

Research Paper

Hemispherical Confocal Imaging

SEIICHI TAGAWA,^{†1} YASUHIRO MUKAIGAWA,^{†1,†2}
 JAEWON KIM,^{†2,†3} RAMESH RASKAR,^{†2}
 YASUYUKI MATSUSHITA^{†4} and YASUSHI YAGI^{†1}

We propose a new imaging method called *hemispherical confocal imaging* to visualize clearly a particular depth in a 3-D scene. The key optical component is a *turtleback reflector* which is a specially designed polyhedral mirror. To synthesize a hemispherical aperture, we combined the turtleback reflector with a coaxial camera and projector, to create on a hemisphere many virtual cameras and projectors with a uniform density. In such an optical device, high frequency illumination can be focused at a particular depth in the scene to visualize only that depth by employing descattering. The observed views are then factorized into masking, attenuation, reflected light, illuminance, and texture terms to enhance the visualization when obstacles are present. Experiments using a prototype system show that only a particular depth is effectively illuminated, and hazes caused by scattering and attenuation can be recovered even when obstacles are present.

1. Introduction

Significant effort has been made to obtain cross-sectional views of 3-D scenes. Real scenes often include obstacles such as scattering materials or opaque occluders. To visualize the cross-sectional views clearly as if the scene is sliced by a plane, only a particular depth must be illuminated, and haze due to scattering and attenuation should be removed.

The simplest way to observe a particular depth is to use a large aperture lens. The large aperture makes the depth of field (DOF) shallow, and the region outside the DOF is blurred. The synthetic aperture method²⁵⁾ mimics a large virtual

aperture by combining many small apertures. However, the obstacles are still bright and visible, even though they are blurred. The confocal imaging method¹⁶⁾ simultaneously scans two confocal pinholes over a particular depth. Because both the illumination and the observation are focused, clear cross-sectional views are obtained. While still visible, the obstacles are darkened and blurred, and the scanning requires long measuring times.

Recently, Levoy et al.¹²⁾ proposed a new imaging technique that combines the synthetic aperture method with the confocal imaging method. Since this technique is based on light field analysis, only a particular depth can be illuminated without scanning. However, the synthesized aperture size is relatively small because the rectangular mirrors are aligned as a 2D array. Furthermore, unwanted effects such as scattering and attenuation remain.

In this paper, we propose a novel imaging method called *hemispherical confocal imaging*. To improve the imaging performance of synthetic aperture confocal imaging¹²⁾, we designed a *turtleback reflector* which is a polyhedral mirror used to approximate a hemispherical aperture with a 180° FOV (field of view). We introduce *focused high frequency illumination* using the turtleback reflector with a projector. This method can eliminate scattering at the focused depth, and makes the unfocused depth almost invisible. We also introduce *factorization* of the observed views to eliminate attenuation.

Contribution

- By using the new optical device, the unfocused depth becomes almost invisible, the scattering is eliminated and the measurement is very fast, because no scanning is required.
- We have designed a *turtleback reflector*, which is a polyhedral mirror circumscribed in an ellipsoid. The combination of the turtleback reflector, a projector, and a camera can synthesize a hemispherical wide aperture. The optical device can also be used for measuring the complete 8-D reflectance field on the hemisphere.
- A new analysis of the 8-D reflectance field is introduced to obtain clear views. This analysis removes the unnecessary effects of scattering and attenuation in captured images.

^{†1} ISIR, Osaka University

^{†2} MIT Media Lab

^{†3} Korea Institute of Science and Technology (KIST)

^{†4} Microsoft Research Asia

2. Related Work

2.1 Reflectance Field Measurement

The optical device proposed in this paper can be regarded as an 8-D reflectance field measuring device. A 4-D slice of the 8-D reflectance field under static illumination can be recorded by a scanning camera¹³⁾ or installing multiple cameras²⁵⁾. Alternatively, a high-resolution camera can be combined with a micro lens array¹⁾, a micro mirror array²³⁾, or masks²⁶⁾. To vary the illumination, Debevec et al.⁴⁾ rotated the light source, and Sen et al.²¹⁾ used a projector as the light source. Masselus et al.¹⁴⁾ rotated the projector, and Matusik et al.¹⁵⁾ rotated both the light source and the camera to measure a 6-D reflectance field. Müller et al.¹⁸⁾ used 151 cameras with flashes.

In principle, the complete 8-D reflectance field can be measured by installing many projectors and cameras densely on a hemisphere. However, it is difficult to realize such a system because of the cost and the physical interference between the devices. While a rotating projector and camera solves these problems, the capture process is impractically long. Recently, Garg et al.⁶⁾ and Levoy et al.¹²⁾ used multiple planar mirrors and Cossairt et al.²⁾ used a lens array to measure a part of the 8-D reflectance field, but the direction of the illumination and observation was limited to a narrow angle.

However, our system can measure on the hemisphere the complete 8-D reflectance field covering the scene. Because our system uses the geometric properties of an ellipsoid, many virtual cameras and projectors can be produced on a hemisphere with *uniform density*.

2.2 BRDF Measurement Using Mirrors

For measurement of the bidirectional reflectance distribution function (BRDF), mirrors are often used to replace a mechanical motion. Reflected light can be measured effectively from all directions using a hemispherical mirror²⁷⁾ or a paraboloidal mirror³⁾. Recently, a wide variety of mirrors such as a cylindrical mirror¹¹⁾, several plane mirrors⁹⁾, an ellipsoidal mirror¹⁷⁾, and a combination of a paraboloidal mirror and a specially-designed dome mirror⁷⁾ have been used in conjunction with a projector and a camera for these measurements.

Our turtleback reflector design was inspired by the BRDF measurement method

using an ellipsoidal mirror¹⁷⁾. We used the geometric property of a rotationally symmetrical ellipsoid where all rays from one focal point reflect from the ellipsoidal mirror and reach the other focal point. However, we used a different geometrical property of the ellipsoid, where the total length from one focal point to the other through any surface point is constant. By using this characteristic, we can produce virtual cameras and projectors at a constant distance from the target just as they would be on a hemisphere. Because our purpose is not BRDF measurement but visualization of cross-sectional views, we designed a polyhedral mirror circumscribed in an ellipsoid.

2.3 Descattering

Incident light upon murky liquids or translucent media scatters, and the appearance becomes blurred. To obtain clear views, descattering methods have been developed. Treibitz and Schechner²²⁾ used a polarizer under water. Kim et al.¹⁰⁾ used a lenslet array. Assuming that only single scattering is observed in optically thin media, Narasimhan et al.¹⁹⁾ estimated a 3-D shape with descattering and Gu et al.⁸⁾ estimated the 3-D distribution of inhomogeneous scattering media.

Recently, Fuchs et al.⁵⁾ combined confocal imaging with descattering, which captured a descattered image at a particular depth in a 3-D scene. This technique consisted of the synthetic aperture confocal imaging proposed by Levoy et al.¹²⁾ and the high frequency illumination proposed by Nayar et al.²⁰⁾. While our method is based on this idea, we combine a factorization method to remove the effect of any attenuation remaining in the descattered images.

3. Hemispherical Confocal Imaging

Let us assume that a 3-D scene is illuminated by a light source and observed by a camera as shown in **Fig. 1**. Even if the camera is focused on a particular depth in the scene, the captured image includes reflections from the entire scene. To observe the particular depth, only that depth should be illuminated. This means that both the illumination and the observation should have a shallow DOF.

