Identifying Surface BRDF From a Single 4-D Light Field Image via Deep Neural Network

Feng Lu, Lei He, Shaodi You, Xiaowu Chen, and Zhixiang Hao

***Abstract*—Bidirectional reflectance distribution function (BRDF) defines how light is reflected at a surface patch to produce the surface appearance, and thus, modeling/recognizing BRDFs is of great importance for various tasks in computer vision and graphics. However, such tasks are usually ill-posed or require heavy labor on image capture from different viewing angles. In this paper, we focus on the problem of remote BRDF type identification, by delivering novel techniques that capture and use a single light field image. The key is that a light field image captures both the spatial and angular information by a single shot, and the angular information enables effective samplings of the four-dimensional (4-D) BRDF. To implement the idea, we propose convolutional neural network based architectures for BRDF identification from a single 4-D light field image. Specifically, a StackNet and an Ang-convNet are introduced. The StackNet stacks the angular information of the light field images in an independent dimension, whereas the Ang-convNet uses angular filters to encode the angular information. In addition, we propose a large light field BRDF dataset containing 47 650 high-quality 4-D light field image patches, with different 3-D shapes, BRDFs, and illuminations. Experimental results show significant accuracy improvement in BRDF identification by using the proposed methods.**

***Index Terms*—BRDF, CNN, light field.**

I. Introduction

LIGHT field imaging captures both the spatial and angular information of a scenario with a single snap shot. By properly balancing the spatial aliasing and angular aliasing, a single light field image can be used to enable various tasks, e.g., depth estimation, refocusing, illumination estimation, material estimation, BRDF (bidirectional reflectance distribution function) estimation, which are not easy for traditional methods with a common image. As a result, researches on light field imaging and analysis are able to benefit various applications of computer vision, computer graphics, virtual reality and augmented reality.

In this paper, we focus on the BRDF identification task. BRDF is the function that describes how a surface reflects light. It determines the surface appearance and thus is of great importance for various tasks in computer vision and graphics. However, modeling/identifying BRDF from a single image is not trivial because a BRDF is a four dimensional function that involves different view angles, light source angles, and surface normal. Therefore, measuring a BRDF typically requires up to hundreds of images [1].

Unlike traditional BRDF estimation methods, which are either ill-posed from a single shot image, or require heavy labor on image capture from different viewing angles [1], we propose a novel method for BRDF identification by using a single light field image. The key is that a light field image captures not only the spatial information, but also the angular information through a single shot. The additional angular information in the 4D light field image provides key information to capture the 4D known BRDF.

In order to fully exploit the information in light field image, this paper proposes convolutional neural network (CNN) based architectures for BRDF identification from a single 4D light field image, together with a new light field BRDF dataset. In particular, we propose two network architectures, namely StackNet and Ang-convNet. The StackNet stacks up the angular information of the light field images in an independent dimension, while the Ang-convNet uses angular filters to directly learn the angular information from each angular block. By doing so, the two methods utilize the angular information in different ways.

As for the dataset, it is produced by using different 3D shapes, BRDFs and illuminations. For the BRDF, we use five representative types: plastic, diffuse, narrow-lobe-metal, wide-lobe-metal, and mid-lobe-metal. Each type contains three sub-types which are similar but not identical. Totally 1200 4D light field images are synthesized, producing 47650 valid light field image patches to form a large light field dataset. To our knowledge, this is the first light field BRDF dataset of this size. The proposed CNNs can be directly trained on this dataset, and accomplish the BRDF identification task. Fig. 1 provides an overview of the proposed architecture.

Manuscript received January 5, 2017; revised April 17, 2017; accepted June 19, 2017. Date of publication July 17, 2017; date of current version October 23, 2017. This work was supported in part by the National Natural Science Foundation of China (No. 61602020), the Joint Funds of NSFC-CARFC (No. U1533129), and the NSFC (No. 61532003, 61325011). The guest editor coordinating the review of this paper and approving it for publication was Dr. Lu Fang. *(Corresponding authors: Feng Lu and Xiaowu Chen.)*

F. Lu is with the State Key Laboratory of Virtual Reality Technology and Systems, School of Computer Science and Engineering, Beihang University, Beijing 100191, China, and also with the International Research Institute for Multidisciplinary Science, Beihang University, Beijing 100191, China (e-mail: lufeng@buaa.edu.cn).

X. Chen is with the State Key Laboratory of Virtual Reality Technology and Systems, School of Computer Science and Engineering, Beihang University, Beijing 100191, China (e-mail: chen@buaa.edu.cn).

