera.

Multiple points with the same depth are used to calculate L, t_0, and k. The number of unknown variables is five: (L_x, L_y, L_z, t_0, k). Pixels with the same depth are chosen to get those parameters. Some pixels with the same depth value are obtained shown as Fig.5. By taking the depth values into Eq.(1), we obtain t_0 and k.

Since the laser can not be a divergent light source, a diffuser is used to be a virtual light source. In this situation, L can be easily calculated. We assume that the virtual light's coordinate is (L_x, L_y, L_z), and the

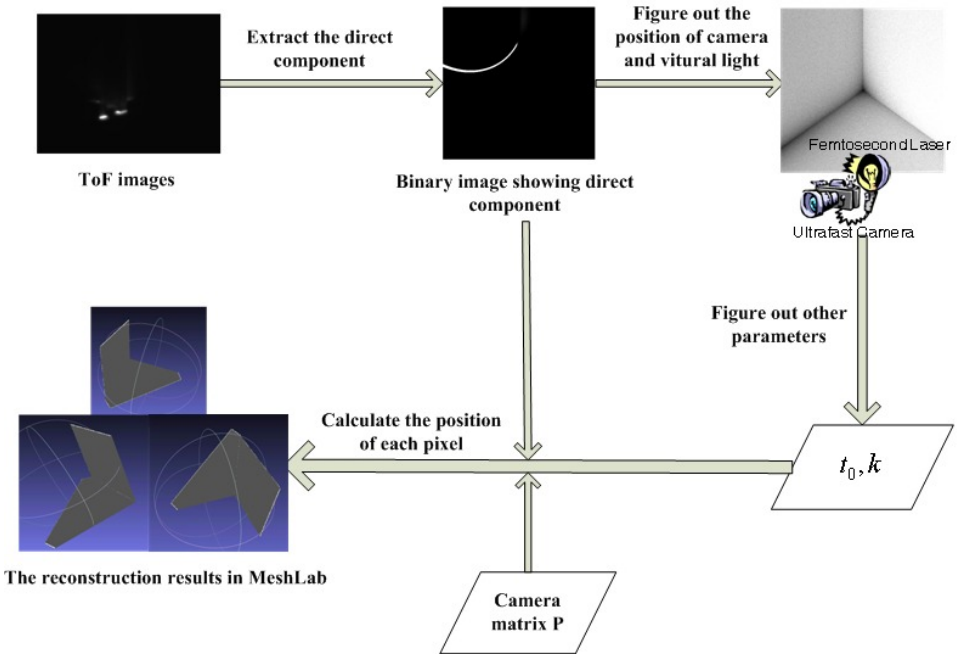


Figure 2. The flow diagram.

camera's coordinate is (0,0,0). In practice, we get a point pair with the same depth in an image and the geometric relationship is expressed as follows:

$$\begin{aligned}&\sqrt{(X_1-L_x)^2+(Y_1-L_y)^2+(Z_1-L_z)^2}+\sqrt{X_1^2+Y_1^2+Z_1^2}\\&=\sqrt{(X_2-L_x)^2+(Y_2-L_y)^2+(Z_2-L_z)^2}+\sqrt{X_2^2+Y_2^2+Z_2^2}\end{aligned}\tag{2}$$

where

$$\begin{cases}X_i=d_i\frac{x_i-P_{13}-P_{12}\frac{y_i-P_{23}}{P_{22}}}{P_{11}}\\Y_i=d_i\frac{y_i-P_{23}}{P_{22}}\\Z_i=d_i\end{cases}\tag{3}$$

Four point pairs with the same depth shown in Fig.5 are chosen. After L and d are carried out, k and t_0 can be calculated. The flow diagram is shown as Fig.2.