Even if we succeed in illuminating only the particular depth, clear views cannot be observed. The major reasons for this are *scattering* and *attenuation*. The scattering is caused by multi-bounce reflections in the translucent media, which

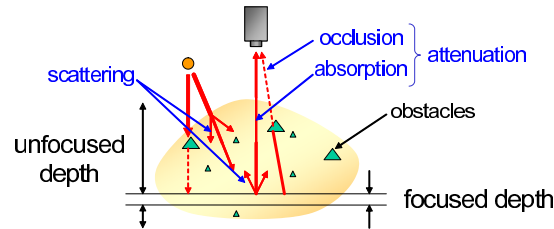


Fig. 1 Illumination and reflection in a 3-D scene. It is difficult to observe a particular depth due to scattering and attenuation.

cause the views to become blurred. The attenuation is caused by *occlusion* due to obstacles or *absorption* due to low transparency media. Under this attenuation, the illumination becomes nonuniform and the reflections are partially darkened. The following four functions are required to obtain clear views of a particular depth in a 3-D scene.

- The DOF should be as shallow as possible.
- Only the particular depth of interest should be illuminated.
- Scattering should be eliminated.
- Attenuation should be eliminated.

To satisfy these requirements, we propose *hemispherical confocal imaging* consisting of (1) a specially designed *turtleback reflector*, (2) *focused high frequency illumination*, and (3) *factorization* of observed views.

The turtleback reflector with coaxial camera and projector synthesizes a hemispherical aperture for both illumination and observation to satisfy (a). The focused high frequency illumination eliminates reflections from the unfocused depth and global reflection to solve (b) and (c). We then factorized the observed views into masking, attenuation, reflected light, illuminance, and texture terms to solve for (d). The advantages and disadvantages and the numbers of projectors and cameras for several imaging methods are summarized in **Table 1** and **Fig. 2**. Although we explained that unfocused depths are ‘unilluminated’ by the synthetic aperture confocal imaging and our hemispherical confocal imaging, the depths are actually illuminated, but these illuminations are eliminated by subtractive calculation. Hence, image noise may increase.

Table 1 Comparison of several imaging methods.

	unfocused depth	scanning	scattering
Synthetic aperture	bright	unnecessary	remaining
Confocal imaging	darken	necessary	remaining
Synthetic aperture confocal imaging ¹²⁾	unilluminated	unnecessary	remaining
Confocal imaging with descattering ⁵⁾	darken	necessary	reduced
Our hemispherical confocal imaging	unilluminated	unnecessary	reduced

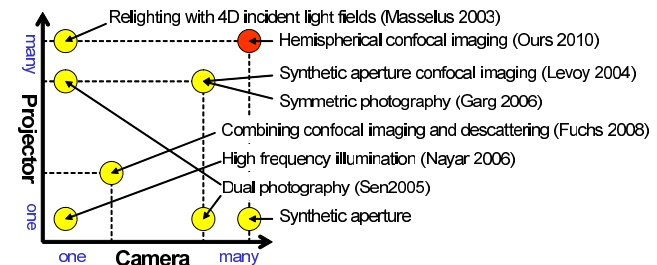


Fig. 2 The numbers of projectors and cameras of several imaging methods which use projector(s) and camera(s) for reflection analysis or reflectance field measurement.

4. Turtleback Reflector

4.1 Hemispherical Synthetic Aperture

To illuminate and capture a particular depth in the scene, the light source and the camera must have an extremely shallow DOF. Such a DOF can be realized using a lens with a large aperture. Theoretically, the DOF can be zero if the aperture size is infinity. On such an aperture, the FOV becomes 180° and the F-number becomes zero. We call this the *hemispherical aperture*. A lens with a *hemispherical aperture* would focus on only one particular depth. Although such a lens does not exist, the aperture can be produced by the synthetic aperture technique.

To synthesize the *hemispherical aperture*, cameras should be placed uniformly on a hemispherical surface. For this purpose, there are two solutions. One is to place real cameras on a hemisphere. The other is to rotate a camera mechanically on a hemisphere. The former approach is difficult because of cost and physical conflicts. The latter approach is impractical because of time costs.

To realize the hemispherical aperture, we used planar mirrors. By combining a camera with many planar mirrors, a number of virtual cameras with low resolution can be generated and can then capture the scene from various directions.

Views from the virtual cameras are included in a view from the real camera. Hence, the corresponding region for each virtual camera must be cropped and warped to obtain images. The geometric calibration of the virtual cameras is performed by finding corresponding points in a captured image using the real camera. The geometric conversion from the real camera to virtual cameras can be achieved using simple homography. The pixel intensities of the virtual image are resampled from the captured image.

4.2 Mirror Design

For the hemispherical synthetic aperture, the planar mirrors should be positioned, so that they place the virtual cameras on a hemisphere. The distance between the target and each virtual camera should be constant. The approach used to produce such an arrangement of planar mirrors varies between the different projections of the real camera.

For a perspective projection, the planar mirrors are located along an ellipsoid as shown in **Fig. 3** (a). All rays from one focal point reflect from the ellipsoid and reach the other focal point with a constant length. A target object is placed at one focal point, and a real camera is placed at the other. The virtual cameras are then placed on a hemisphere.

On the other hand, if the projection is orthographic, the planar mirrors are located along a paraboloid as shown in **Fig. 3** (b). At a point on the paraboloid, the distances from the focal point and the directrix are equivalent. The virtual cameras, therefore, are placed on a hemisphere.

In the design using the paraboloid, a telecentric lens is needed. However, the telecentric lens is expensive. Hence, we adopted the design using an ellipsoid because it only requires an ordinary lens.

4.3 Polyhedral Mirror Sampling a Hemispherical Aperture

In practice, the number of virtual cameras is finite. These cameras are discretely placed on a hemisphere. For the synthetic aperture technique, the virtual cameras should be placed with as uniform a density as possible.

As shown in **Fig. 4** (a), a geodesic dome is a well-known polyhedron that uni-

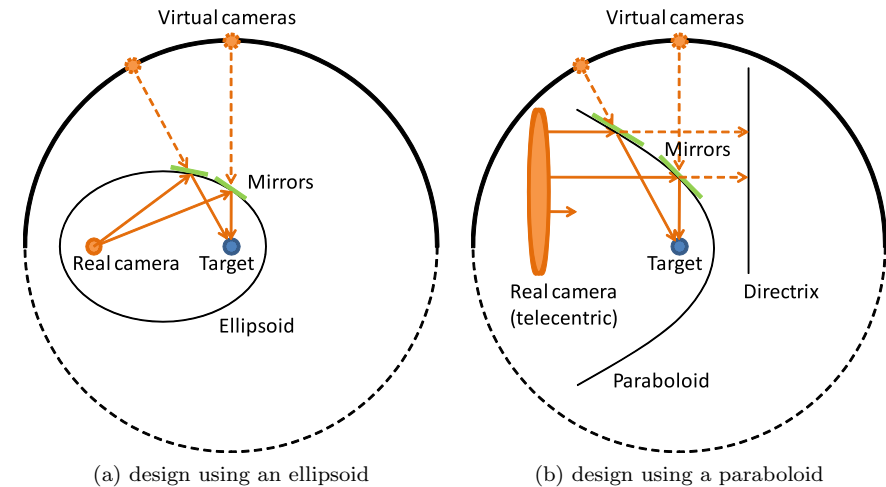


Fig. 3 Principle of the proposed optical device. Virtual cameras are placed on a hemisphere by using planar mirrors.