L. He and Z. Hao are with the School of Computer Science and Engineering, Beihang University, Beijing 100191, China (e-mail: helei01470@buaa.edu.cn; haozx0@buaa.edu.cn.).

S. You is with the Data61-CSIRO, Canberra, ACT 2601, Australia, and also with Australian National University, Canberra, ACT 0200, Australia (e-mail: shaodi.you@anu.edu.au).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTSP.2017.2728001

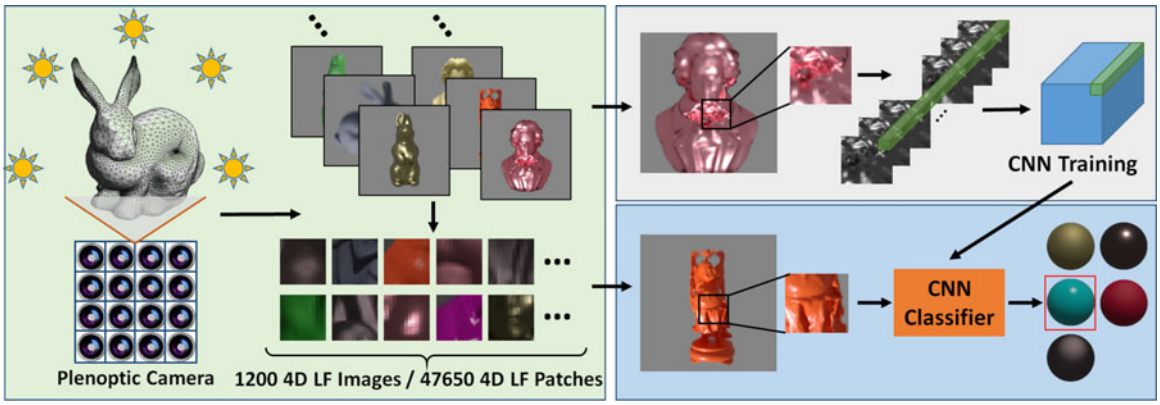


Fig. 1. An overview of the proposed dataset and method for identification of BRDF. Light field images allows simultaneous spatial and angular sampling of an object, which provide rich information for BRDF identification. We collect a dataset containing 47650 4D light field image patches, and train an CNN network with both spatial and angular awareness that automatically identifies the surface BRDF.

The contributions of this paper are summarized as follows. 1) We propose to accomplish the BRDF type identification task by using a single light field image. To the best of our knowledge, this is novel. 2) We propose two CNN based architectures that can effectively utilize light field data for BRDF type identification, which outperform conventional methods by a large margin. 3) We create a large 4D light field dataset that enables efficient training and evaluation.

The rest of the paper is organized as follows: Section II introduces the related works on light field and BRDF. Section III provides the theoretical analysis to show the potential to bridge BRDF identification with light field imaging. Section IV introduces our 4D light field image dataset for BRDF identification followed by the CNN architectures proposed in Section V. Finally, Section VI provides systematical evaluations and Section VII concludes the paper.

II. Related Work

A. Light Field

A light field camera or plenoptic camera allows capturing of spatial and angular information of a scene from a single shot [2]–[7]. This important trait makes light field imaging an interesting topic in computer vision, computer graphics and virtual reality, where the material estimation, illumination estimation, refocusing are essential.

The light field camera was first presented by Lippmann *et al.* [8], and they used lenslets. Ives *et al.*in [9] used pinhole screens. Both of the devices and techniques are based upon the idea that each of the lenslet or pinholes can be considered as a single perspective camera. Because all the lenslets are arranged by varying the spatial positions, they can be viewed as a set of camera array. Stewart *et al.* [10] considered filter to improve sampling aliasing.

A benefit of light field camera is that it avoids angular aliasing as those usually found in camera arrays and multi-view stereo cameras [4]. However, the angular sampling accuracy is achieved through a trade-off with its spatial resolution. Thus, a light field camera usually suffers from spatial aliasing [11], and spatial aliasing is dependent on the camera geometry [3], For a better solution of the trade-off, Levoy and Hanrahan [5] considered the use of large apertures. Bishop *et al.* [11] considered reducing the spatial aliasing by making use of a space-varying filter of the light field in an iterative fashion. Li *et al.* [12] considered deblurring directly. Similarly, Chai *et al.* [13] considered the image-based rendering where a minimum sampling rate for light field rendering is presented. Ng *et al.* [6] exterminated the spatial aliasing and considers them as the artifacts introduced by the reconstruction of the light field. As somewhat related, You *et al.*considered utilizing water-drops as lenslets for light field estimation [14].

