# Next Generation CAT System

by

## Jaewon Kim

M.S., Korea Advanced Institute of Science and Technology (2001)
B.S., Korea Advanced Institute of Science and Technology (1999)

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning,
in partial fulfillment of the requirements for the degree of

Master of Science in Media Arts and Sciences

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2010

Author⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯
Program in Media Arts and Sciences
August 6, 2010

Certified by⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯
Ramesh Raskar
Associate Professor of Media Arts and Sciences
Program in Media Arts and Sciences
Thesis Supervisor

Accepted by⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯
Pattie Maes
Associate Academic Head
Program in Media Arts and Sciences

**Next Generation CAT System**

by

Jaewon Kim

The following people served as readers for this thesis:

Thesis Reader
Ramesh Raskar
Associate Professor of Media Arts and Sciences
Program in Media Arts and Sciences

Thesis Reader
V. Michael Bove
Principal Research Scientist of Media Arts and Sciences
Program in Media Arts and Sciences

Thesis Reader
Fredo Durand
Associate Professor of Electrical Engineering and Computer Science
MIT CSAIL

Thesis Reader
Yasuhiro Mukaigawa
Associate Professor of Institute of Scientific and Industrial Research
Osaka University

# Next Generation CAT System

by

Jaewon Kim

## Abstract

Two novel techniques for future CAT system are presented. Transmission descattering is a single-shot method to differentiate unscattered and scattered components of light transmission through a scattering material. Directly-transmitted components travel in a straight line from the light source, while scattered components originate from multiple scattering centers in the volume. Angularly varying scattered light is strategically captured via a lenslet array placed close to the image plane and the unscattered direct component is computed based on separable scattered components. The disadvantage is a reduction in spatial resolution. As an application, the enhanced tomographic reconstruction is demonstrated using estimated direct transmission images. The other technique is single-shot 3D reconstruction of a translucent object. Multiple light sources form images of a translucent object at different projection angles onto a screen. Those images are captured by a single-photo in a coded format via lenslet array. The projection image casted from each light source is separated from each other by a decoding process and in turn the images are combined to reconstruct 3D shape of the translucent object by ART method.

Thesis Supervisor: Ramesh Raskar
Title: Associate Professor of Media Arts and Sciences, Program in Media Arts and Sciences

# Acknowledgements

This thesis could be come out by many people's help. It would be my duty and great pleasure to remember them in this page.

Ramesh Raskar, my academic advisor, showed me new research direction in computational photography. It was my great chance to extend my research career from computer vision to computational photography under his excellent guidance. I learned a lot of new techniques and performed novel researches in the field. This thesis is the condensed output of my research projects. I appreciate Ramesh's advising and management on the projects and this thesis. Thanks a lot, Ramesh.

Michael Bove gave me very useful ideas and comments on this thesis. It contributed to improve the weak points in this thesis research. Also, discussion with him brought new ideas for future research projects. I appreciate Michael on sharing his professional views and opinions on this thesis work.

Yasuhiro Mukaigawa helped me a lot on my thesis work and research projects. Although he was a just visiting professor in our group, he discussed my research in very detail and spent gladly much time to teach me various concepts and theoretical background. It was my great luck that you visited our group at that time period. I appreciate all of your help on teaching and advising.

Fredo Durand readily accepted my asking for being a thesis reader although he was on his sabbatical year. He gave me useful comments and feedback at distance. Also, one more thing I appreciate to Fredo is that his class teaching significantly contributed to my TOG paper submission. Thanks a lot, Fredo, for your teaching and all comments on my research projects and this thesis.

I should expose there have been much helps on this thesis work from many other people including my group members. I appreciate all of them to share willingly their time for my requests on thesis work and projects.

Lastly, I hope to deliver the entire honor on this thesis to my wife, Juyoung Mun. Without her help and devotion on my study and work, I wouldn't have completed Media Lab MS course. I deeply appreciate Juyoung again to endure all the hard time with my lovely kids, Minchai and Dongwoo, in new life environment.

# Contents

# List of Figures

# Chapter 1

# Introduction

Generally CAT system is applied to acquire 3D inner views of a volume which cannot be visualized by human eyes. Despite of its useful characteristics, its usage has been limited to specific professional applications in medical, industrial and security fields. The major reason of such limitation is that X-ray, the typical emitting source in CAT system, is hazard to a human body because long exposure to X-ray radiation can cause critical diseases by destroying or transforming human cells. Another limitation in CAT usage is caused by its bulky and expensive system. X-ray scanning, typical mechanical operation in CAT, requires huge mechanical components to rotate and translate a heavy X-ray tube with enforcing the overall system bulky, slow and expensive. If CAT system can be built in an easily-accessible way for daily life, there would be a wide range of useful applications such as regular body-check for our health condition by ourselves and biometric/HCI applications using inside shape information of our bodies. This thesis addresses two novel techniques necessary for such new concept of CAT system.

The first technique, transmission-mode descattering, enhances transmitted light signal through a volume in which multiple scattering is dominant. This technique contributes to acquire the enhanced image of the inner volume with low-power X-ray or unharmful light sources. Many researchers have used NIR (Near Infrared) sources to look inside a human body by analyzing the reflected light from the body in wavelength domain. However, those approaches are still limited to only thin parts of a body due to insufficient penetrating ability of the light sources. The technique presented in this

thesis enhances SNR of transmitted light through a thick volume by eliminating scattering effect which generally makes a sensed image blur. This technique can be applied to electromagnetic radiation such as X-ray and terahertz wave as well as any kind of light sources including IR and visible light sources. In experiments, transmitted light through a scattering media, milky water or human skin, from a normal LED or an IR LED source is separated into direct and scattered component. Then, the enhanced 3D shape of an object inside the volume is reconstructed by combining only direct component images which contain the sharpest shape information.

The other technique is a single-shot 3D reconstruction method using 4D light field photography. Current CAT system requires multiple images taken at different locations of X-ray source with scanning it rotationally. 4D light field capturing technique allows sensing rays from multiple light sources with a single-shot photo and generating image set independently contributed by each light source. This thesis explains how to adopt the 4D light field capturing method into CAT system for the sake of designing a scan-free system.

## 1.1 Contributions

This thesis addresses novel techniques to realize next generation CAT system which is scan-free, fast and less harmful. The primary technical contributions are as follows.

- This thesis presents a single-shot CAT technique which provides a real-time 3D inner view of a volume in a scan-free system. Such fast CAT system can be used to monitor the movement of organs (like a heart) for medical treatments or research. Also, scan-free CAT technique reduces the system size without using mechanical components for scanning. In the future goal, wearable or portable CAT system will benefit by this technique.

- This thesis describes a method for single-exposure separation of direct and scattered components of transmitted light passing through scattering media using a lenslet or pinhole array placed close to the image sensor. In the direct-only image, high-frequency details are restored and provide strong edge cues for scattering objects. Due to its single-shot nature, this method is well-suited for analyzing dynamic scenes.

14

- This thesis demonstrates enhanced tomographic reconstruction of scattering objects using direct component images. These separation methods are well-suited for applications in medical imaging, providing an internal view of scattering objects such as human skin using visible or near-visible wavelength light sources, rather than X-rays.

## 1.2 Related Work

**Light Field Capturing:** Adelson[1] presents a single lens stereo method to find depth cues by a single-shot photo using a pinhole or lenslet array. The concept of 4D light field sensing is presented by Levoy[2] and Gortler[3]. Isaksen et al.[4] describe a practical method to compute 4D light field. Levoy et al.[5] and Vaish et al.[6] introduce improved ways using micro lens arrays and camera arrays. Veeraraghavan et al.[7] present an inexpensive way to capture 4D light field by a 2D thin mask. Coded images and light field techniques have been used in many ways [8][9][10][11].

**Visible light tomorgraphy:** Tomographic methods using visible light have been widely researched in medical imaging field. Optical coherence tomography [12] is a high-resolution cross-sectional imaging method based on the echo time of backscattered light. Interferometer with coherent light is adopted for the precise measurement of the echo time. Diffuse optical imaging [13] is another approach to view inside a body using NIR light. This technique basically quantifies the change of hemoglobin concentration in blood based on the fact that the absorbed amount of NIR light is proportional to the hemoglobin concentration. To compensate the sensing error, MRI (Magnetic Resonance Imaging) is combined with this method. Wang and Zhang [14] describe various terahertz(THz) tomography methods including THz diffraction tomography, THz computed tomography, THz binary lens tomography and THz digital holography.

**Shield Field Imaging:** Douglas et al. [15] present a way to reconstruct 3D shape of an occluder by a single-shot photo. In the research, multiple LEDs cast silhouettes of the occluder at different directions onto a screen. The silhouettes are imaged in a coded format by a mask inside the screen [16]. From the captured image, each silhouette is separated from one another by a decoding process, and then they are combined to reconstruct 3D shape of the occlude by Visual Hull method.

**Direct-Scattered Separation:** Direct-scattered separation of light is widely studied in diverse fields spanning computer vision, graphics, optics, and physics. Due to the complexities of scattering, reflection, and refraction, analytical methods do not achieve satisfactory results in practical situations. In computer vision and graphics, Nayar et al. [17] present an effective method to separate direct and scattered components from a scene by projecting a sequence of high-frequency patterns. Their work is one of the first to handle arbitrary natural scenes. However, it requires temporally-multiplexed illumination, limiting the utility for dynamic scenes. Nasu et al. [18] present an accelerated method using a sequence of three patterns. In addition, Rosen and Abookasis [19] present a descattering method using speckle analysis.