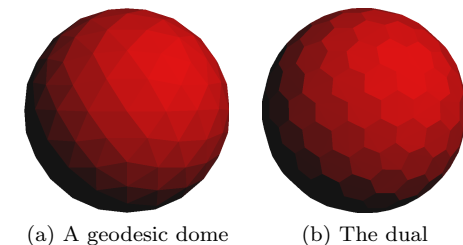


Fig. 4 Geodesic domes.

formly samples a sphere. The virtual cameras are located on the facets of the geodesic dome. The hemispherical aperture is then divided by the facets as uniformly as possible. The mirrors are located on tangent planes on the ellipsoid. The tangent point is given by the projection of the center of the facet from the center of the dome. To decide the mirror shape, the vertices of the facet are projected on to the tangent plane. These mirrors become the polyhedral mirrors, which place virtual cameras on a hemisphere with uniform density.

Each facet of the polyhedron acts as a viewing window corresponding to the

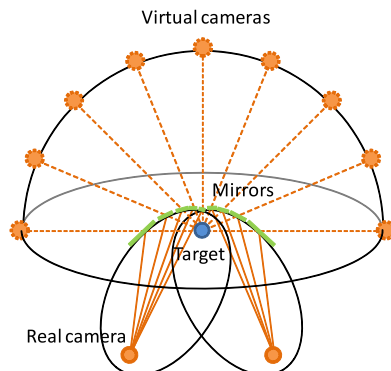


Fig. 5 Design of polyhedral mirrors using two ellipsoids.

virtual camera. A region commonly observed from all virtual cameras is defined by a logical AND of all the views of the virtual camera. That is, the shape of a facet should be close to a circle in order to observe a large area. Therefore, we used the dual geodesic dome shown in Fig. 4 (b) instead of the original geodesic dome, because the dual dome has pentagonal and hexagonal facets, while the original geodesic dome had triangular facets.

4.4 Design of the Turtleback Reflector

We now show an implementation of our polyhedral mirror. We first decided on the number of mirrors. The geodesic dome is made by dividing 20 facets of an icosahedron into 20×4^k ($k \in \mathbf{N}$) facets. The number of facets of the dual dome is then $N = 10 \times 4^k + 2$. We set $k = 2$ and $N = 162$. The upper half hemisphere consists of 81 facets. However, 81 facets do not completely cover the upper hemisphere. Hence we added 10 horizontal facets, which are located at the border of the upper and lower hemispheres. In total we used 91 facets.

In the design shown in Fig. 3 (a), the target object occludes some mirrors. Hence, we combined two ellipsoids as shown in **Fig. 5**. In this design, the two polyhedral mirrors place the virtual cameras over the whole surface of the hemisphere.

We designed one polyhedral mirror because pairs of polyhedral mirrors are symmetrical. To fix the mirror patches, we designed a frame as shown in **Fig. 6**.

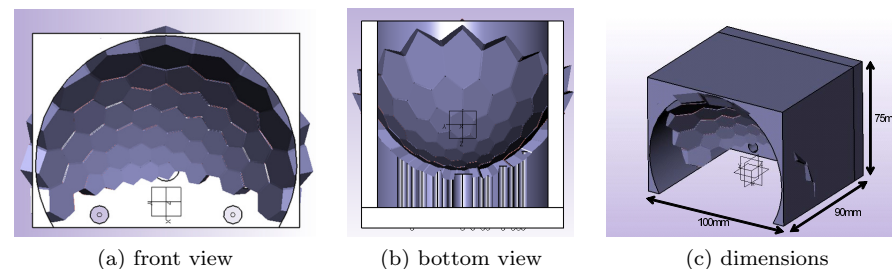


Fig. 6 Design of the frame for the turtleback reflector.



Fig. 7 Turtleback reflector.

This frame is made by stereolithography and planar mirror patches are attached to the frame. Fifty mirror patches completely cover half of the hemisphere. The frame with mirror patches is the *turtleback reflector* as shown in **Fig. 7**.

4.5 Total Optical Device

In this research, we combine the turtleback reflector with a coaxial pairing of a high-resolution camera (PointGrey, Grass-50S5C, $2,448 \times 2,048$) and a small projector (KAIREN Projector X Pro920, 640×480) using a beam splitter as shown in **Fig. 8**.

The device is designed to measure a small area of 6×6 mm. By resampling the captured image using a real camera, images of the virtual cameras whose resolution is 60×60 pixels are generated. For a virtual projector, we resample

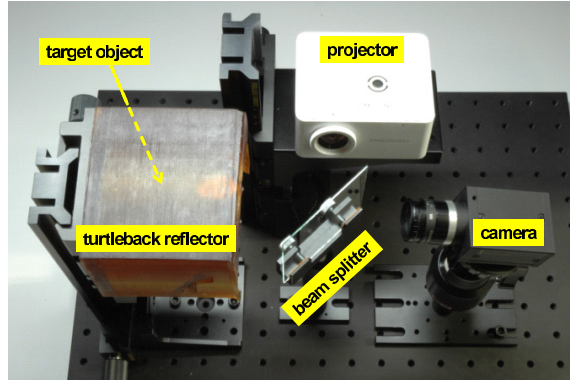


Fig. 8 Total optical device for hemispherical confocal imaging.

the target area using a 20×20 format. Hence, the resolution of the virtual projector corresponds to 20×20 pixels.

Theoretically, both the illumination and observation can be focused only on a particular depth in the scene and the DOF becomes zero. However, we implemented only half of the whole turtleback reflector. Therefore, strictly the DOF of our prototype system is not zero but shallow.

5. Focused High Frequency Illumination

5.1 Illumination and Reflection in a 3-D Scene

To analyze the reflectance field in a 3-D scene, we need to know how light illuminates points in a scene, and how the reflections are observed. We divide the 3-D scene into a set of small voxels. Let L be an illumination on the scene, and $R_k(L)$ be an image component that consists of a set of reflected and scattered lights from the k -th voxel under the illumination L as shown in **Fig. 9** (a). Because the observed image of the entire scene by a camera is expressed as a sum of the reflected and scattered lights from all voxels, the image is represented by $\sum_i R_i(L)$.

Illuminations and reflections can be regarded as the sum of the direct and global components²⁰⁾. As shown in **Fig. 9** (b), the illumination can be decomposed into direct L^D and global illuminations L^G as

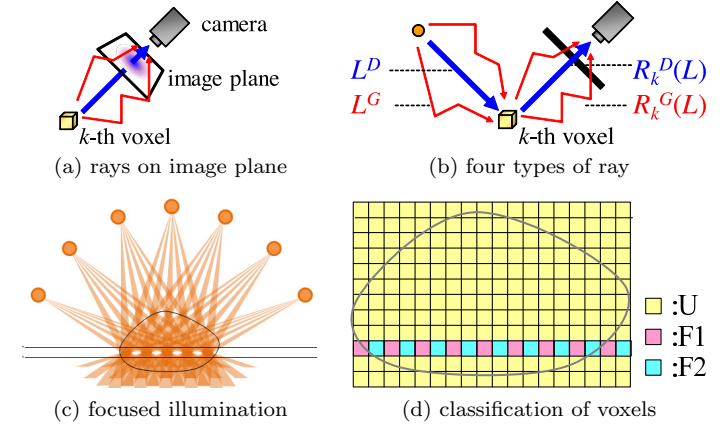


Fig. 9 Focused high frequency illumination. The high frequency patterns are focused only on the particular depth. The projection is blurred out of the DOF.

$$L = L^D + L^G. \quad (1)$$

Similarly, the reflection can also be decomposed into the direct $R_k^D(L)$ and global reflections $R_k^G(L)$. The R_k^D is recorded as the pixel intensity at the intersection of the image plane and the line between the k -th voxel and the camera, while the R_k^G is recorded as the part of the intensity of other pixels as shown in **Fig. 9** (a).