Light field enables various applications, other than traditional applications in depth estimation [15], [16], Wang *et al.* [17] considered textured material recognition using a 4D light field dataset along with a CNN architecture. However, their method only works for Lambert material where the angular aliasing barely exists. Besides, Li *et al.* [18] considered using saliency detection.

B. Material/BRDF Recognition

Material recognition is an important task for computer vision. Researches in this direction can be roughly divided into two categories: those based on image appearances (e.g., texture), and those based on reflectance property/BRDF. In the first category, Li *et al.* [19] proposed a data-driven method that uses virtual examples to help material recognition. Liu *et al.* [20] proposed a Bayesian framework using low and mid-level features for material recognition. Recently, Sharan *et al.* [21] exploited the perceptually inspired features for this task while Cimpoi *et al.* [22] used deep filter banks for feature extraction. Wang *et al.* [17] proposed a 4D light field dataset and a CNN architecture to recognize materials using 4D data.

Existing methods in the second category exploit the surface reflectance property, i.e., the BRDF property. For accurate BRDF measurement, several works have been done by using professional capture devices [1], [23], and having exact surface BRDFs, material classification can be done nicely [24]. However, devices for BRDF capture are not widely available for

common users. Dana *et al.* proposed a special mirror used for BRDF/BTF measurement [25], and for convenient measurement of spatially varying bidirectional reflectance [26]. By using similar devices, Han *et al.* [27] proposed to measure bidirectional texture reflectance with a kaleidoscope, and Zhang *et al.* [28] captured reflectance disks and proposed a reflectance hashing method for material recognition. Without requiring additional devices, some other methods need to assume varying illuminations [29]–[34], multiview capturing [35], or known object shapes [36]. Overall, recognize surface BRDFs from a single shot is still challenging.

C. Material/BRDF Recognition Dataset

There are various public datasets produced for material recognition. For instance, the Flickr Material Database (FMD) [37] contains 10 categories with 100 images in each category, the Context Database (MINC) contains 3 million real world image patches classified into 23 materials, and the Describable Textures Dataset (DTD) [38] is another popular one. There are also synthetic datasets [39] compose of computer-generated images, and 4D light-field dataset [17] for material recognition. These datasets all focus on images appearances rather than reflectance property.

There are only a few datasets designed for reflectance/BRDF recognition. Zhang *et al.* [28] proposed a reflectance disk dataset which captures intensities of different viewing angles for 20 materials. However, their dataset lacks spatial information and thus is small.

III. Capturing BRDF in 4D Light Field Image

In this section, we aim to formulate the BRDF identification task under a light field setting. First, the basic concept of BRDF is introduced in Section III-A. Then, the light field imaging setup is provided in Section III-B. Finally, the idea of BRDF identification using a single 4D light field image is analyzed.

A. Surface Appearance and BRDF

BRDFs measure the ratio of the reflected radiance from a surface area. It is a function $f(\boldsymbol{\omega}_{in}, \boldsymbol{\omega}_{out})$ of incoming light direction $\boldsymbol{\omega}_{in}$ and outgoing light direction $\boldsymbol{\omega}_{out}$ in a local coordinate system defined by the surface normal. According to its definition, the pixel value captured at a surface point can be computed by

$$I = f(\boldsymbol{\omega}_{in}, \boldsymbol{\omega}_{out})(\boldsymbol{n}^{\mathrm{T}}\boldsymbol{s}), \tag{1}$$

where $\boldsymbol{n}$ is the unit normal vector of the surface point and $\boldsymbol{s}$ indicates the point light source.

As explained above, the BRDF is a 4-dimensional function which can be highly complex. In most cases, we only consider simple isotropic BRDFs, e.g., the commonly used 100 different isotropic BRDFs in the MERL database [1]. Moreover, those isotropic BRDFs can be categorized into several groups, and the BRDFs inside each group appearing similarly. Accordingly, in this paper we propose five typical BRDF types and use them for BRDF identification.

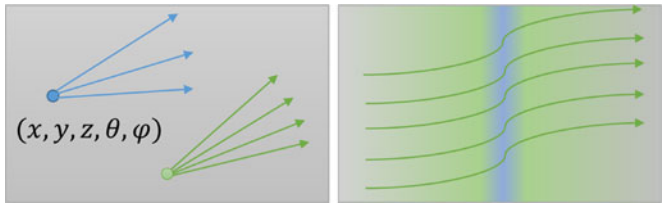


Fig. 2. Light transfer in a media. Left: the power at the given point and given direction can be defined. Right: non-uniform media can cause detours, refraction and reflection of the light.