**Microscopy:** The scattering in microscopic objects is addressed by careful optical methods. Hisashi [20] presents a method to achieve a sharp in-focus signal in a confocal microscope setup. He uses two pinholes to sense in- and out-of-focus signals and acquire a sharp in-focus signal by subtracting the two. This requires two exposures and careful alignment for each spot. In addition, scanning process is required to recover a whole scene. Sheppard et al. [21] present a method to improve lateral and axial resolution. They achieve enhanced lateral resolution by subtracting a weighted traditional signal from a confocal imaging signal. Their method also increases axial resolution by using a number of detectors in different sizes. These are multi-exposure methods but are similar to this thesis's method where a simple linear transformation of intensities in a neighborhood recovers the sharp component. Levoy et al. [22] record 4D light field using a microlens array for digital refocusing and acquiring angular views with a single-snapshot photo in microscope. In this thesis's approach, the angular variation recorded by the similar way is exploited explicitly making it robust to non-homogeneous local variations. Also, the approach requires no complicated light sources or mechanical scanning or change in aperture settings. Jaewon et al. [23] applied confocal microscopy to normal photography to generate highlight effect on focused objects using projector-camera setup.

**3D Recovery in Scattering Media**: Narasimhan et al. [24] and Gu et al. [25] use sequential structured light patterns to recover 3D shape of static opaque objects in low density scattering media. This thesis's method requires simple light sources and only a single photo per view. Atcheson et

16

al. [26] recover non-stationary gas flows using Schlieren imaging and multiple cameras. The method is suitable for refracting but not scattering media. Rosen and Abookasis [19] proposed a method to recovery shape of binary objects between 2 layers of scattering media based on refocusing principles. Trifonov et al. [27] consider tomographic reconstruction of transparent objects using large number of photos and index matching liquids.

# Chapter 2

# Descattering Transmission via Angular Filtering

The separation of direct and scattered components of incident light is a challenging topic in computer vision and graphics, a task that is confounded by the complex behavior of light in participating media, e.g., reflection, refraction, and scattering in haze, underwater or in volumetric translucent objects. These complex characteristics of light are one of the main factors hindering an analytical solution for direct-scattered separation. For this reason, active coding methods have been proposed. Nayar et al. [17] project high-frequency patterns onto a reflective scene. Such active coding methods achieve accurate and robust separation. Narasimhan et al. [24] use structured light to estimate the 3D shape of objects in scattering media, including diluted suspensions. Gu et al. [25] also use structured light, exploiting compressive sensing techniques, to decrease the data acquisition time. Atcheson et al. [26] estimate the 3D shape of non-stationary gas flows. In many existing approaches, only scattering scenes composed of low density materials (eg. smoke, liquid, and powder) are allowed, such that a single scattering mode is dominant. This thesis demonstrates direct-scattered separation for scenes in which multiple scattering is predominant. In this thesis's approach a passive, single-shot imaging method is used to achieve separation of *transmitted* light for a scene containing heterogeneous scattering media. Specifically, a lenslet (or pinhole) array is placed close to the image plane to separate direct and scattered components of incident light (albeit while reducing the resolution of

(a) Acutal Setup        (b) Schematic Diagram
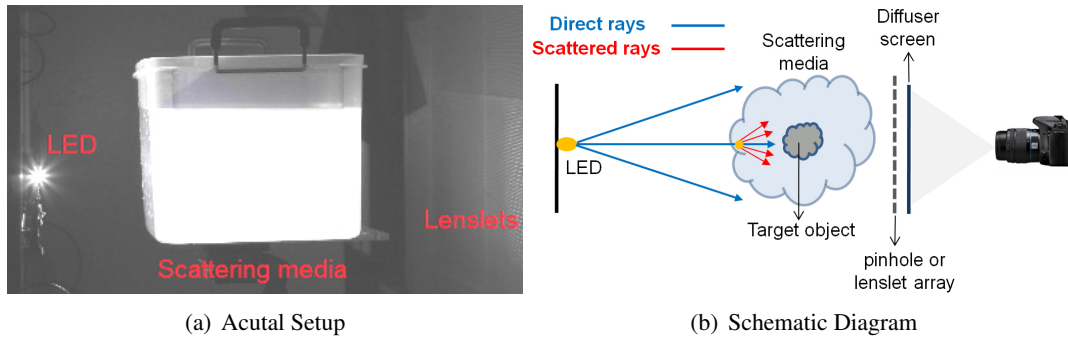
**Figure 2-1:** *Imaging system for single-shot descattering. (a) Imaging system consisting of an LED and a lenslet array. A single LED is used to back-illuminate a scattering scene. A diffuser is used to form an image through a lenslet array. A high-resolution camera captures the array of lenslet images in a single exposure. (b) A schematic diagram of the actual setup.*

the recovered direct and scattering component images since the projection of each lenslet or pinhole provides a single pixel in the recovered images). Using a sequence of such images, an estimate of the volumetric attenuation can be recovered using existing tomographic reconstruction methods, demonstrating benefits for both dehazing and 3D shape recovery.

## 2.1 Imaging System

### 2.1.1 Overview

The actual setup and schematic diagram of the proposed imaging system is shown in Figure 2-1. Note that the direct-scattered separation method handles solid objects and liquid mixtures. In particular, this thesis addresses the case when a solid object is enclosed by a scattering media, as is typical in medical imaging applications. The property of the solid object can be most types of scattering except significant internal refraction. Refraction is treated as scattering and appropriately separated from direct component but the low frequency fitting of this thesis's approach (RTE scattering model) becomes inaccurate. Thin glass objects and thin boundary of media with minor refraction are fine. Under a simple geometric scattering model, light rays are emitted from a point source. When each direct ray impinges on a scattering center, a new scattering light source is effectively created(Figure 2-1(b)). Both direct and scattered rays form an image through a lenslet or pinhole ar-

(a) Without any optical component         (b) With lenslet or pinhole
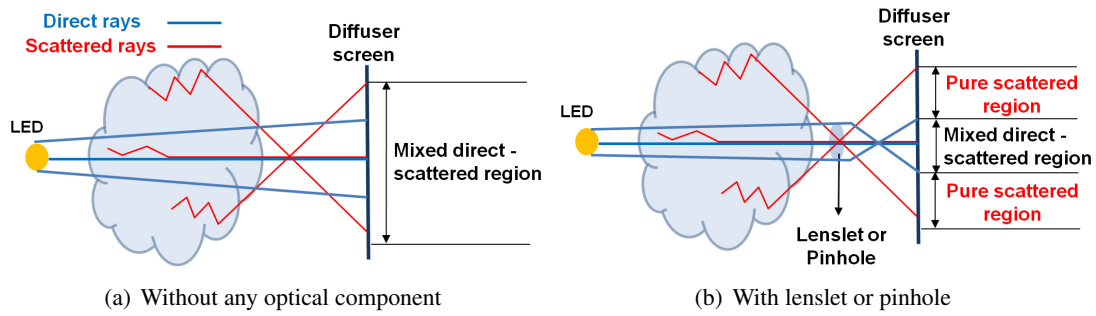
**Figure 2-2:** *Recovering the direct component from a mixed direct-scattered region. There is no way to separate direct and scattered rays in (a). The rays are spatially separated by a lenslet or pinhole in (b). The estimate from the pure-scattered region is used to subtract the scattered component in the central region.*

ray onto a diffuser screen which is captured by a camera. Radiative transport equation (RTE) [28] is applied to model the angular variation of this scattering center. This thesis assumes, at a basic level, that the heterogenous scattering media will be dominated by multiple scattering events [29] [30].

### 2.1.2   Imaging with Lenslets or Pinhole Arrays

In the actual experimental setup, Figure 2-1(a), a Canon EOS Digital Rebel XSi, of which sensor resolution is $4272 \times 2848$ pixels, with a lens of 50mm focal length is located at 1m from the diffuser. The lenslet array(Focal length: 3mm, lens pitch: 5mm) is separated from the diffuser by 5mm in order to form an image of the scattering volume(depth: 200mm) focused on the entire volume with large DOF (centimeters almost infinite) of lenslet (Figure 2-1(b)). Lanman et al. [15] used a similar imaging setup to compute a single-shot light field of opaque objects while this thesis addresses scattering to compute direct component of translucent objects. From Figure 2-2(b), it is infered there are two regions in the image under each lenslet. The first region consists of a mixed signal due to cross-talk between the direct and scattered components. The second region represents a pure scattered component. The following section shows a simple method for analyzing such imagery to separate direct and scattered components for multiple-scattering media. As shown in Figure 2-2(b), the angular sample directly under each lenslet can be used to estimate the combined direct plus scattered transmission along the ray between a given pixel and the light source. Similarly, any non-zero neighboring pixels (not beneath the lenslet) can be fully attributed to scattered illumination due

to volumetric scattering.