The image component $R_k(L)$ can be decomposed into the direct and global components as

$$R_k(L) = R_k^D(L) + R_k^G(L). \quad (2)$$

It is easy to assume that the R_k is a linear function by disabling the gamma correction of the camera. Hence, the observed image can be modeled as the sum of four components by

$$\sum_i R_i(L) = \sum_i R_i^D(L^D) + \sum_i R_i^D(L^G) + \sum_i R_i^G(L^D) + \sum_i R_i^G(L^G). \quad (3)$$

5.2 Focused Illumination by Multiple Projectors

To obtain clear views of a particular depth in a 3-D scene, only that depth should be illuminated. Any global illuminations and global reflections should be eliminated to reduce scattering in the media. That is, direct reflections by direct

Table 2 Relative intensities of four reflection components for each voxel type.

	$R_k^D(L^D)$	$R_k^D(L^G)$	$R_k^G(L^D)$	$R_k^G(L^G)$
U (unfocused)	1/2	1/2	1/2	1/2
F1 (focused and illuminated)	1	1/2	1	1/2
F2 (focused and unilluminated)	0	1/2	0	1/2

illuminations, $R_k^D(L^D)$, from the particular depth should be measured separately.

Fuchs et al.⁵⁾ proposed the descattered confocal imaging technique to capture a clear view of a particular depth in scattering media. They combined the focused illumination technique proposed by Levoy et al.¹²⁾ and the high frequency illumination technique proposed by Nayar et al.²⁰⁾. They captured clear descattered images by line scanning confocal imaging. However, this technique could not deal with reflections from opaque objects in the unfocused depth. The reflections were introduced as the floodlit contribution by Levoy et al.¹²⁾. They removed them using many observations under a lot of random pattern illuminations.

Here, we propose a new imaging method, called *focused high frequency illumination* (**FHFI**), based on descattered confocal imaging⁵⁾. Our technique can remove the floodlit contribution and requires only constant measurement time. The FHFI uses a lot of projectors placed over a wide range on a hemisphere.

For the FHFI, high frequency checkerboard patterns are projected from each projector. The positions of the white and black pixels are aligned at the depth as shown in Fig. 9 (c). This means that the high frequency illumination is focused only at a particular depth. The voxels in the scene are classified into unfocused voxels *U*, focused and illuminated voxels *F1*, and focused but unilluminated voxels *F2* as shown in Fig. 9 (d).

It is noted that synthetic aperture illumination with regular patterns like checkerboards often cause artifacts of periodic patterns at depths other than the focal plane²⁸⁾. This artifact occurs when the projected patterns are accidentally aligned at different depths. This problem can be reduced by the hemispherical placement of the virtual projectors because projected patterns do not align at different depths.

When compared with a white pattern, the average intensity of the high frequency illumination is darker because half the pixels are black. **Table 2** shows

the relative intensities of the four reflection components for each voxel type. The global illumination to each voxel decreases by half. The direct illumination to *U* voxels also decreases by half because the projected patterns are blurred. Therefore, the floodlit contribution decreases by half under the FHFI using the turtleback reflector. The *F1* voxels receive full direct illumination, while the *F2* voxels receive no direct illumination. By combining these differences, $\sum_{i \in F1 \cup F2} R_i^D(L^D)$ which enables only direct components from voxels at the focused depth to be separated.

Let I_P be a captured image when voxels of *F1* are illuminated but voxels of *F2* are not illuminated. Let I_N be a captured image when the inverse pattern is projected. Then, these images can be expressed as

$$I_P = \sum_{i \in F1} R_i \left(L^D + \frac{L^G}{2} \right) + \sum_{i \in F2} R_i \left(\frac{L^G}{2} \right) + \sum_{i \in U} R_i \left(\frac{L^D + L^G}{2} \right), \quad (4)$$

$$I_N = \sum_{i \in F1} R_i \left(\frac{L^G}{2} \right) + \sum_{i \in F2} R_i \left(L^D + \frac{L^G}{2} \right) + \sum_{i \in U} R_i \left(\frac{L^D + L^G}{2} \right). \quad (5)$$

The absolute difference of two intensities in I_P and I_N is then calculated for each pixel. Since this is a pixel-wise process, the pixel intensities at coordinates (x, y) in I_P , I_N , and $R(L)$ are expressed as $I_P(x, y)$, $I_N(x, y)$, and $R(L, x, y)$. The absolute difference I_D is calculated by

$$\begin{aligned} I_D(x, y) &= |I_P(x, y) - I_N(x, y)| \\ &= \left| \sum_{i \in F1} R_i(L^D, x, y) - \sum_{i \in F2} R_i(L^D, x, y) \right| \\ &= \left| \sum_{i \in F1} R_i^D(L^D, x, y) + \sum_{i \in F1} R_i^G(L^D, x, y) \right. \\ &\quad \left. - \sum_{i \in F2} R_i^D(L^D, x, y) - \sum_{i \in F2} R_i^G(L^D, x, y) \right|. \end{aligned} \quad (6)$$

If we can assume a locally homogeneous surface, the global component does not change even if the phase of the high frequency pattern shifts^{5),20)}. Therefore,

$$\sum_{i \in F1} R_i^G(L^D, x, y) \simeq \sum_{i \in F2} R_i^G(L^D, x, y). \quad (7)$$

Moreover, $\sum_{i \in F1} R_i(L^D, x, y)$ and $\sum_{i \in F2} R_i(L^D, x, y)$ are exclusive. That is,

when the voxels of $F1$ are illuminated, the former becomes large while the latter becomes zero and vice versa. Hence,

$$\begin{aligned}
 I_D(x, y) &= \left| \sum_{i \in F1} R_i^D(L^D, x, y) - \sum_{i \in F2} R_i^D(L^D, x, y) \right| \\
 &= \sum_{i \in F1} R_i^D(L^D, x, y) + \sum_{i \in F2} R_i^D(L^D, x, y) \\
 &= \sum_{i \in F1 \cup F2} R_i^D(L^D, x, y).
 \end{aligned} \tag{8}$$

This means that only the particular depth ($F1 \cup F2$) can be directly illuminated without global illuminations, and only the direct reflections can be measured without global reflections. As shown in Table 1, our method does not illuminate the unfocused depth. Because no scanning is necessary, the measurement is fast. Furthermore, scattering, which is a major global component in translucent media, is eliminated.