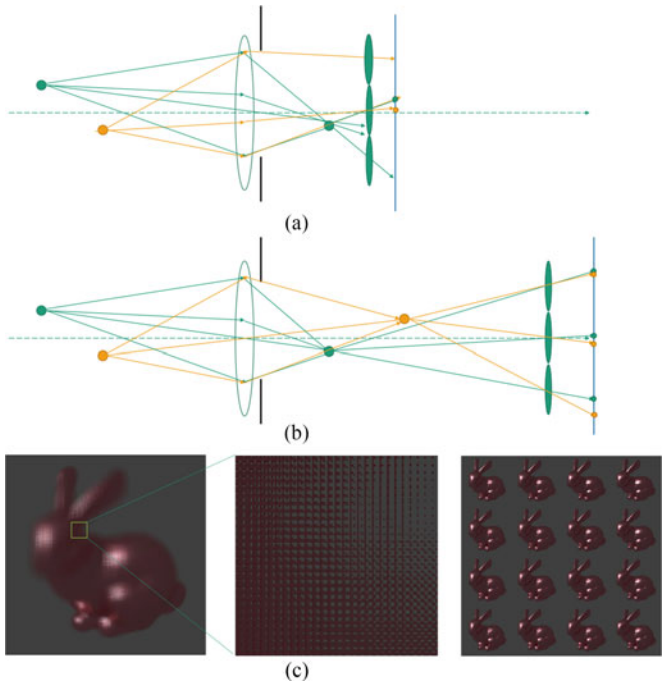


Fig. 3. Light Field Capturing Device (Light Field Cameras) (a) Standard light field camera, the device captures the light field at the 2D place located at the front lens. Light transfers through the front lens, and later the back lens array re-projects the light captured from different directions into the image plane. (b) Equivalence of light field cameras as a set of perspective camera arrays, and one of their captured images. (c) Examples of light field image and multi-view images.

B. Light Field Imaging

We briefly describe the light field imaging technique. This will help understand why light field imaging can benefit the BRDF identification.

Light Field: Light field is a fundamental idea in optics. As illustrated in Fig. 2, light transfers in spaces as a particle. For any media allows a light transfer, it forms a light field. For any given position (x, y, z) in the media, and any given direction (θ, φ), one can observe the field transfer, and in particular, the power (luminescence) of the light transferring at the given point and direction, which is usually defined as $I(x, y, z, \theta, \varphi)$. For a given media with quasi-stable light transferring, the light field is the power of the light transferring among all the positions and directions:

$$I(x, y, z, \theta, \varphi), x, y, z \in \Omega, \tag{2}$$

where ω is the optical media where allows the light transfer.

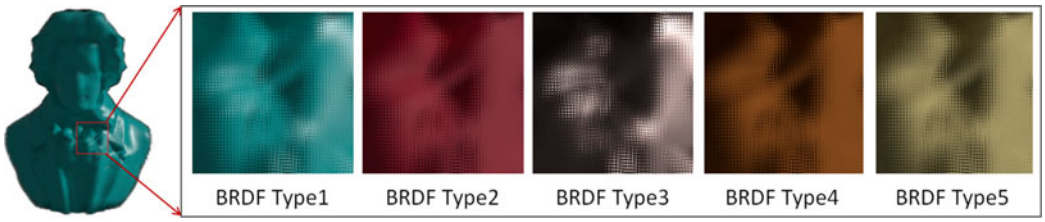


Fig. 4. Different appearances of light field images with different BRDFs. The five BRDF types are: plastic, diffuse, narrow-lobe-metal, wide-lobe-metal, and mid-lobe-metal correspondingly.

Usually, the media is uniform, and light transfers in straight rays. However, in other situations the light transfer undergoes detours, refraction and reflection, as shown in Fig. 2.

Light Field Imaging: As illustrated in Fig. 3, a standard light field camera captures the light field at the front lens, and then re-projects the light from different directions into the image plane via the back lens array. Each lenslet, which can be considered as a perspective camera and captures the light from a certain direction, only covers a portion of all pixels on the sensor due to the limited spatial scale.

As shown in Fig. 3, we can define the light field on the image sensor. As a common practice, we denote the position of lenslets using s, t axis, and the pixel coordinates in each image using u, v axis. Thus we have:

$$I(s,t,u,v), s,t \in \Omega_S, u,v \in \Omega_A, \tag{3}$$

where Ω_S indicates the spatial area, it can be converted between the area on the image sensor and the area on the front lens [40]. Similarly, Ω_A defines the possible angular area, and it is also determined by the distance from the object observed and the front lens aperture.