2.3 Scene points calculation

Since the camera calibration is completed, the geometric relationships are used in the acquisition of scene points' coordinates. The geometric constraints are expressed as follows:

$$||X-L||+||X-C||=k(t-t_0)\tag{4}$$

$$x=PX\tag{5}$$

where t_0 and k are known, X denotes the coordinate of scene point in world space, x is the corresponding point in the image. Eq.(4) and Eq.(5) denote that the point is the crosspoint of the line which is through the camera center, the pixel and the ellipsoid. As X=(x,y,z), C=(0,0,0) and L=(L_x,L_y,L_z), Eq.(4) can be described as

$$\sqrt{(x-L_x)^2+(y-L_y)^2+(z-L_z)^2}+\sqrt{x^2+y^2+z^2}=k(t-t_0)\tag{6}$$

Since there is only one camera, the camera's coordinate is (0,0,0). The camera matrix P is defined as

$$P = \begin{pmatrix} P_{11} & P_{12} & P_{13} & 0 \\ 0 & P_{22} & P_{23} & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \tag{7}$$

By taking Eq.(7) into Eq.(5), Eq.(5) is rewritten as:

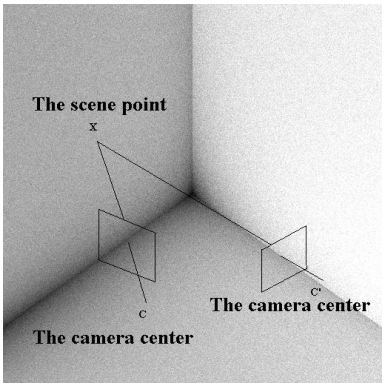
$$\begin{cases} P_{11}X + P_{12}Y + P_{13}Z = xd \\ P_{22}Y + P_{23}Z = yd \\ Z = d \end{cases} \tag{8}$$

where x and y are the coordinates of the pixel in the image and d is the depth of the scene point. (X,Y,Z) denotes the scene point and takes the following form:

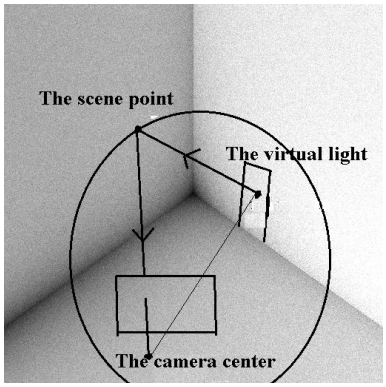
$$\begin{cases} X = d\frac{x - P_{13} - P_{12}\frac{y-P_{23}}{P_{22}}}{P_{11}} \\ Y = d\frac{y-P_{23}}{P_{22}} \\ Z = d \end{cases} \tag{9}$$

By taking (X,Y,Z) into Eq.(2), the function takes the form of Ad^2+Bd+C=0. The function has two solutions, and they represent the two points that the line intersect with the ellipsoid. The solution greater than 0 makes sense. By taking d into Eq.(9), (X,Y,Z) can be obtained. The algorithm refers to Algorithm 2.

The scene is reconstructed from only one view, while the traditional reconstruction methods use multi-view images. Fig.3 gives an example of the leading difference. The previous reconstruction method is shown in Fig.3a, and our scheme is shown in Fig.3b.



(a) The traditional reconstruction method which is required at least two views



(b) Our scheme which is required only one view

Figure 3. The comparison of our work and other reconstruction methods

Algorithm 2 Initial parameters and point clouds acquisition

Require: The direct Component D(t)

1. Represent (X_i,Y_i,Z_i) by Eq.(3) with the unknown depth d.
2. Take four groups of (X_i,Y_i,Z_i) into Eq.(2) and (L_x,L_y,L_z) can be calculated.
3. Take (L_x,L_y,L_z) into Eq.(1) and compute k and t_0.
4. For a pixel (x,y) with a direct component, represent the corresponding scene point with Eq.(8).
5. Take Eq.(9) into Eq.(4) and Eq.(5), and obtain the coordinates of the scene points.

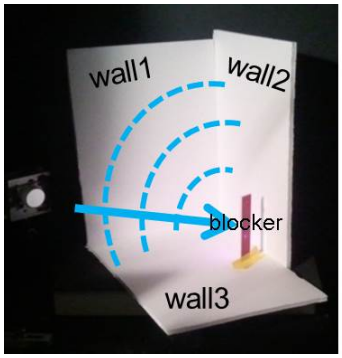


Figure 4. The structure of the scene.

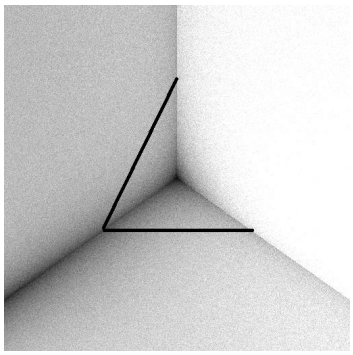


Figure 5. The pixels with the same depth.