6. Factorization of the Observed Views

6.1 Attenuation of Incident and Reflected Light

By the FHFI introduced in the previous section, only the focused depth is illuminated and global components such as scattering are eliminated. However, the obtained view may still be unclear because illumination and reflected light attenuate. The reason for the attenuation is the *occlusion* and *absorption* as shown in Fig. 1.

These are similar but different optical phenomena. Occlusion directly interrupts reflected light by opaque obstacles. Hence, it causes very dark regions with sharp edges in the observed image. On the other hand, absorption decreases lighting powers by low transparency media. It causes smooth unevenness in the observed image. To obtain a clear view at a particular depth of the scene, the optical effects of occlusion and absorption should be eliminated.

6.2 Factorization Using Multiple Images

The attenuation occurs in both observation and illumination. However, if the observation and illumination are coaxial, the attenuation effects are the same. Hence, we can estimate the attenuation of the observation.

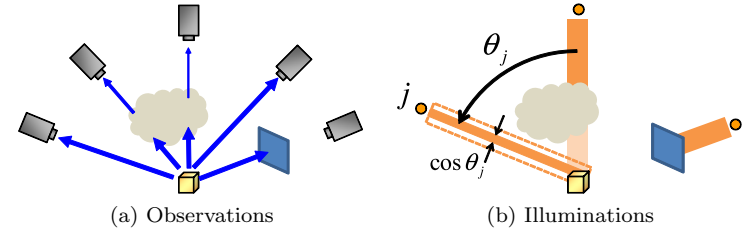


Fig. 10 Attenuation effects.

When a region in a captured image is dark, we cannot directly know why the reflectance is low or the illumination is weak. It is difficult to eliminate attenuation effects using only a single image. Fortunately, the scene is observed from many virtual cameras. Even if some lights are not observed from a camera, other cameras may observe the scene without attenuation as shown in Fig. 10 (a). Hence, we try to estimate texture that is not affected by attenuation based on observation from multiple cameras.

To factorize the observed intensities into reflected light and the attenuation, we assume that surface reflections can be expressed by the Lambert model. That is, the observed brightness is independent of the viewing angle. We assume that there are K virtual cameras and each camera has N pixels. Let O_{ij} be the intensity of the i -th pixel in the j -th camera. In our model, the observed intensities are factorized as

$$O_{ij} = M_{ij} A_{ij} R_i. \tag{9}$$

Here, M_{ij} is the *masking term* which has a value of 0 or 1. If the light is occluded by obstacles, the value becomes 0 otherwise it becomes 1. A_{ij} is the *attenuation term*, which expresses light attenuation due to absorption. R_i is the *reflected light term* which expresses reflected light of the particular depth under attenuated illumination. It is noted that only the reflected light term is independent of the viewing direction, assuming Lambertian reflection. Figure 11 illustrates this relationship.

The flow of the factorization process is as follows

STEP-1: First, the masking term is decided. Since unfocused depths are not illuminated by the FHFI, obstacles can easily be distinguished

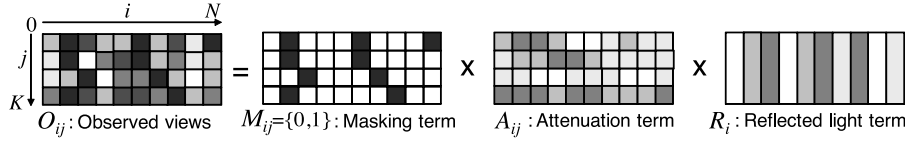


Fig. 11 Concept of factorization. The observed intensities are factorized into masking, attenuation, and reflected light terms to reduce attenuation.

using a simple threshold. After making the decision on the masking term, the following processes are conducted for pixels satisfying $M_{ij} = 1$.

STEP-2: The initial attenuation term is decided as $A_{ij} = 1$.

STEP-3: The reflected light term is calculated. Ideally, a unique reflected light should be estimated despite the different camera j , but the observed intensities vary. This kind of problem is often seen in stereoscopy²⁴⁾, so we used a median filter in a similar fashion by $R_i = \text{Median}(O_{ij}/A_{ij})$.

STEP-4: The attenuation term is updated by $A_{ij} = O_{ij}/R_i$ to satisfy Eq. (9).

STEP-5: The attenuation term is smoothed using a Gaussian function, because the attenuation varies smoothly over the 3-D scene. After this, return to **STEP-3** and repeat until the texture term does not change.

Using this factorization process, the observed views are decomposed to three terms and we can obtain the reflected light of the particular depth without attenuation of the observation.

Although the reflected light term R_i is estimated, it includes the attenuation effects of the illumination. Hence, we estimate the texture term T_i , which is independent of the illumination, by eliminating attenuation effects in the reflected light term. First, a maximum illuminance on the surface is estimated. The maximum illuminance is the illuminance when there are no obstacles. The illuminance from the j -th virtual projector weakens in proportion to $\cos \theta_j$, where θ_j is the zenith angle as shown in Fig. 10 (b). Here, we can assume that the illuminance is uniform over the surface by assuming that the observation area is

small. Hence, the maximum illuminance L_i^m for i -th pixel can be expressed by the sum of illuminance from all virtual projectors as follows,

$$L_i^m = \sum_j \cos \theta_j. \quad (10)$$

Since the j -th virtual projector and the j -th virtual camera are placed at the same position, we can regard the attenuation term A_{ij} as the attenuations of the illumination from the j -th virtual projector on the i -th pixel. Hence, the sum of the attenuated illuminance L_i^a is expressed as

$$L_i^a = \sum_j M_{ij} A_{ij} \cos \theta_j. \quad (11)$$

Therefore, the texture term T_i is restored by

$$T_i = R_i L_i^m / L_i^a. \quad (12)$$

Through these processes, the observed views are decomposed into a masking term, attenuation term, illuminance, and texture term. In this texture term, both the attenuation effects of the observation and the illumination are eliminated.

6.3 Stability of the Masking Term

Pixel-wise masking is often used for merging multiple images. Wilburn et al.²⁸⁾ used a mask to reconstruct an alias-free synthetic aperture image. They decided on a mask based on the variance of corresponding pixels.

On the other hand, we used a simple threshold. In general, shadow regions are not perfectly black because of global illumination. However, our FHFI can effectively eliminate the global components in advance. Hence, pixel intensities in shadow regions become close to zero, and the regions can be segmented stably.

Our factorization process strongly depends on the Lambert model. We assume that intensities observed by multiple cameras become constant if there is no attenuation. Hence, if the diffuse assumption is violated and strong specular reflections are observed, the factorization does not work well.

7. Experiments

7.1 Synthetic Aperture Using Virtual Cameras

First, we evaluated the ability of the synthetic aperture using the prototype system. A textured paper was covered using an obstacle of yellow dense mesh, as shown in **Fig. 12** (a). A white uniform pattern was projected onto the scene.