Assuming the total number of lenslets are S and T, and the number of pixels in each lenslet as U and V. One can formulate the trade-off that $SU = constant$ and $TV = constant$.

Note that the light field imaging is not necessary to be equivalent to the camera array. However, in this paper we use the camera array model to formulate the light field imaging.

C. Light Field of BRDF

Based on the above discussion, we describe the benefit of identifying the BRDFs from light field images.

Referring to (1), the imaging process is determined by the $geometry: \boldsymbol{n}, observingdirection: \boldsymbol{\omega}_{out}$, and the $BRDF: f$ and the $illumination$.

The BRDF identification task can be defined as finding the BRDF which best fits the light field observation.

$$\max_{f \in B} p(I_{LF}|I_f). \tag{4}$$

Where B is the possible BRDF collection, I_f denotes the appearance prior given the BRDF f, and $I_L F$ denotes the observed light field image.

Fig. 4 provides examples of how different BRDFs (plastic, lambertian and metal with different gloss levels) vary the final light field images. Detailed analysis on how BRDF can be effectively captured by the light field image is given below.

Geometry: The light field information enables us to capture the surface geometry from a single image. This is because the light field camera can be considered as perspective camera arrays in some stand [41], [42] to allow shape from light field. Consequently, the surface geometry, especially the surface normal, is an essential factor for measuring BRDFs as described in Section III-A.

Angular Sampling BRDFs are defined as the inbound and outbound light relations, meaning that the angular sampling is crucial. Traditional cameras can only observe one inbound-outbound relation in one image, while a light field image can capture multiple observations in angular domain.

Spatial Resolution: One aspect of light field imaging is that the spatial resolution is not as high as a perceptive camera. However, it is not a substantial problem in BRDF estimation. A light field camera can still have a spatial resolution as high as 1000 by 1000 pixels or more. Which is sufficient for BRDF estimation.

Following the above discussions, in this paper, we propose to tackle the BRDF identification problem by using a light field image. Rather than explicitly measuring the geometric and angular values, we take the advantage of deep convolutional neural networks and learn the most distinguishing features from a single light field image automatically.

IV. Light Field BRDF Dataset

Different from common 2D images, the light field images are neither easy to capture in real world or widely available in Internet. Therefore, for the sake of studying how light field images can benefit the BRDF identification, we provide a novel synthetic light field BRDF dataset in this paper. All images in our dataset are 4D light field images produced by image synthesis. Different combinations of BRDFs, 3D shapes, and illuminations are covered by the dataset.

A. Light Field Image Rendering

We render a large amount 4D light field images by using the following configuration.

BRDFs: We define five typical types of BRDFs, namely, plastic, diffuse, narrow-lobe-metal, wide-lobe-metal, and mid-lobe-metal BRDFs, as seen in the top-left of Fig. 5. Each BRDF type also contains three representative BRDFs, which are similar but not identical. This is to ensure the diversity inside each type. As a result, there are totally fifteen different BRDFs in our dataset.