## 2.2 Direct-Scattered Separation

### 2.2.1 Separation via Angular Filtering

This section explains direct-scattered separation for a 1-D sensor and a 2-D scene, while the results can be trivially extended to 2-D sensors and 3D volumes. In the following analysis, only a lenslet array is considered, however a similar analysis holds for pinhole arrays. As shown in Figure 2-3(a), the diffuser-plane image consists of a set of sharp peaks under each lenslet in the absence of any scattering media between the light source and diffuser. As shown on Figure 2-3(b), the lenslet images contain extended, blurred patterns when a scattering object is placed between the light source and camera. Ultimately, the scattered light causes samples to appear in pixels neighboring the central pixel under each lenslet. A single lenslet image is defined by two separate regions: a pure scattered component region and a region of mixed direct and scattered components. The received intensity at each diffuser-plane pixel is represented as, $\{L_0, L_1, \ldots, L_n\}$, when a scattering object is placed between the light source and the diffuser. The individual sensor values are modeled as

$$L_0 = G_0 + D_0$$
$$\vdots \tag{2.1}$$
$$L_n = G_n + D_n,$$

where $\{G_n\}$ and $\{D_n\}$ represent the underlying scattered and direct intensities measured in the sensor plane, respectively. As shown in Figure 2-3(b), a straightforward algorithm can be used to estimate the direct and scattered components received at each lenslet. First, the non-zero region is estimated in Figure 2-3(a) which is captured with no object present. Next, values of the scattering component $\{G_n\}$ in the region are approximated using a scattering model, described in next section, as shown in Figure 2-3(c). Note that this region is subject to mixing in Figure 2-3(b) and

Photo without scattering

Direct region

Intensity

Pixel

(a) Without Scattering

Photo with scattering

Pure scattered region | Mixed direct & scattered region | Pure scattered region

Intensity

$...L_{-3}L_{-2}L_{-1}L_0\ L_1\ L_2L_3...$

Pixel

(b) With Scattering

Known scattered values | Unknown scattered values | Known scattered values

Intensity

$...G_{-3}G_{-2}G_{-1}G_0\ G_1\ G_2\ G_3...$

Pixel

(c) Estimation

Computed Direct values

Intensity

$...D_{-3}D_{-2}D_{-1}D_0\ D_1\ D_2\ D_3...$

Pixel
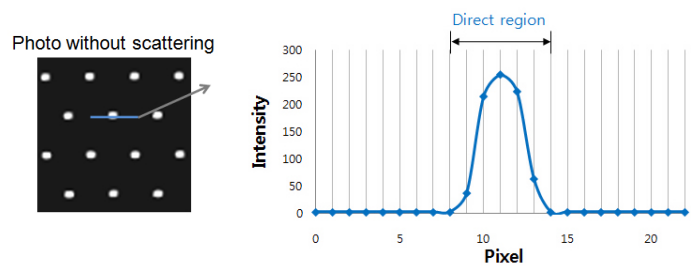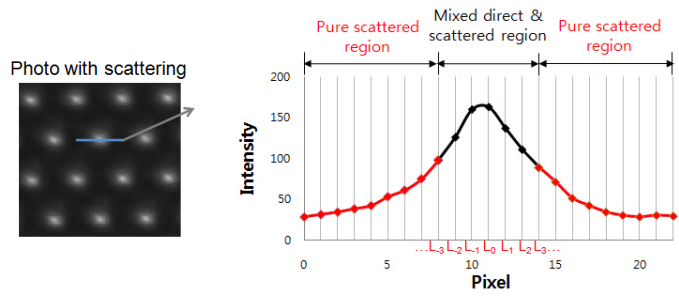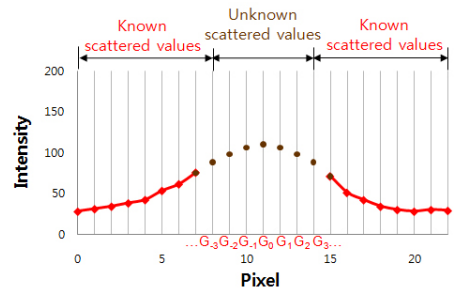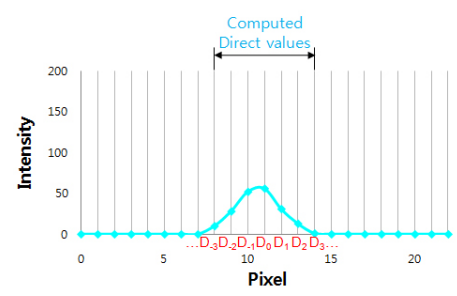
(d) Subtraction

**Figure 2-3:** *Comparison of direct and scatter components without and with scattering media. (a) Central region under each lenslet is sharp without scattering. (b) Direct as well as scattered component is included in the central region. (c) Measured (red) and estimated (brown) values for scattering-only component. (d) The direct-only image formed by subtracting (c) from (b).*

(a) Measured Data     (b) q=0.9, T=1.4     (c) q=0.7, T=1.4     (d) q=0.2, T=1.4

(e) q=0.9, T=2.0     (f) q=0.2, T=2.0     (g) q=0.9, T=4.0     (h) q=0.2, T=4.0
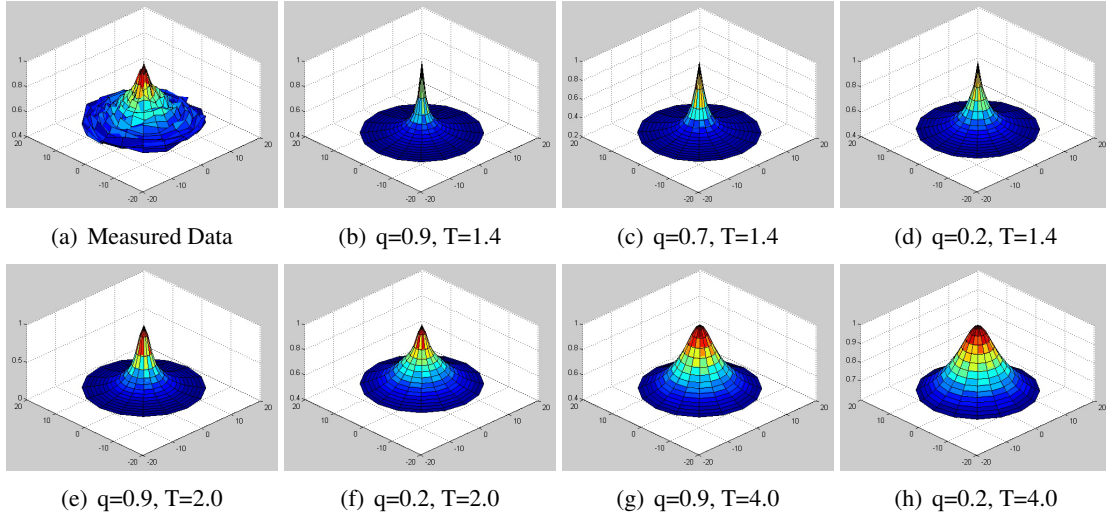
**Figure 2-4:** *RTE(Radiative Transport Equation) modeling of scattering values through a pinhole. (a) Measured data (b)-(h) RTE Modeling with different q and T. (f) is in minimum fitting error with the measured data.*

the scattering component must be approximated from the known scattered values in Figure 2-3(c). Finally, a direct-only image can be estimated by subtracting the estimated scattering component for the central pixel under a lenslet, such that $D_0 \approx L_0 - G_0$.

### 2.2.2 Mathematical Model for Multiple Scattering

This section describes the multiple scattering model used in the descattering algorithm described in the previous section. Numerical Monte-Carlo techniques have been widely used for tracing scattered rays but it needs high computational cost for a large number of rays. To implement efficient descattering algorithm, this thesis's approach uses the physics-based model presented by Narasimhan and Nayar [28]. Multiple scattered intensity through a lenslet can be described by RTE (Radiative Transport Equation) and the solution of it is a function of three parameters, T(optical thickness), q(forward scattering parameter) and x(spatial position) as explained in the Narasimhan and Nayar's paper. RTE is used to fit measured 2D data, Figure 2-4(a), under each lenslet of the imaging condition. (b)-(h) show varied intensity distributions according to different T and q. By an iterative error-minimization method, the best matching profile, (f), can be found for the measured 2D signal (a) and any unknown scattered value for nearby regions can be approximately calculated by the

0.03%  0.05%  0.07%  0.09%

(a) Normal Photo

(b) Direct-only Image

(c) Scattered-only Image

0.11%  0.13%  0.15%  0.17%

(d) Normal Photo

(e) Direct-only Image

(f) Scattered-only Image
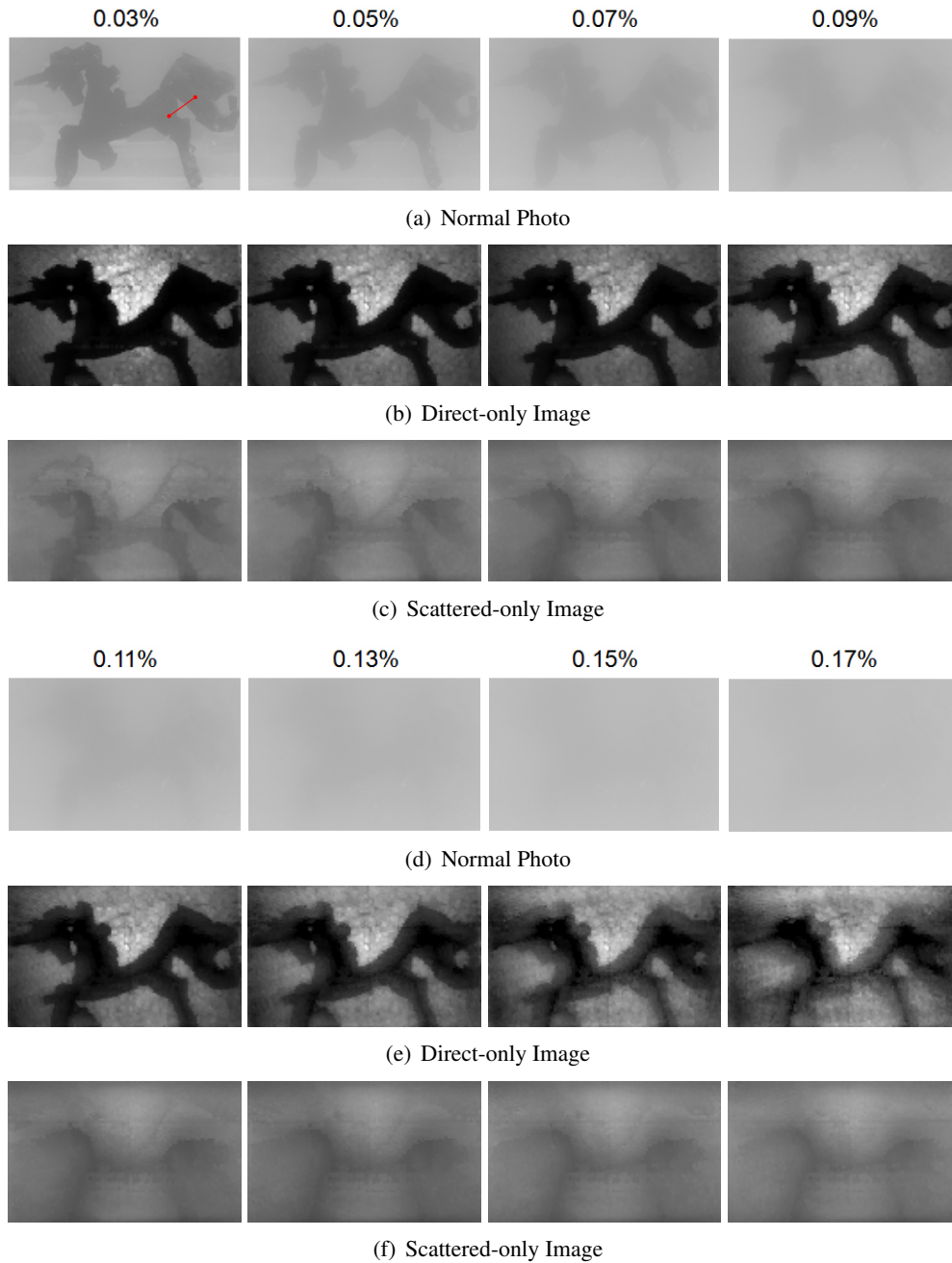
**Figure 2-5:** *Recovery of an opaque object in participating media with milky water. (a) and (d) Normal photos according to concentration 0.03%–0.17% in which water is 7500ml and milk is increased by 1.5ml from 2ml. (b) and (e) Recovered direct-only images computed using angular-domain filtering with a lenslet array. Note enhanced visibility for sharp features and edges in the descattered image. (c) and (f) scattered-only images preprocessed to acquire the direct-only images*