3. EXPERIMENTAL RESULTS

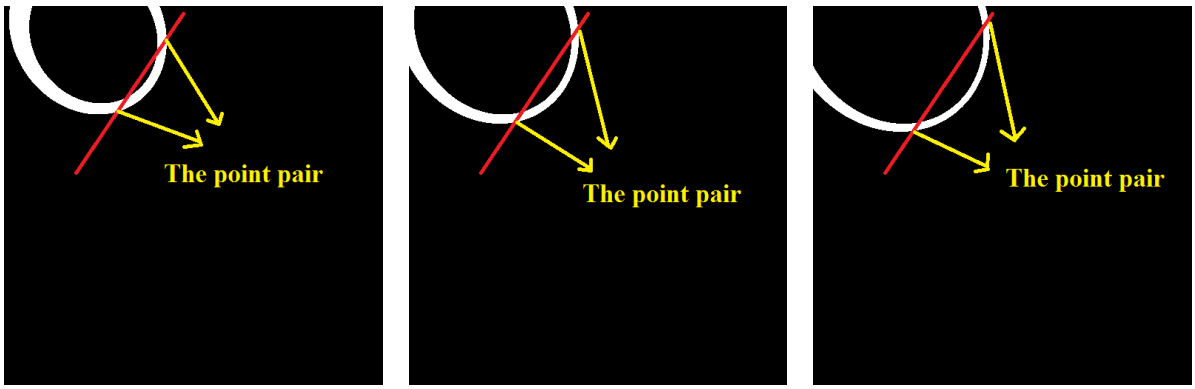
In this section, we illustrate partial scene reconstruction from a number of ToF images. The scene is composed of three walls and a diffuser in the third wall, which is used to reflect the light of laser to the scene. Fig.4 shows the scene.

At first, we compute the camera's and the virtual light source's coordinates. As the three walls are symmetric to the camera, and each two sides intersect with a 120-degrees angle. The plane with equal depth to the camera and the three walls intersect to form an equilateral triangle in the image. In each wall, the line parallel to the corresponding side of the equilateral triangle consists of pixels with equal depth. A corresponding side can be used to find a point pair in an image. By keeping the side stable in four images, a point pair could be obtained in each image. These four point pairs all have the same depth. The pixels with the same depth in an image are shown in Fig.5. Fig.6 shows three images with the line of the same depth. The point pairs with the same depth are chosen to yield the unknowns.

Each coordinate of the scene pixel with a direct component could be calculated by using geometric constraints. The scene point is located in the intersection of the line and the ellipsoid. As the third wall has a virtual source light, only two walls can be reconstructed. By showing each scene point in MeshLab, we can get the scene structure shown in Fig.7 and Fig.8. The whole structure is displayed in Fig.7 and Fig.8 shows the detail sections. The experimental results validate the effectiveness of the proposed approach.

4. CONCLUSION

This paper is devoted to reconstructing partial scene. The proposed techniques realized partial scene reconstruction by using ToF images. It is different from traditional scene reconstruction methods with multi-view images. In the previous research, the direct component could be extracted from ToF images. The direct component containing two rays is easier to be analyzed. Incorporating the ToF images, we are able to reconstruct a scene easily. For a solution, the reconstruction is achieved by using the geometric relationships between the focus of the camera, the scene point and the virtual light source. The distance that light travels between the two focuses



(a) (b) (c)

Figure 6. Three Point pairs used to calculate the unknowns with the same depth.