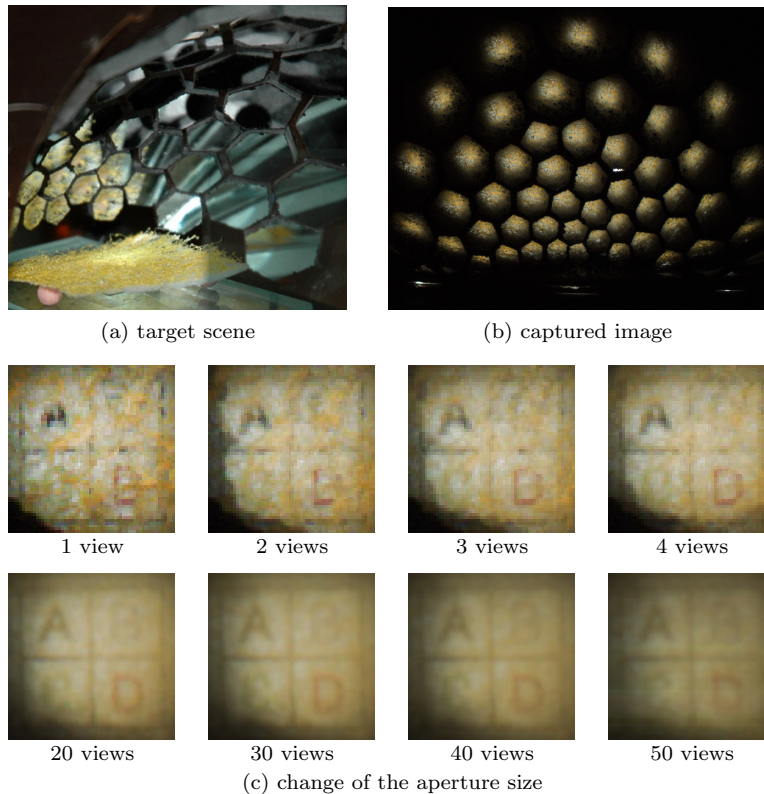


Fig. 12 Results for synthetic aperture using our optical device.

Figure 12 (b) shows the captured image from the real camera. This image includes 50 views corresponding to 50 virtual cameras. Because all the views are affected by the obstacle, it is difficult to see the texture of the sheet. Figure 12 (c) shows the change in appearance when the number of virtual cameras increases to synthesize a large aperture. Because our optical device can synthesize half the hemispherical aperture, the obstacle is completely blurred and the texture becomes clear with an increasing number of virtual cameras.

7.2 Shallow DOF

The turtleback reflector can produce an extremely large aperture using the

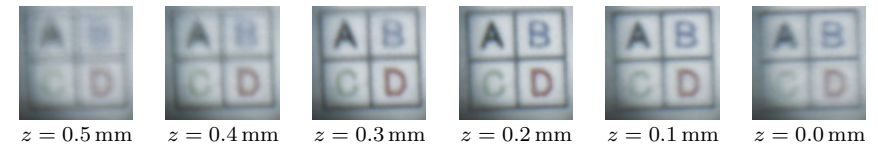


Fig. 13 Sweeping focal plane.

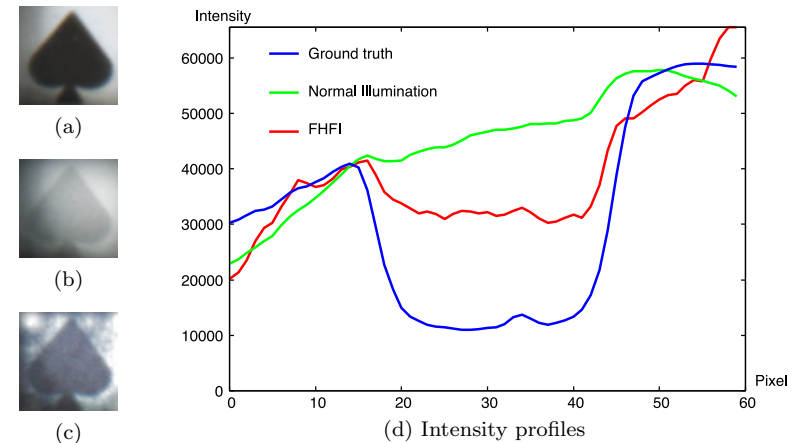


Fig. 14 Descattering by focused high frequency illumination.

synthetic aperture technique. To evaluate the DOF, a textured paper is captured using a hemispherical synthetic aperture. We synthesize an aperture focused on a plane with 0.1 mm increments between $z = 0.0$ mm and $z = 0.5$ mm. As shown in **Fig. 13**, the image at $z = 0.2$ mm is the most focused and the images at $z = 0.0$ mm and $z = 0.4$ mm are blurred explicitly. Therefore, the DOF provided by the turtleback reflector is about 0.4 mm. This DOF is not very shallow because we implemented only half of the turtleback reflector. However, we confirmed the ability to make a shallow DOF.

7.3 Descattering by Focused High Frequency Illumination

Next, we confirmed that the FHFI is effective for descattering in a 3-D volume. The textured paper in **Fig. 14** (a) is covered by a sponge as in (b). Since the sponge scatters lights strongly, edges are blurred.

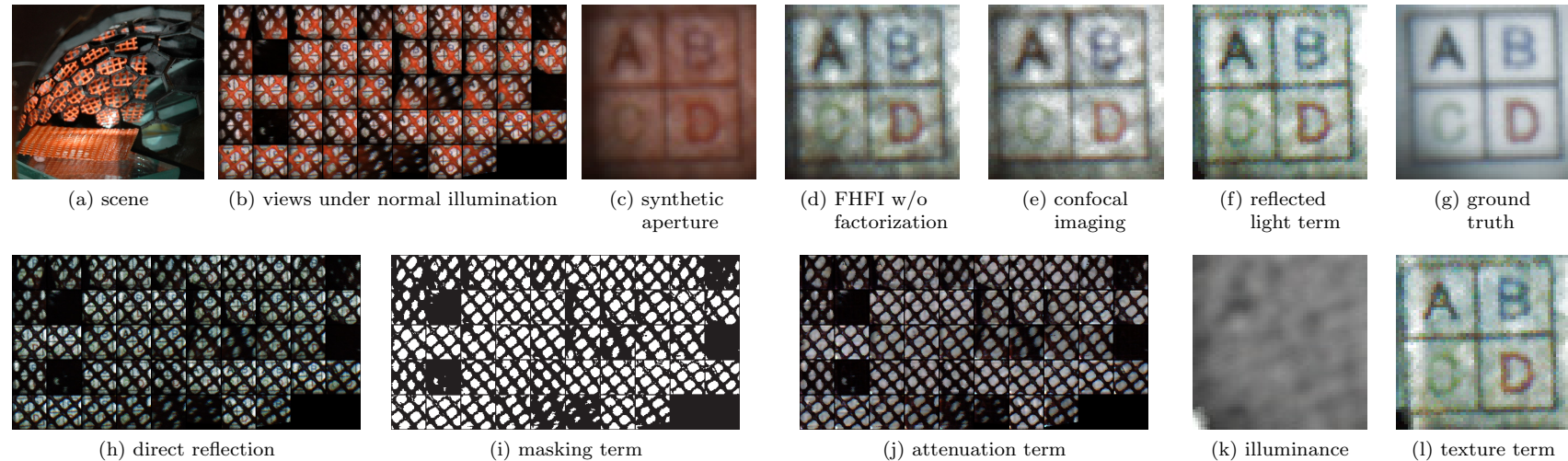


Fig. 15 Result of the combination of the FHFI and the factorization with an occluded scene.

In theory, only two illumination patterns are needed for the FHFI. However, the illuminated pattern becomes blurred at the edges. Hence, we used checkered patterns, in which white and black are replaced every three pixels and shifted the pattern by one pixel. In total, 18 illumination patterns are projected from the virtual projectors so that these patterns are aligned on the paper. As Nayar et al.²⁰⁾ did, we took maximum and minimum intensities for each pixel position in the 18 images. Then, the difference image can be obtained.