(a) Normal Photo  (b) Direct-only Image

(c) Scattered-only Image  (d) SNR Comp.
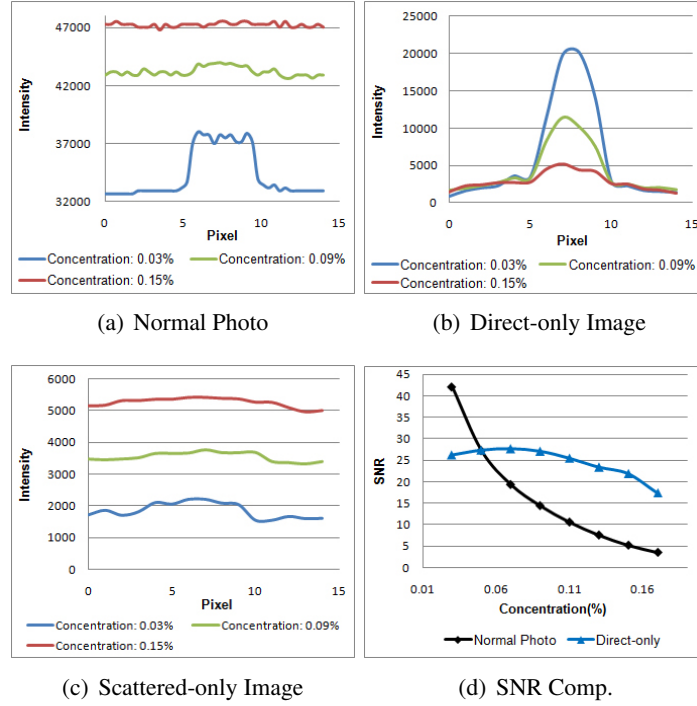
**Figure 2-6:** *(a)-(c) Intensity profiles show that signals in direct-only images are significantly enhanced compared with those in normal photos(Intensity scale 0-65535) (d) SNR comparison between normal and direct-only images shows that the proposed method is effective at scattering-dominent scene.*

fitted model.

### 2.2.3 Experimental Results

From Section 3.1, direct signals $\{D_n\}$ and scattered signals $\{G_n\}$ are separated in each lenslet region. By collecting and combining the direct signals in each lenslet (or pinhole) region, a direct image can be generated. The scattered image is obtained by a similar process, collecting scattered signals. The original image (or normal photo) can be considered as the summed image of the direct and scattered signals. The resolution of the direct and scattered component images are identical to the number of lenslets (or pinholes), because there is only one signal value for direct and scattered components for each lenslet region. In the experiment, the image size is $150 \times 100$. Figure 2-5 compare normal photos of a scattering scene, consisting of an opaque horse-shape object enclosed in an aquarium with milky water, and direct-only images generated by the proposed separation process in

26

(a) Room-light Photo    (b)    Direct+Scattered    (c) Direct-only
                                  with a backlight

**Figure 2-7:** *Direct-scattered separation with a pinhole array. Direct-only images enhance high frequency features of an object enclosed by a scattering solid object.*

lenslet array setup. From left to right, the images show results acquired at higher concentrations of milky water. Figure 2-6 (a)-(c) compare signals at the position of the red line in Figure 2-5(a) for normal photos, direct-only images and scattered only images at different co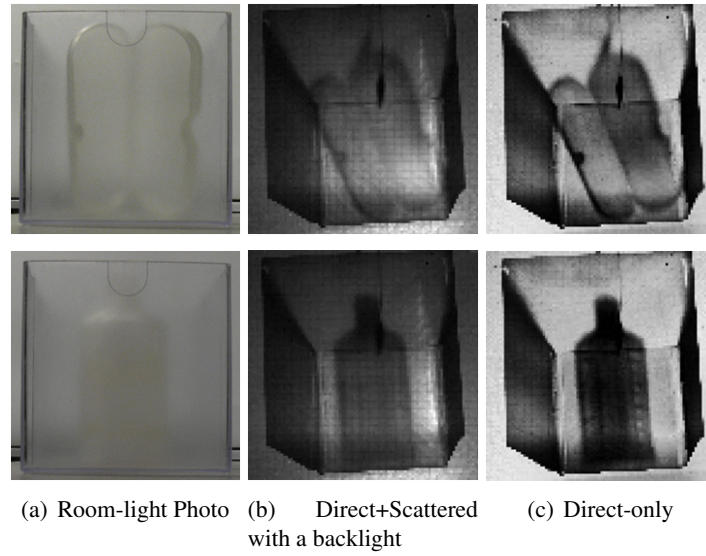ncentrations. Figure 2-6 (b) shows the signals are enhanced compared with signals in normal photos, (a). As the concentration of milky water is increased, the intensity of the signal in direct-only images, (b), is decreased. The opposite effect is observed in scattered-only images, (c), which follows physical reasoning. (d) compares the signal-to-noise ratio (SNR) between normal photos and direct-only images according to concentration. At low concentration, the SNR of a normal photo is larger than one of a direct-only image. However, as concentration is increased, the SNR of a direct-only image gradually becomes higher than the SNR of a normal photo. Note that the signal of normal photos, (a), is significantly decreased from the concentration 0.03% to 0.09% compared with the signal change in direct-only images in (b).

Figure 2-7 shows experimental results using a pinhole array instead of a lenslet array. (a) shows the room-light photo of a solid object placed in the scattering medium. (b) displays ground-truth photos which are acquired by summing all direct and scattered values under each lenslet. (c) contains the direct-only image. By comparing (b) and (c), note that the direct-only images give the sharpest image boundaries for the scattering objects.

27

(a) IR Imaging Setup



(b) Normal Photo    (c) Direct-only    (d) Scattered-only
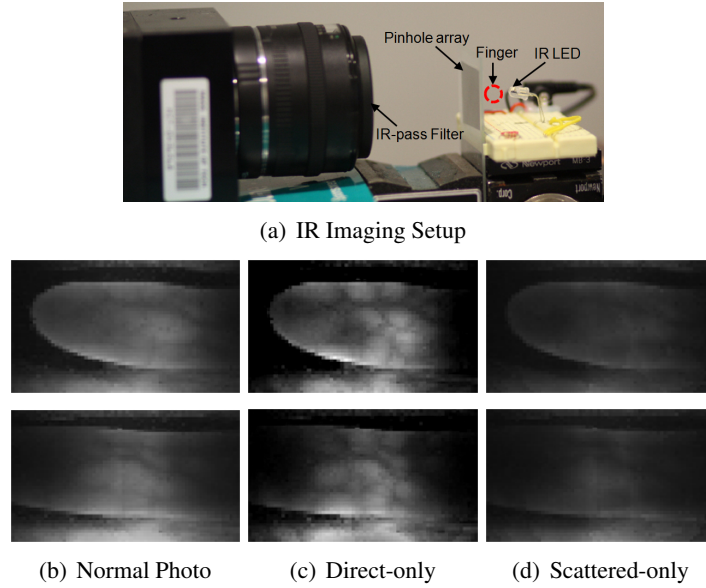
**Figure 2-8:** *Direct-scattered separation images for human fingers using infrared imaging setup. Direct-only images show sharper shapes of finger veins than normal photos. (a) The camera focuses on the pinhole array plane. The IR-pass filter cut visible light and only pass IR light.*

This thesis's method has been tested for human fingers with a near-infrared imaging setup where finger veins are well visualized with infrared light. Direct-only images in Figure 2-8 (c) shows sharper shape of the veins than normal photos do. This is an initial experiment for a strongly scattering challenging object but the image formation model is identical to Figure 2-5 and the results are comparable to [31] which uses specialized narrow wavelength profile light source. As in Figure 2-5, veins closer to the skin are more prominently visible as they are decomposed in the direct component although the finger is strongly scattering. Such enhanced visualization of a human body will benefit medial, biometrics and HCI applications.