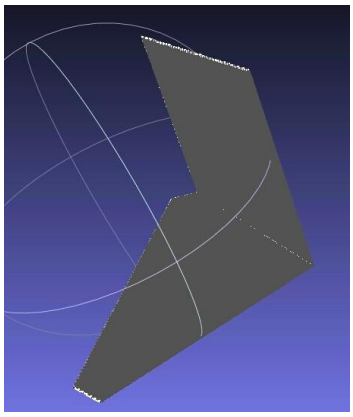
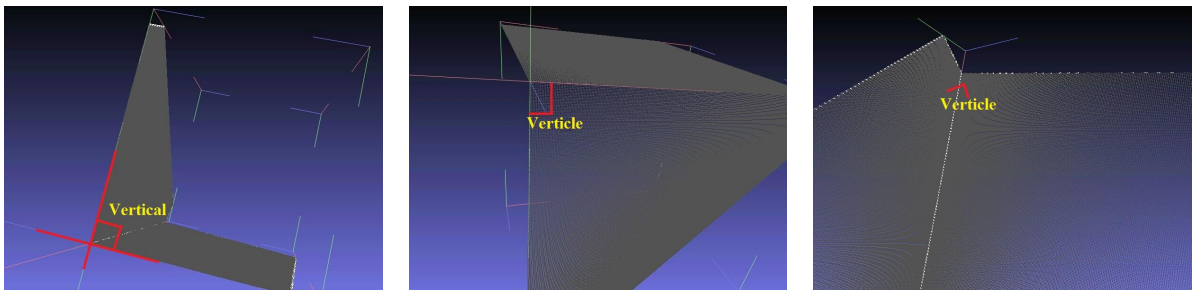


Figure 7. The whole structure in MeshLab.

and the scene point is a constant. The focus of the camera, the scene point and the corresponding image point are supposed to be collinear. The previous reconstruction algorithms require kinect cameras or camera matrix, while our approach only requires a stable camera.

REFERENCES

[1] Vogiatzis, G., Torr, P. H., and Cipolla, R., "Multi-view stereo via volumetric graph-cuts," in [*Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*], **2**, 391–398, IEEE (2005).



(a) (b) (c)

Figure 8. Scene reconstruction shown in details. The gray regions are reconstructed models and red lines mark the regions with vertical geometric relationship

[2] Hernández Esteban, C. and Schmitt, F., "Silhouette and stereo fusion for 3d object modeling," *Computer Vision and Image Understanding* **96**(3), 367–392 (2004).

[3] Goesele, M., Snavely, N., Curless, B., Hoppe, H., and Seitz, S. M., "Multi-view stereo for community photo collections," in [*Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*], 1–8, IEEE (2007).

[4] Merrell, P., Akbarzadeh, A., Wang, L., Mordohai, P., Frahm, J.-M., Yang, R., Nistér, D., and Pollefeys, M., "Real-time visibility-based fusion of depth maps," in [*Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*], 1–8, IEEE (2007).

[5] Wu, D., Velten, A., OToole, M., Masia, B., Agrawal, A., Dai, Q., and Raskar, R., "Decomposing global light transport using time of flight imaging," *International Journal of Computer Vision* **107**(2), 123–138 (2014).

[6] Bai, J., Chandraker, M., Ng, T.-T., and Ramamoorthi, R., "A dual theory of inverse and forward light transport," in [*Computer Vision–ECCV 2010*], 294–307, Springer (2010).

[7] Seitz, S. M., Matsushita, Y., and Kutulakos, K. N., "A theory of inverse light transport," in [*Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*], **2**, 1440–1447, IEEE (2005).

[8] Liu, S., Ng, T.-T., and Matsushita, Y., "Shape from second-bounce of light transport," in [*Computer Vision–ECCV 2010*], 280–293, Springer (2010).

[9] Nayar, S. K., Krishnan, G., Grossberg, M. D., and Raskar, R., "Fast separation of direct and global components of a scene using high frequency illumination," in [*ACM Transactions on Graphics (TOG)*], **25**(3), 935–944, ACM (2006).

[10] Chen, T., Lensch, H., Fuchs, C., and Seidel, H.-P., "Polarization and phase-shifting for 3d scanning of translucent objects," in [*Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*], 1–8, IEEE (2007).

[11] Gupta, M., Tian, Y., Narasimhan, S. G., and Zhang, L., "(de) focusing on global light transport for active scene recovery," in [*Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*], 2969–2976, IEEE (2009).

[12] Zhang, L. and Nayar, S., "Projection defocus analysis for scene capture and image display," in [*ACM Transactions on Graphics (TOG)*], **25**(3), 907–915, ACM (2006).

[13] Gupta, M., Agrawal, A., Veeraraghavan, A., and Narasimhan, S. G., "Structured light 3d scanning in the presence of global illumination," in [*Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*], 713–720, IEEE (2011).

[14] Raskar, R., "Femto-photography: Visualizing photons in motion at a trillion frames per second." `http://web.media.mit.edu/~raskar/trillionfps/`.