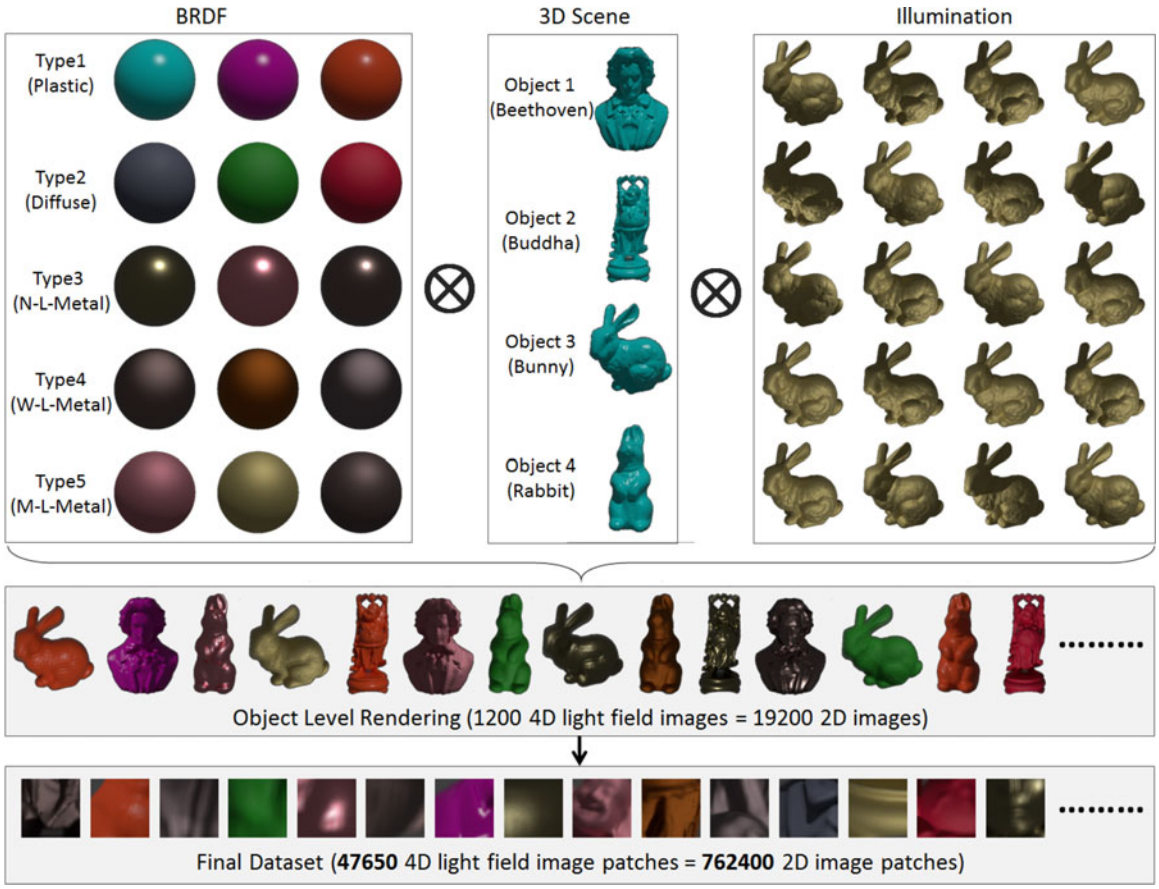


Fig. 5. The proposed dataset for the BRDF identification task. The dataset is generated via image synthesis with five typical types of BRDFs, namely, plastic, diffuse, narrow-lobe-metal, wide-lobe-metal, and mid-lobe-metal BRDFs, four typical 3D objects, namely, Beethoven, Buddha, Bunny and Rabbit, and twenty different illumination conditions. Totally 1200 4D light field images are synthesized which comprise 19200 2D images. Finally, we carefully crop out 47650 4D light field image patches (comprising 762400 2D image patches) from the images to form our dataset.

3D objects: For rendering the images, we use four different and complex 3D objects, including beethoven, happy buddha, bunny, and rabbit. They are easily accessible from the Internet, as visualized in Fig. 5.

Illumination: We design twenty different illumination conditions. Each illumination condition is produced by a distant point light source with a constant intensity and also an ambient light which is uniform from all directions. The directions of the point light sources roughly cover the hemisphere in the camera side. It is important to ensure the diversity of light source directions since different light source directions could produce significantly different surface appearances. The illuminations are also visualized in Fig. 5.

Light Field Imaging Setting: We render each of the 4D light field images by assuming each lenslet covers 4×4 pixels on the image sensor, and thus the angular resolution of the light field image is 4×4. By setting the resolution of any image captured by a lenslet to be 512×512, the final resolution of the 4D light field image is 2048×2048. Note that all the 16 lenslets/cameras have the same distance to the 3D model, and the angular difference of viewing directions of adjacent lenslets/cameras is set to be 2°.

Overall, by considering all the combinations of BRDFs, 3D objects, and illuminations, we render $15 \times 4 \times 20 = 1200$ different 4D light field images, and use them to constitute our dataset. If considering the 2D sub-images captured by different lenslets, our dataset contains $1200 \times 16 = 19200$ 2D images. Some representative images are shown in Fig. 5, whose appearances differ from each other due to different surface BRDFs, 3D shapes, and the illuminations.

B. Final Dataset With Light Field Image Patches

In order to use the rendered light field data in a more efficient way, we extract image patches from the entire images.

First, each 4D light field image is decomposed into 16 2D images. Then, images patches with a size of 64×64 are extracted at the same position of the 16 2D images. Finally, by merging all the 16 2D image patches, the 4D light field image patch with a size of 256×256 is produced.