### 2.2.4  Volumetric Reconstruction using ART

$$I = I_0 exp(-\sum_{i=1}^{N} a_i f_i) \qquad (2.2)$$

This thesis uses an algebraic reconstruction technique (ART) presented by Roh et al.[32] to reconstruct 3D shape of scattering objects following tradition short-baseline tomography approaches.

Generally, when a ray passes through an object, the change in intensity can be modeled by Equation 2.2. $I_0$ is the original intensity of the ray and I is the resultant intensity after penetrating $N$ layers inside the object. In this equation, $a_i$ means the distance penetrating at $i^{th}$ material of which absorption coefficient is $f_i$ as depicted in Figure 2-9. Equation 2.3 is the logarithmic expression of Equation 2.2. Note that Equation 2.3 can be rewritten for the $j^{th}$ ray as Equation 2.4.

$$h = log(I_0/I) = \sum_{i=1}^{N} a_i f_i \tag{2.3}$$

$$h^j(t) = \sum_{i=1}^{N} a_i^j f_i \tag{2.4}$$
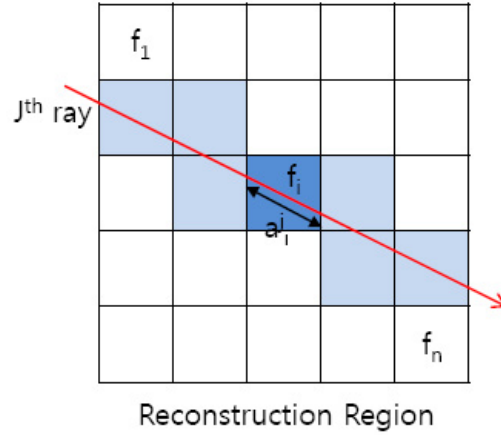


**Figure 2-9:** *Projection model of a ray.*

Now, the problem is finding $f_i$ values which correspond to the density information within the reconstruction region. Equation 2.4 can be described using Equation 2.5 in matrix form. A matrix represents the projective geometry of rays calculated for the emitting position of rays, and the received position for a predetermined reconstruction region in which the object is assumed to be placed. The vector $h$ is related to the sensed intensity values. In practice, the operating of ART takes approximately five minutes for this thesis's data sets.

$$\mathbf{h} = \mathbf{AF} \tag{2.5}$$

$$where\ \mathbf{A} = \begin{bmatrix} a_1^1 & a_2^1 & \cdots & a_N^1 \\ a_1^2 & a_2^2 & \cdots & a_N^2 \\ \vdots & \vdots & \cdots & \vdots \\ a_1^M & a_2^M & \cdots & a_N^M \end{bmatrix} = \begin{bmatrix} \mathbf{a}^1 \\ \mathbf{a}^2 \\ \vdots \\ \mathbf{a}^M \end{bmatrix}$$

$$\mathbf{F} \in R^N, \mathbf{h} \in R^M, \mathbf{A} \in R^{M \times N}$$

Note that Equation 2.6 can be used to get the next step value, $f_i$(t+1), from the parameters at the current $i^{th}$ step. In this equation, t and $\lambda$ are the step index and a coefficient related with convergence parameter, respectively. The values of $g$ and $h$ are the measured value from sensing and the calculation value from Equation 2.5 using $f$ at the current step. As the iteration step, t, increases, the error term, $g^j$-$h^j$(t), decreases and $f_i$(t) gets closer to the exact value. Finally, approximated reconstruction result, f, is acquired.

$$f_i(t+1) = f_i(t) + \lambda \frac{g^j - h^j(t)}{\sum_i^N (a_i^j)^2} a_i^j \tag{2.6}$$

Figure 2-10 (b) and (d) compare two 3D reconstruction results using eight normal photos and eight descattered images, respectively. Eight photos are sequentially captured with a LED mounted at different position to get multiple angular views of the targeted inside object, the bottle, in Figure 2-7 (bottom). In the 3D reconstruction, 2-10 (b), the bottle is rendered by blue color to add distinguishably from the outer scattering container. Note that the rendering isn't accurate for the bottle since the bottle shape in the captured images has been hazed by scattering. The 3D result using direct-only images in 2-10 (d) shows a more accurate rendering result for the inside bottle object.

(a) Normal Photo

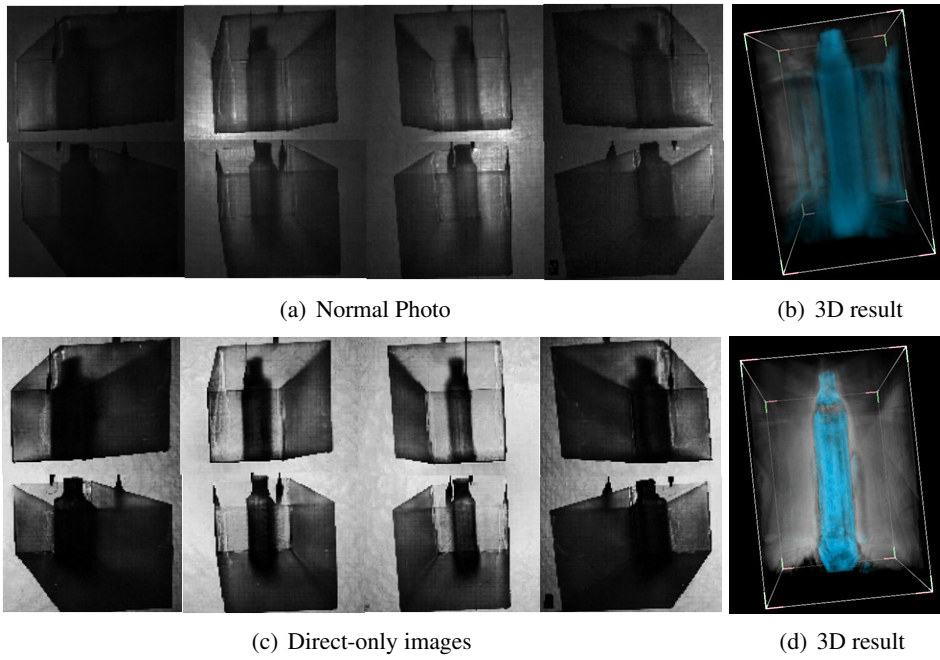(b) 3D result



(c) Direct-only images

(d) 3D result

**Figure 2-10:** *Tomographic reconstruction resutls. (a) and (b) show eight normal photos captured at different light-projection angles and 3D result using them, respectively. (c) and (d) are direct-only images for the each normal photo and 3D result using the direct-only images.*

# Chapter 3

# Single-shot 3D Reconstruction

This section presents a single-shot 3D reconstruction method using 4D light field photography. Current CAT system requires multiple images taken at different locations of X-ray source with scanning it rotationally. 4D light field capturing technique allows sensing rays from multiple light sources with a single-shot photo and generating image set independently contributed by each light source. The way to adopt the 4D light field capturing method into CAT system is described for the sake of designing a scan-free system.

## 3.1    4D Light Field Capturing Technique

4D light field capturing technique has been explored in various literatures in computational photography, computer vision and computer graphics. The representative way to capture 4D light field is using a lenslet array, a pinhole array or a coded mask. Each way has its own advantages and disadvantages. When a lenslet or pinhole array is used, processing is performed in spatial domain with low computational cost. In the other hand, a coded mask requires frequency-domain processing demanding high computation cost. In the aspect of system cost, manufacturing a coded mask or a pinhole array is much cheaper than a lenslet array. A pinhole array, advantageous in computational and manufacturing cost, demands long exposure time due to the nature of low light efficiency. In this thesis, a lenslet and pinhole array is used to accomplish efficient computation.
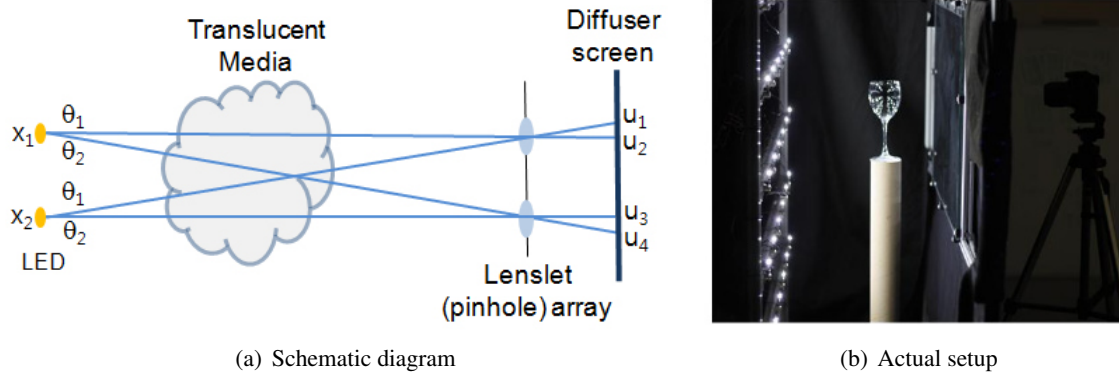
(a) Schematic diagram                   (b) Actual setup