Figure 14(c) shows the direct component obtained by the FHFI. We can see that scattering in the 3-D scene is reduced and the contrast is improved. To show how much the contrast was improved, we analyzed the intensity profiles of the images as shown in Fig. 14(d). This graph shows intensity profiles along a horizontal line at the mid height of the images. The central region from 20 to 40 corresponds to the black part of the spade figure. We can see that the contrast was improved by descattering. The descattering effect is not perfect, which is attributed to the low resolution of the virtual projectors in the current prototype system.

7.4 Factorization of the Observed Views

We confirmed the ability to visualize a particular depth in a 3-D scene by combining the FHFI and the factorization. The factorization reduces attenuation effects of occlusion and absorption. Hence, we evaluated them separately.

In the first experiment, we used a scene with occluders. **Figure 15** (a) shows the scene where an orange mesh covers a textured paper, and (b) shows all views from the virtual cameras under normal illumination^{*1}. By simply averaging these views, a synthetic aperture image can be generated, as shown in (c). Although the obstacle is blurred, the orange color of the mesh affects the paper. Confocal imaging¹⁶⁾ can generate a better image as shown in (e) because the orange mesh is less illuminated.

The mesh becomes dark under the FHFI because it is not illuminated, while the paper is bright, as shown in (h). By averaging these views, the dark mesh is blurred and the orange color correctly disappears, as shown in (d). However,

^{*1} Although there are 50 mirror patches, only 48 patches were used because two patches were misaligned.

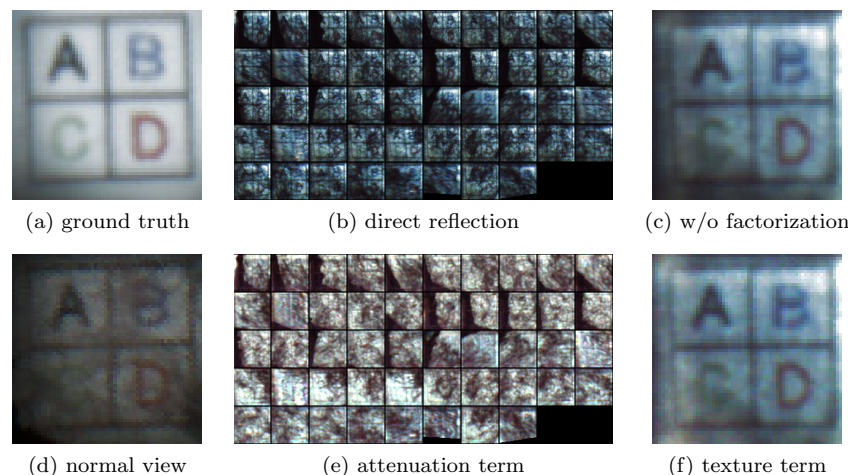


Fig. 16 Result of the combination of the FHFI and the factorization with an absorbed scene.

there are uneven dark regions due to attenuation. The factorization decomposes the observed views (h) into the masking term (i), the attenuation term (j), and the reflected light term (f). We then estimate the illuminance (k) and the texture term (l) using these decomposed terms. We see that the attenuation is reduced, especially around the letters of the black ‘A’ and the red ‘D’, since the occlusion due to the mesh is regarded as masking. On the other hand, comparing (l) with (f), the attenuation is too reduced under the letter ‘B’ because of the inaccurate masking term. Around the edges of obstacles, the masking term cannot be accurately estimated due to the low resolution of the virtual cameras.

While the difference compared with the ground truth image (g) that was captured without the orange mesh is not small, we confirmed that the image without the DOF can be made to almost disappear by hemispherical confocal imaging.

In the second experiment, we used a scene with an inhomogeneous translucent absorber. A target scene consists of a textured paper of **Fig. 16** (a) with cotton waste on the paper. A normal view (d) is unevenly attenuated. In the direct components obtained by the FHFI, attenuation still remains as in (b). Figure 16 (e) shows the attenuation term, and expresses smooth change of absorption and (f) shows the texture term.

While the attenuation and texture terms were factorized, the effectiveness was weak. Figure 16 (c) shows a simple average of (d). The difference between (c) and (f) is small. One of the possible reasons is calibration errors. In fact, the hemispherical synthetic aperture and the FHFI effectively improved the appearance. However, accurate geometric calibration of virtual cameras is required for the factorization. Another reason is that the FHFI did not completely remove scatterings due to the low resolution, while the factorization assumes that images have no scattering.

8. Limitations

- The resolution of the virtual projectors and cameras is low, because the imaging areas of the real projectors and cameras are divided into the virtual areas.
- The observable area is narrow because all of the projectors must illuminate and all of the cameras must observe a common area. To enlarge the area, a large turtleback reflector is necessary, and it may be difficult to construct.
- The factorization is basically an ill-posed problem. For example, we cannot distinguish between two different scenes in which a red texture is covered with a colorless sheet and a white texture is covered with a red sheet. Some empirical constraints such as the smoothness of attenuation are necessary.

9. Conclusions

We propose a new method of hemispherical confocal imaging. This new imaging technique enables us to observe clear views of a particular depth in a 3-D scene. The turtleback reflector designed here can divide the imaging areas so that a projector and a camera mimic a number of virtual projectors and cameras surrounding the scene. The combination of the focused high frequency illumination and the factorization can illuminate only the particular depth of interest, and eliminate scattering and attenuation. We have constructed a prototype system and confirmed the principles of the hemispherical aperture, descattering, and factorization.

In our future work, we aim to build a more accurate optical device using a high resolution projector and evaluate the total performance, because the cur-

rent prototype system only showed the principles separately. Development of applications to visualize cross-sectional views of a translucent object is important, and our future work also aims to visualize the inside of the human body using infrared light.

Acknowledgments This research is granted by the Japan Society for the Promotion of Science (JSPS) through the “Funding Program for Next Generation World-Leading Researchers (NEXT Program),” and by the Microsoft Research CORE5 project.