The positions where the 64×64 image patches are extracted from the 2D image are determined as follows. For each 2D image, we traverse through its different pixel positions by a step of 4 pixels. At each valid pixel position, a crop should be done if more than 95% area of the cropped image patch can be occupied by the object/foreground region. Because difference in BRDF/illumination does not affect the foreground/background distribution in the image, the numbers of cropped image patches must be the same for images rendered for the same 3D object. In particular, Table I shows the numbers of image

TABLE I
NUMBERS OF IMAGE PATCHES THAT CAN BE CROPPED OUT FROM A SINGLE IMAGE OF DIFFERENT 3D OBJECTS

3D object	Beethoven	Buddha	Bunny	Rabbit	Total
# of pathes per image	3062	1458	3366	1644	9530

patches that can be cropped out from one image of different 3D objects.

Theoretically, we can produce $9530 \times 15 \times 20 = 2859000$ image patches as described above. However, notice that there is a change that two patches share the same 3D shape (the same crop position) and the same (or similar) illumination, while their BRDFs are not the same but belonging to the same BRDF type. In such cases, the appearances of the two patches become quite similar. In order to avoid such cases, we propose the following strategy. At each crop position, we randomly select five BRDFs with different BRDF types; at the same time, we also randomly select five illumination conditions. In other words, we only produce five image patches (out of all the 300 possible ones) at each crop position, and their appearances are expected to be quite different. In this way, $9530 \times 5 = 47650$ 4D image patches are obtained, which are ensured to be significantly different from each other.

Each 4D light field image patch is then annotated by its BRDF type label. They constitute our final dataset. To our knowledge, this is the first light field BRDF dataset of this large for the BRDF identification task.

V. CNN ARCHITECTURE FOR BRDF IDENTIFICATION

Following the theoretical analysis in Section III, we propose two CNN architectures for BRDF identification in this section.

A. System Overview

The input to the CNNs is a 4D light field image. In particular, we design two different schemes to process the input in the early stage of the CNN.

StackNet: We convert the light field image into a 3D image volume before sending it to the CNN. In particular, the original light field image is re-organized such that the first two dimensions represent the spatial observation s, t and the angular observation u, v is mapped to the third dimension. Intuitively, the 3D image volume is a stack of 16 gray-scale images of size 64×64 , each of which can be considered as being captured by one of the 16 lenslets, as shown in Fig. 6. Then, the 3D volume is input to a convolutional layer with 64 kernels. Finally, the output is sent to the rest of the network, which is described later.

Ang-convNet: To better utilize the angular information, we propose an Ang-convNet that processes the 4D light field image as a whole. In this method, we introduce 64 so-called angular convolutional kernels (ACKs). Each ACK is a 4 by 4 filter. It differs from common convolutional kernels in that it only filters the angular area in the (gray-scale) light field image. In other words, the stride of convolution operations is 4 in both the horizontal and vertical dimensions. Therefore, the filtering result is a $64 \times 64 \times 64$ data volume, as that in the StackNet. The output is then sent to the rest of the network, as shown in Fig. 6. This method directly learns to handle angular information in the light field image, and thus is expected to have better performance.

As mentioned above, both the StackNet and the Ang-convNet share a large portion of the same network structure, as shown in Fig. 6. In particular, the shared part contains two convolution layers, each of which is followed by a ReLu layer, a normalization layer and a max pooling layer. These operations convert the $64 \times 64 \times 64$ input data into a $16 \times 16 \times 128$ feature, which is then reshaped to form a 1D feature vector. Then, the feature vectors dimension is further reduced to 384×1, 192×1, and 5×1, respectively, after two fully connected layers and a SoftMax classifier. By checking the final output, the BRDF type label can be determined. Note that the stride of all convolution operations in the network is set to 1, and the padding type is zero padding. The image size is changed only during the max pooling operations.

B. Training the CNNs

After finishing the data collection and the CNN structure design, we train the networks by using the ground truth data. In particular, we use the softmax cross-entropy loss function defined as

$$J = -\sum_{i=1}^{n} \sum_{c \in C} t_{ic} \cdot log(y_{ic}), \tag{5}$$

where i indicates different samples and c indicates different BRDF type labels. Note that t_{ic} is 1 only when sample i belongs to BRDF type c, and the softmax output y_{ic} is the probability that sample i belongs to BRDF type c. The loss function computes the difference between the output label and the ground truth, which is iteratively minimized by using the gradient descent algorithm. Besides, we choose to add a dropout layer after each fully connected layer to discard a portion of learnt weights randomly. This helps avoid local optimal, according to Krizhevsky *et al.* [43]. Note that during test, the dropout layers are removed. By setting the learning rate to 0.001, the network learns the relation between the input light field image and its corresponding BRDF type label. The network is then used for the BRDF identification task.