**Figure 3-1:** *2D Light field is parameterized in 1D diagram (a) of the actual setup (b).*

### 3.1.1   Lenslet or Pinhole Array Method

2D light field, geometric domain of rays, is described by x and $\theta$ coordinates in Figure 3-1 (a). A lenslet (or pinhole) array transforms the 2D light field to 1D spatial signal which is formed onto a diffuser screen in u coordinate. The gap between a lenslet(pinhole) array and the diffuser is 5mm. A camera captures the formed image at the other side of the diffuser screen. In real world, light field is defined as 4D coordinates and transformed to 2D spatial coordinates by a lenslet array. Figure 3-1 (b) shows the actual imaging setup consisting of a LED array, a translucent object, a lenslet(pinhole) array, a diffuser screen and the same camera in previous chapter. The LED array is composed of 6x6 Philips Luxeon Rebel cool-white LEDs distributed uniformly over a 1.2 m x 1.2 m region, with each 3 mm wide LED producing 180 lumens at 700 mA. We use the same lenslet array described in section 2.1.2 and 150x100 pinhole array(pinhole diameter of 428 microns) printed at 5,080 DPI on 100 micron polyester base using a Heidelberg Herkules emulsion printer. Figure 3-2 (a) is a captured 4D light field image of a wine glass using the imaging setup and Figure 3-2 (b) is an inset image of it. Each white dot in the repetitive pattern of the red box region gives the intensity value of a ray at a specific angular and a spatial domain. In this setup the angular resolution, the number of columns and rows in the red box, is 6x6 which is same with the number of LEDs.
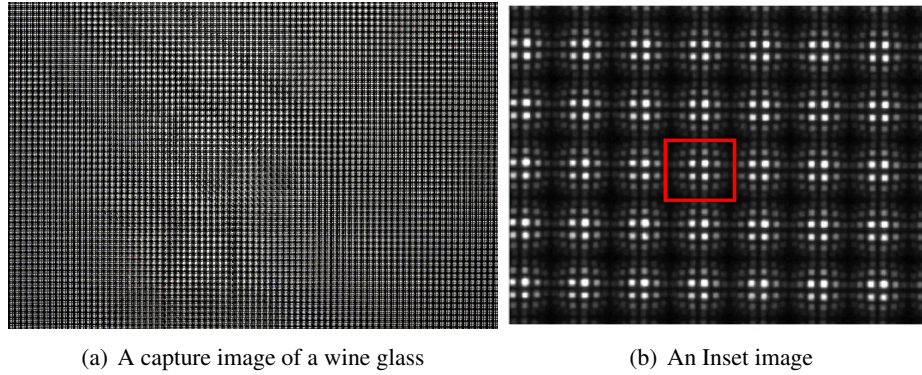
(a) A capture image of a wine glass          (b) An Inset image

**Figure 3-2:** *(a) shows captured 4D light field image of a wine glass (Resolution: 3872x2592). (b) exposes the coded format of 4D light field into 2D spatial domain*

### 3.1.2 Decoding Process for Captured Light Field

This section presents the decoding method to separate 2D image contributed by a single LED from the captured image. In the imaging setup, LED positions and the gap (4mm in actual setup) between a diffuser and a lenslet array are carefully chosen to meet the condition that there is no overlap between white dots in the captured image (Figure 3-2 (b)). In the decoding process, shown in Figure 3-3, the center pixel in the white dot at same angular coordinate is collected to generate a 2D image, which is formed by a single LED. By repeating this process, N images (same with the number of LEDs) are obtained. The resolution of each decoded image is the number of lenses in the lenslet array (or the number of pinholes when a pinhole array is used). Figure 3-4 shows the decoded image set from Figure 3-2 (a) by this process. The resolutions of the original image, Figure 3-2 (a), and each decoded image, Figure 3-4, are 3872x2592 and 150x100, respectively. Each decoded image provides different angular views of the object, and consequently it proves that the multiple views of any object are instantaneously obtained by a single-shot photo with this thesis's method.

### 3.1.3 Preprocessing for Enhanced 3D Reconstruction

In the decoded images (Figure 3-4), the intensity variation is not significant between foreground (object) region and background region because thin translucent objects were tested to minimize refraction not handled by this research. To enhance the 3D reconstruction of the objects, it is required to exclude the background region from reconstruction process. For this job, reference images
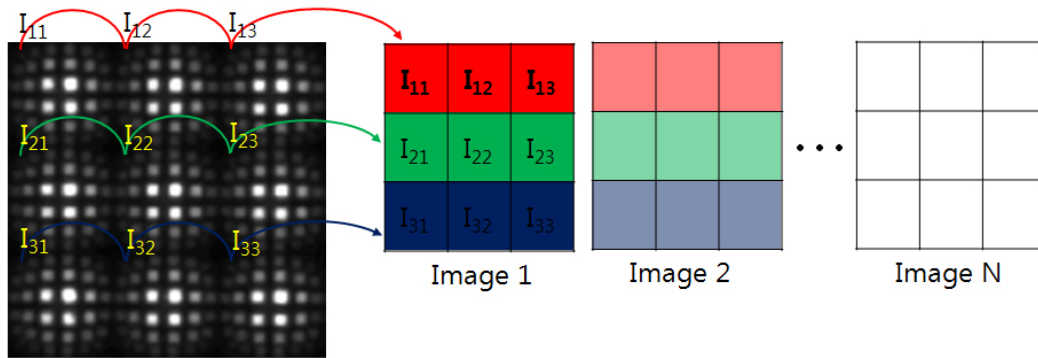
**Figure 3-3:** *Decoding process of a coded 4D light field image*
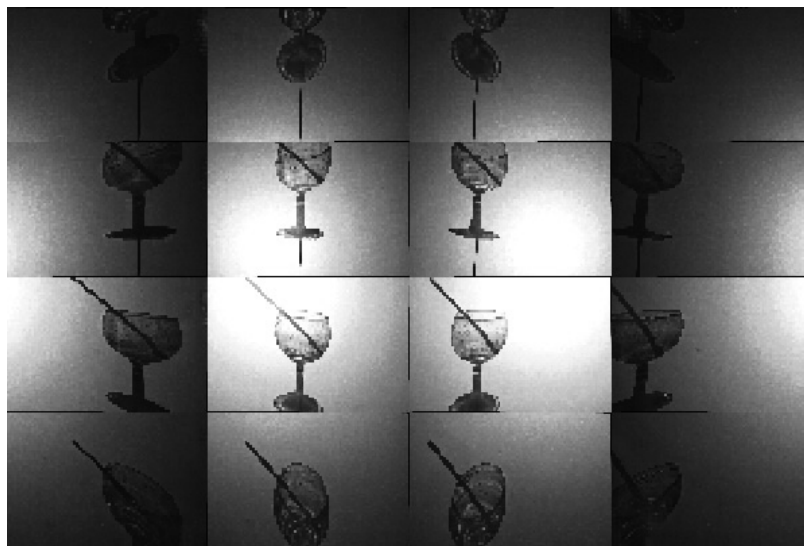


**Figure 3-4:** *16 Decoded images from a coded light field image of a wine glass and a straw(the resolution of each image is 150x100)*

(Figure 3-5 (a)) are acquired by decoding an image taken without any object. Subtracted images and Region-filled images, Figure 3-5 (c) and (d), are generated by subtracting the original decoded images from the reference images and morphological processing on the subtracted images. Finally, foreground-only images, Figure 3-5 (e), are achieved by segmentation on the Region-filled images.
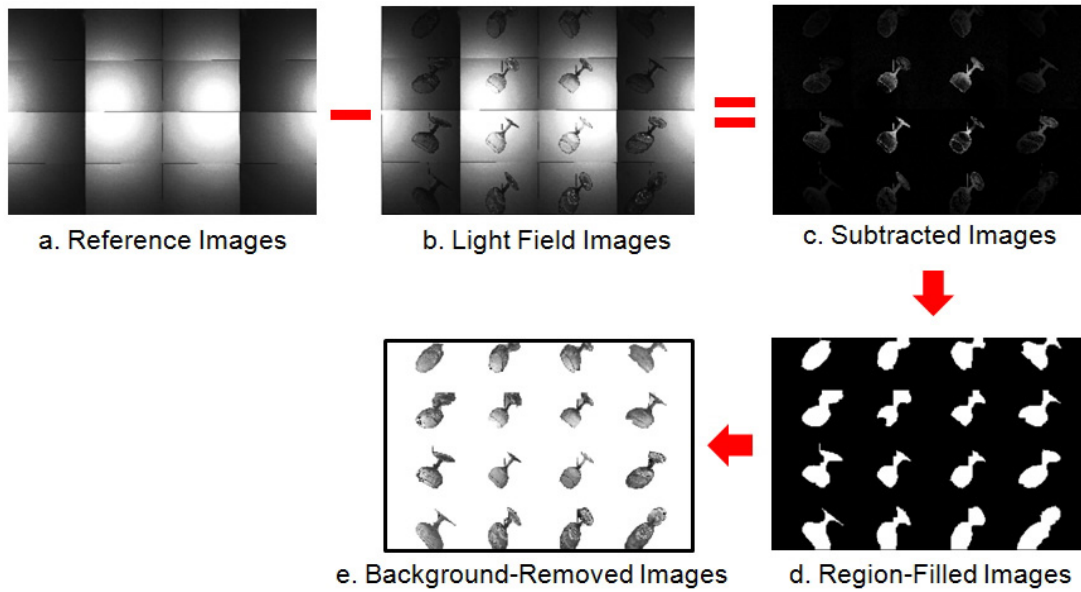
**Figure 3-5:** *Image processing to remove background of light field images*