References

- 1) Adelson, E.H. and Wang, J.Y.A.: Single Lens Stereo with a Plenoptic Camera, *IEEE Tran. PAMI*, pp.99–106 (1992).
- 2) Cossairt, O., Nayar, S.K. and Ramamoorthi, R.: Light Field Transfer: Global Illumination Between Real and Synthetic Objects, *Proc. SIGGRAPH2008*, Article No.57 (2008).
- 3) Dana, K.J. and Wang, J.: Device for convenient measurement of spatially varying bidirectional reflectance, *J. Opt. Soc. Am. A*, Vol.21, No.1, pp.1–12 (2004).
- 4) Debevec, P., Hawkins, T., Tchou, C., Duiker, H.P., Sarokin, W. and Sagar, M.: Acquiring the Reflectance Field of a Human Face, *Proc. SIGGRAPH2000*, pp.145–156 (2000).
- 5) Fuchs, C., Heinz, M., Levoy, M., Seidel, H. and Lensch, H.: Combining Confocal Imaging and Descattering, *Proc. Computer Graphics Forum, Special Issue for the Eurographics Symposium on Rendering*, Vol.27, No.4, pp.1245–1253 (2008).
- 6) Garg, G., Talvala, E.V., Levoy, M. and Lensch, H.P.A.: Symmetric Photography: Exploiting Data-sparseness in Reflectance Fields, *Proc. EGSR2006*, pp.251–262 (2006).
- 7) Ghosh, A., Achutha, S., Heidrich, W. and O’Toole, M.: BRDF Acquisition with Basis Illumination, *Proc. ICCV2007* (2007).
- 8) Gu, J., Nayar, S.K., Grinspun, E., Belhumeur P.N. and Ramamoorthi, R.: Compressive Structured Light for Recovering Inhomogeneous Participating Media, *Proc. ECCV2008*, pp.845–858 (2008).
- 9) Han, J.Y. and Perlin, K.: Measuring Bidirectional Texture Reflectance with a Kaleidoscope, *ACM Trans. Gr.*, Vol.22, No.3, pp.741–748 (2003).
- 10) Kim, J., Lanman, D., Mukaigawa, Y. and Raskar, R.: Descattering Transmission via Angular Filtering, *Proc. ECCV2010*, pp.86–99 (2010).
- 11) Kuthirummal, S. and Nayar, S.K.: Multiview Radial Catadioptric Imaging for Scene Capture, *Proc. SIGGRAPH2006*, pp.916–923 (2006).
- 12) Levoy, M., Chen, B., Vaish, V., Horowitz, M., McDowall, I. and Bolas, M.: Synthetic Aperture Confocal Imaging, *Proc. SIGGRAPH2004*, pp.825–834 (2004).
- 13) Levoy, M. and Hanrahan, P.: Light field rendering, *Proc. SIGGRAPH’96*, pp.31–42 (1996).
- 14) Masselus, V., Peers, P., Dutré, P. and Willems, Y.D.: Relighting with 4D incident light fields, *Proc. SIGGRAPH2003*, pp.613–620 (2003).
- 15) Matusik, W., Pfister, H., Ngan, A., Beardsley, P., Ziegler, R. and McMillan, L.: Image-Based 3D Photography using Opacity Hulls, *Proc. SIGGRAPH2002*, pp.427–437 (2002).
- 16) Minsky, M.: Microscopy apparatus, US Patent 3013467 (1961).
- 17) Mukaigawa, Y., Sumino, K. and Yagi, Y.: Multiplexed Illumination for Measuring BRDF using an Ellipsoidal Mirror and a Projector, *Proc. ACCV2007*, pp.246–257 (2007).
- 18) Müller, G., Bendels, G.H. and Klein, R.: Rapid Synchronous Acquisition of Geometry and Appearance of Cultural Heritage Artefacts, *Proc. VAST2005*, pp.13–20 (2005).
- 19) Narasimhan, S.G., Nayar, S.K., Sun, B. and Koppal, S.J.: Structured Light in Scattering Media, *Proc. ICCV2005*, Vol.1, pp.420–427 (2005).
- 20) 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, *Proc. SIGGRAPH2006*, pp.935–944 (2006).
- 21) Sen, P., Chen, B., Garg, G., Marschner, S., Horowitz, M., Levoy, M. and Lensch, H.: Dual Photography, *Proc. SIGGRAPH2005*, pp.745–755 (2005).
- 22) Treibitz, T. and Schechner, Y.Y.: Active Polarization Descattering, *IEEE Tran. PAMI*, Vol.31, No.3, pp.385–399 (2009).
- 23) Unger, J., Wenger, A., Hawkins, T., Gardner, A. and Debevec, P.: Capturing and Rendering With Incident Light Fields, *Proc. EGRW 2003*, pp.141–149 (2003).
- 24) Vaish, V., Szeliski, R., Zitnick, C.L., Kang, S.B. and Levoy, M.: Reconstructing Occluded Surfaces using Synthetic Apertures: Stereo, Focus and Robust Measures, *CVPR2006*, Vol.II, pp.2331–2338 (2006).
- 25) Vaish, V., Wilburn, B., Joshi, N. and Levoy, M.: Using Plane + Parallax for Calibrate Dense Camera Arrays, *Proc. CVPR 2004*, Vol.I, pp.2–9 (2004).
- 26) Veeraraghavan, A., Raskar, R., Agrawal, A., Mohan, A. and Tumblin, J.: Dappled Photography: Mask Enhanced Cameras for Heterodyned Light Fields and Coded Aperture Refocusing, *Proc. SIGGRAPH2007*, Article No.69 (2007).
- 27) Ward, G.J.: Measuring and Modeling Anisotropic Reflection, *Proc. SIGGRAPH’92*, pp.255–272 (1992).
- 28) Wilburn, B., Vaish, V., Talvala, E., Antunes, E., Barth, A., Adams, A., Horowitz, M. and Levoy, M.: High Performance Imaging Using Large Camera Arrays, *Proc. SIGGRAPH2005*, pp.765–776 (2005).

(Received November 10, 2010)

(Accepted October 7, 2011)

(Released December 28, 2011)

(Communicated by *Takahiro Okabe*)



Seiichi Tagawa received his M.E. degree in Information and Computer Sciences from Osaka University in 2010. He is currently a Ph.D. candidate at Osaka University.



Yasuhiro Mukaigawa received his M.E. and Ph.D. degrees from University of Tsukuba in 1994 and 1997, respectively. He became a Research Associate at Okayama University in 1997, an assistant Professor at University of Tsukuba in 2003, and an Associate Professor at Osaka University in 2004. He joined the MIT Media Lab as a visiting associate professor from 2009 to 2010. His current research interests include computer vision and computational photography. He is a member of IEICE, VRSJ, and IEEE.



Jaewon Kim is currently working at Korea Institute of Science and Technology (KIST) as a research scientist. He received his B.S. and M.S. degrees in mechanical engineering from Korea Advanced Institute of Science and Technology (KAIST) and M.S. degree in media arts and sciences from Massachusetts Institute of Technology (MIT). His research interests are in computational photography, computer vision/graphics and human-computer interaction (HCI).



Ramesh Raskar joined the MIT Media Lab from Mitsubishi Electric Research Laboratories in 2008 as a Head of the Lab's Camera Culture research group. His research interests span the fields of computational light transport, computational photography, inverse problems in imaging and human-computer interaction.



Yasuyuki Matsushita received his B.S., M.S., and Ph.D. degrees in EECS from the University of Tokyo in 1998, 2000, and 2003, respectively. He joined Microsoft Research Asia in April 2003, where he is now a Lead Researcher. His major areas of research are photometric techniques in computer vision and graphics. He is on the editorial board member of IJCV, PAMI, IPSJ CVA, The Visual Computer Journal. He is a senior member of IEEE.



Yasushi Yagi is a Professor at the Institute of Scientific and Industrial Research, Osaka university. He received his Ph.D. degree from Osaka University in 1991. After working at the Product Development Laboratory, Mitsubishi Electric Corporation, he joined Osaka University in 1990. The international conferences for which he served as the program/general chair include: ROBIO2006 (PC), ACCV2007 (PC), ACCV2009 (GC) and ACPR2011 (PC). He was an Editor of IEEE ICRA CEB (2008–2011). He is an associate Editor-in-Chief of IPSJ Transactions on CVA. He has received several awards, including ACM VRST2003 Honorable Mention Award and PSIVT2010 Best Paper Award. He is a fellow of IPSJ and a member of IEICE, RSJ, and IEEE.