## 3.2 Experimental Results

ART method (Roh et al. [32]) is applied to reconstruct 3D shape of objects following traditional short-baseline tomography approaches. Figure 3-6 (a) shows decoded 10 images from a single coded light field image of a translucent dog-shaped toy situated in a translucent plastic cup and (b) is the 3D reconstruction result using the images. The inside dog and the outer cup are rendered in cyan and white, respectively. Figure 3-7 (b) shows 3D shaping result for a wine glass with a straw using the decoded images in (a). The wine glass and the straw are rendered in white and red, respectively. Note that the inside objects in both cases are separable from the outer medium as being rendered by different colors in the 3D reconstruction results (Figure 3-6 (b) and Figure 3-7 (b)) reflecting intensity difference between them in decoded images. Such intensity difference depends on material density. Also, note that the whole 3D shape of inside objects is fully recovered in the both results.
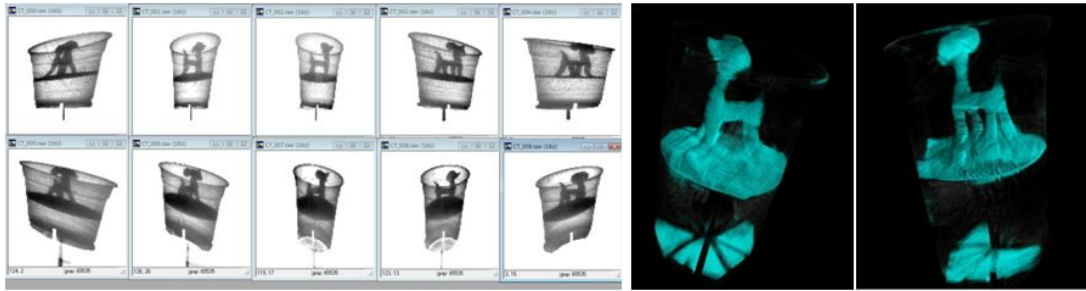
(a) Captured images at 8 different viewing angles      (b) 3D reconstrction by ART

**Figure 3-6:** *Tomographic reconstruction for a dog-shape objects.*



(a) Captured images at 8 different viewing angles      (b) 3D reconstrction by ART

**Figure 3-7:** *Tomographic reconstruction for a wine glass.*

# Chapter 4

# Benefits, Limitations and Conclusion

## 4.1 Benefits

This thesis makes four primary contributions: (1) robust separation of direct and scattered components of incident light passing through heterogenous scattering media, (2) 3D volumetric reconstruction of the mixed scattering objects, (3) a novel technique to enable effective volumetric analysis of solid scattering objects in multiple-scattering conditions, and (4) a single-shot 3D reconstruction technique applicable to translucent objects. For direct and scattered separation, this thesis's approach requires only a simple system consisting of a lenslet (or pinhole) array and point light sources. Compared with other methods, like using a projector to generate temporally-multiplexed patterns, this method can achieve separation in a single exposure. Also, the method requires simple local computations, performed independently on each lenslet or pinhole image. Furthermore, dynamic applications are possible.

(a) Inset Photo          (b) Well-focused direct-only          (c) Poor-focused direct-only

**Figure 4-1:** *Direct-only images by Nayar et al. method for a horse-shaped object enclosed in an aquarium with diluted milk(Concentration 0.11%) (a) Inset of a captured photo with a projected high-frequency pattern (b) Direct-only image in a well-focused case (c) Direct-only image in a poor-focused case*

This thesis's transmission-mode descattering technique has potential applications extending beyond computer vision and graphics, including non-invasive medical imaging [33]. Specific parts of biological organisms, including human fingers, can be modeled by scattering and translucent material which is similar to objects considered in this thesis work. As a result, it is possible to view the inner 3D shape of certain parts in human and animal bodies by this technique. Most importantly, such visible-wavelength separation methods may allow hazardous X-ray imaging to be replaced in certain applications. Such applications include the personal identification field. For example, 3D shape recognition of finger veins can provide strong cues for identification. Furthermore, such features may overcome several limitations of traditional fingerprints, which change due to aging.

The proposed method has several unique advantages in comparison to the closely related method of Nayar et al. [17] of which key limitations is that the assumption of high-frequency projected patterns aren't satisfied in dense media(Figure 4-1 (a)). Another limitation of any projector-based solution arises due to the finite depth of field achieved in practice. For transmission-mode descattering, the projector must focus on the scattering media and the screen at the same time—unlike the case of reflection-mode acquisition dealt in Nayar et al. [17]. Thus, the projector requires a wide depth of field. Figure 4-1(c) shows a direct image by [17] when the projector's DOF isn't wide enough to cover both inside object and screen. This thesis's method is free from such focusing problems as well as the imaging system is much simpler and inexpensive, containing a single LED. Finally, this thesis's approach is well-suited for less dense parts of human bodis, such as fingers as shown in Figure 2-8.
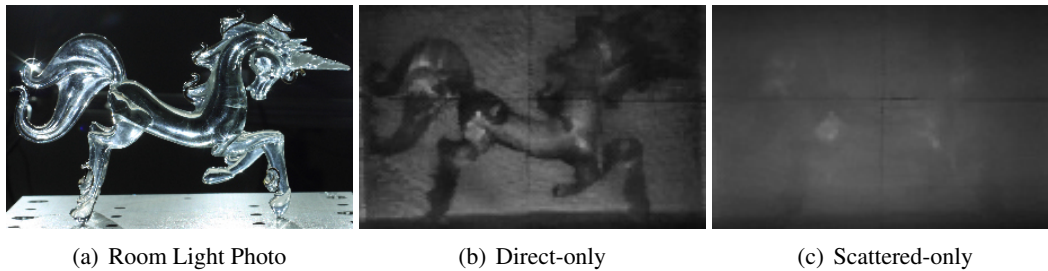
(a) Room Light Photo          (b) Direct-only          (c) Scattered-only

**Figure 4-2:** *Limitation of refraction. Results for a translucent horse-shaped object enclosed in an aquarium with diluted milk(Concentration 0.11%).*

## 4.2  Limitations

In the current method, the primary limitation is due to the loss of resolution incurred by a lenslet or pinhole array. In addition, pinhole arrays require long exposures. While lenslets could be used to overcome exposure issues, loss of resolution remains. Also, the separated results can be affected by refraction. Figure 4-2 shows separated results of a translucent horse-shape object in milky water. Note that the legs and the end of the tail in (b) look dark by refraction although they have similar density with the body area as shown in (a). The proposed 3D reconstruction method requires control of the lighting environment and, as a result, cannot be directly extended to natural environments. Furthermore, this reconstruction method requires a temporally-multiplexed set of images for tomographic reconstruction, limiting dynamic scene reconstruction. This thesis emphasizes, however, that direct-scattered separation can be performed in a single exposure. Most importantly, challenges are anticipated in strongly-scattering environments. In such circumstances, the scattering term will dominate the proposed low-order polynomial approximation and the direct term will not be reliably recovered.

In the single-shot 3D reconstruction method, one of the most important factors for sound results is the quality of decoded images which is affected by the distance between lighting plane and the diffuser in Figure 3-1 (a) and the interval of lenslets (or pinholes). The smaller the distance is, the more distorted the side part of a light field image is, resulting in more overlap of white dots as shown in left image of Figure 4-3. In the other hand, when the distance becomes too large, the overlap among the white dots happens in the whole region as shown in the right of Figure 4-3. Such
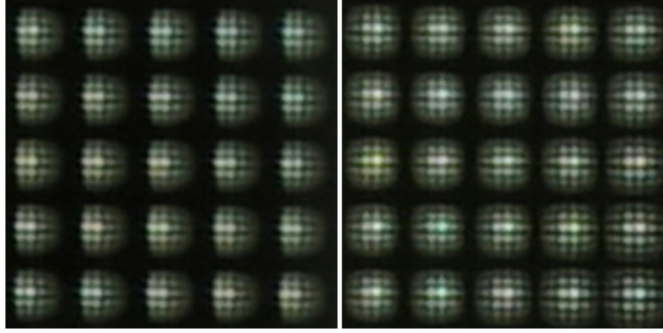
41

**Figure 4-3:** *Incorrect coded image when the distance between light plane and a diffuser is unsuitable (Left and right images are captured when the distance is too short and large, respectively)*
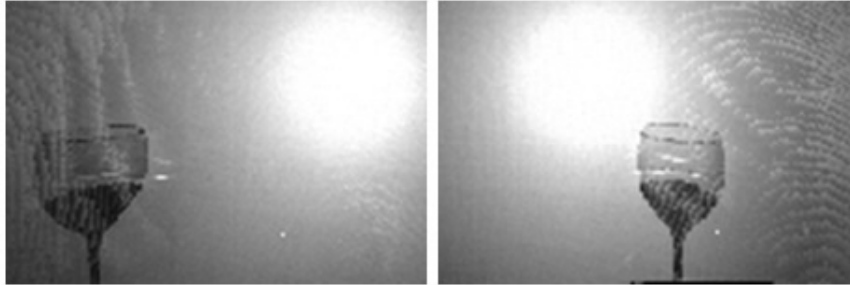


**Figure 4-4:** *Limitation due to overlap of regions (white dots), Figure 4-3, under lenslets*

overlap produces artifacts (Figure 4-4) in decoded images and in turn, degrades 3D reconstruction result.

3D reconstruction of an object with complex shape requires high-resolution decoded images. Current resolution, 150x100, is not enough to express such complex objects. To increase the resolution of decoded images, some factors should be considered. First, the number of lenslets (or pinholes) should be increased. Also, the sensor resolution of a camera should be high enough to capture the whole images formed by lenslets with no overlapping. Ideally, The interval of white dots should be same in the entire region since it depends on fixed parameters in Equation 4.1 assuming that LEDs are perfect point light sources. However, practically there is more possibility that the interval can be affected by limited illumination angle of LEDs in boundary region of a captured image than center region as shown in Figure 4-5. To avoid this matter, the system geometry should meet the condition of Equation 4.2.

$$I_w = \frac{D_d}{D_l} I_l \qquad (4.1)$$

where $I_w$ and $I_l$ represent the interval of white dots and LEDs, respectively. $D_l$ and $D_d$ are distances from a lenslet(pinhole) array to LED plane and diffuser plane, respectively.

$$acos\left(\frac{D_l}{D_{la}}\right) < \frac{\theta_{LED}}{2} \tag{4.2}$$

where $\theta_{LED}$ and $D_{la}$ indicate the angular limit of LED illumination and distance between a LED and a lenslet(pinhole) array.



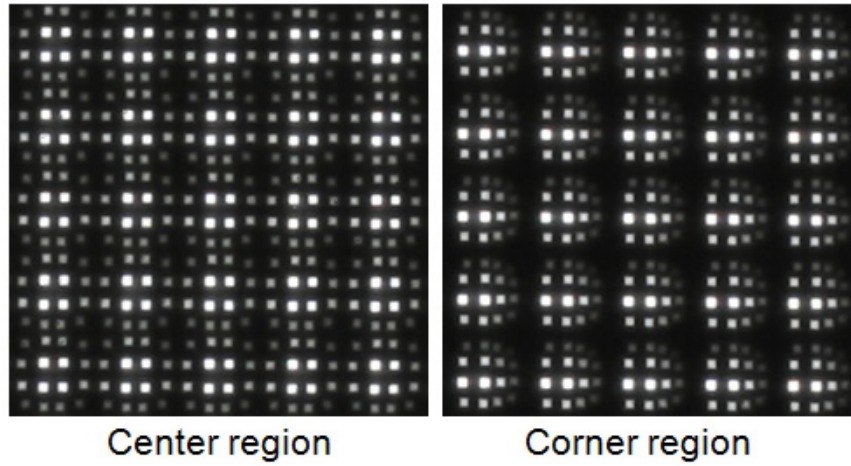Center region          Corner region

**Figure 4-5:** *Irregularity of a light field image to be considered as a setup parameter. The comparison of center and corner regions in a light field image reveals the irregularity in space between white spots.*

## 4.3   Conclusion

This thesis contributes the new concept of CAT system by proposing two novel techniques, transmission-mode descattering and single-shot 3D reconstruction. In the future, volume-inside information will be considered as another commercial paradigm. Currently 3D information is getting attention in various visualization markets. Current CAT technology can achieve high quality volume-inside information however its usage is limited to only inevitable cases due to the harmfulness of X-ray and its bulky system. This thesis hopes to open a new direction for less-harmful or safe CAT system with a compact system. Such features will play an important role in bringing casual usages of CAT system in our daily lives. Future system should consider diffraction /citeOh using augmented light field technique.

The first proposed method separates direct and scattered components of transmitted light through scattering media. The direct-only images provide sharp shape information for inside objects of the media. A volumetric reconstruction technique has been demonstrated following classic methods of limited-baseline tomography, to reconstruct the objects using direct-only images. These results can be achieved with low-cost hardware consisting of LEDs, diffusers, and printed masks with pinholes. In particular, this thesis shows that visible-wavelength radiation can be applied for attenuation-based tomography when such separation methods exist in the transmission-mode. I hope that this thesis's approach will inspire others to pursue low-energy, non-invasive imaging in the medical and biological sciences. The second method, single-shot 3D reconstruction, will be very applicable in scientific and medical imaging fields. Fundamentally, this research can significantly decrease the scanning time in current CAT system. Consequently, high-speed machines without mechanical movement will allow a new design paradigm for future CAT system.

# Bibliography

[1]  Adelson, T., Wang, J.Y.A.: Single lens stereo with a plenoptic camera. IEEE Transactions on Pattern Analysis and Machine Intelligence **14** (1992) 99–106

[2]  Levoy, M., Hanrahan, P.: Light field rendering. SIGGRAPH96 (1996) 31–42

[3]  Gortler, S.J., Grzeszczuk, R., Szeliski, R., Cohen, M.F.: The lumigraph. SIGGRAPH (1996)

[4]  Isaksen, A., Mcmillan, L., Gortler, S.J.: Dynamically reparameterized light fields. Proc. SIGGRAPH (2000)

[5]  Levoy, M., Chen, B., Vaish, V., Horowitz, M., Mcdowall, M., Bolas, M.: Synthetic aperture confocal imaging. ACM Transactions on Graphics **23** (2004) 825–834

[6]  Vaish, V., Wilburn, B., Joshi, N., Levoy, M.: Using plane + parallax for calibrating dense camera arrays. Proc. Conf. Computer Vision and Pattern Recognition (2004)

[7]  Veeraraghavan, A., Raskar, R., Agrawal, A., Mohan, A., Tumblin, J.: Mask enhanced cameras for heterodyned light fields and coded aperture refocusing. ACM SIGGRAPH 2007 (2007)

[8]  Raskar, R., Agrawal, A., Wilson, C.A., Veeraraghavan, A.: Glare aware photography: 4d ray sampling for reducing glare effects of camera lenses. ACM Trans. Graph. **27** (2008) 1–10

[9]  Fuchs, M., Raskar, R., Seidel, H., Lensch, H.: Towards 6d displays. SIGGRAPH (2008)

[10]  Hirsch, M., Lanman, D., Holtzman, H., Raskar, R.: A thin, depth-sensing lcd for 3d interaction using light fields. SIGGRAPH Asia (2009)

[11] Kim, J., Lanman, D., Mukaigawa, Y., Raskar, R.: Descattering transmission via angular filtering. ECCV (2010)

[12] Fujimoto, J.G.: Optical coherence tomography. C. R. Acad. Sci. Paris Serie IV 2 (2001) 10999–1111

[13] Boas, D.A., Dale, A.M., Franceschini, M.A.: Diffuse optical imaging of brain activation: approaches to optimizing image sensitivity, resolution and accuracy. NeuroImage **23** (2004) 275–288

[14] Wang, S., Zhang, X.C.: Pulsed terahertz tomography. JOURNAL OF PHYSICS D: APPLIED PHYSICS **37** (2004)

[15] Lanman, D., Raskar, R., Agrawal, A., Taubin, G.: Modeling and capturing 3d occluders. SIGGRAPH Asia 2008 (2008)

[16] Chai, J., Chan, S., Shum, H., Tong, X.: Plenoptic sampling. In: SIGGRAPH. (2000) 307–318

[17] Nayar, S., Krichnan, G., Grossberg, M., Raskar, R.: Fast separation of direct and global components of a scene using high frequency illumination. Transactions on Graphics **12** (2006) 935–943

[18] Nasu, O., Hiura, S., Sato, K.: Analysis of light transport based on the separation of direct and indirect omponents. IEEE Intl. Workshop on Projector-Comera Systems(ProCams 2007) (2007)

[19] Rosen, J., Abookasis, D.: Noninvasive optical imaging by speckle ensemble. Optics Letters **29** (2004)

[20] Okugawa, H.: A new imaging method for confocal microscopy. Proc. of SPIE (2008)

[21] Sheppard, C.J.R., Cogswell, C.J.: Confocal microscopy with detector arrays. Journal of Modern Optics **37** (1990) 267–279

[22] Levoy, M., Zhang, Z., Mcdowall, I.: Recording and controlling the 4d light field in a microscope using microlens arrays. J. of Microscopy (2009)

[23] Kim, J., Horstmeyer, R., Kim, I., Raskar, R.: Highlighted dof photography: Shining light on focus. TOG in review (2010)

[24] Narasimhan, S.G., Nayar, S.K., Sun, B., Koppal, S.J.: Structured light in scattering media. In Proc. IEEE ICCV **1** (2005) 420–427

[25] Gu, J., Nayar, S., Grinspun, E., Belhumeur, P., Ramamoorthi, R.: Compressive structured light for recovering inhomogeneous participating media. In European Conference on Computer Vision (ECCV) (2008)

[26] Atcheson, B., Ihrke, I., Heidrich, W., Tevs, A., Bradley, D., Magnor, M., Seidel, H.P.: Time-resolved 3d capture of non-stationary gas flows. ACM Transactions on Graphics (2008)

[27] Trifonov, B., Bradley, D., Heidrich, W.: Tomographic reconstruction of transparent objects. Eurographics Symposium on Rendering (2006)

[28] Narasimhan, S.G., Nayar, S.K.: Shedding light on the weather. In Proc. IEEE CVPR **151** (2003) 665–672

[29] Jensen, H., Marschner, S., Leboy, M., Hanrahan, P.: A practical model for subsurface light transport. SIGGRAPH (2001) 511–518

[30] Sun, B., Ramamoorthi, R., Narasimhan, S.G., Nayar, S.K.: A practical analytic single scattering model for real time rendering. ACM Transactions on Graphics (2005) 1040–1049

[31] N. Miura, A.N., Miyatake, T.: Feature extraction of finger-vein patterns based on repeated line tracking and its application to personal identification. Machine Vision and Applications (2004)

[32] Roh, Y.J., Park, W.S., Cho, H.S., Jeon, H.J.: Implementation of uniform and simultaneous art for 3-d reconstruction in an x-ray imaging system. IEEE Proceedings, Vision, Image and Signal Processing **151** (2004)

[33] Tuchin, V.: Tissue optics. SPIE (2000)

[34] Narasimhan, S.G., Nayar, S.K., Sun, B., Koppal, S.J.: Structured light in scattering media. In Proc. IEEE ICCV **1** (2005) 420–427

[35] Ng, R., Leboy, M., Bredif, M., Duval, M., Horowitz, G., Hanrahan, P.: Light field photography with a hand-held plenoptic camera. Tech. rep, Stanford University (2004)

[36] Levoy, M., Ng, R., Adams, A., Footer, M., Horowitz, M.: Light field microscopy. ACM Transactions on Graphics (2006)

[37] Oh, S.B., Kashyap, S., Garg, R., Chandran, S., Raskar, R.: Renderingwave effects with augmented light field. EUROGRAPHICS **29** (2010)