VI. EXPERIMENTS

In this section, we evaluate the proposed CNN-based BRDF identification methods on our light field BRDF dataset. The dataset contains 47650 4D light field image patches, which randomly cover 5 BRDF categories, 9530 different shapes, and 20 illuminations. Our proposed strategy in Section IV-B ensures that any two patches rarely get a chance to share a similar appearance, and thus those 47650 patches are unique.

For experiments, we randomly choose 40000 4D light field image patches as training samples, and the rest 7650 patches as test samples. The training samples are then sent to the StackNet, Ang-convNet and two baseline methods for training, and the test samples are used for evaluation. BRDF type identification accuracy is finally computed from the test output.

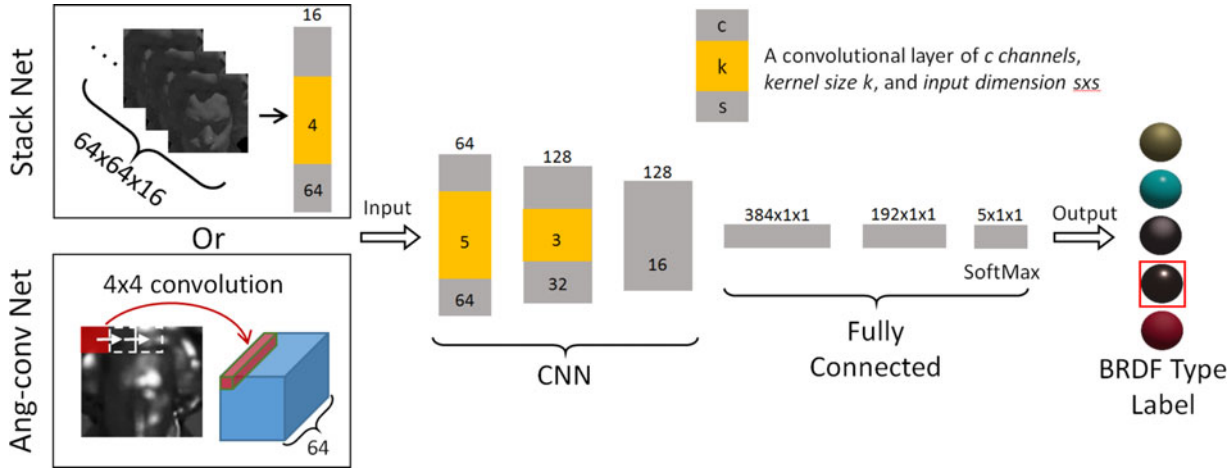


Fig. 6. The proposed CNN structures to utilize 4D light field input for BRDF identification. The StackNet first decomposes one light field image into 16 2D images and then put them into a convolutional layer, while the Ang-convNet uses 4 by 4 angular filters to encode the angular information in the 4D light field image. Then, the output of either the StackNet or the Ang-convNet is passed to the following CNN with 2 typical convolutional layers and 2 typical fully connected layers for BRDF type identification.

TABLE II
ACCURACY W.R.T. BRDF TYPES

	BRDF 1	BRDF 2	BRDF 3	BRDF 4	BRDF 5	Total
Random Forests	56.3%	**75.1%**	90.2%	77.6%	57.9%	71.5%
SingleNet	69.8%	65.1%	82.6%	68.9%	69.8%	61.6%
StackNet	79.9%	74.1%	90.9%	62.0%	70.7%	75.4%
Ang-convNet	**87.6%**	74.7%	**94.9%**	**82.7%**	**78.4%**	**83.7%**

TABLE III
ACCURACY W.R.T. 3D OBJECTS (WHERE THE INPUT PATCH COMES FROM)

	Object 1	Object 2	Object 3	Object 4	Total
Random Forests	68.8%	68.6%	72.5%	77.1%	71.5%
SingleNet	70.7%	70.5%	66.8%	74.1%	69.8%
StackNet	71.6%	76.9%	75.8%	80.8%	75.4%
Ang-convNet	**82.8%**	**84.1%**	**84.4%**	**83.3%**	**83.7%**

Note that our experiments are conducted by using the proposed synthetic dataset. This is because the synthetic dataset can provide different combinations of shapes, BRDFs, and illuminations. Besides, the BRDFs from the MERL dataset were accurately measured from real surfaces, which ensures that the synthetic data can approximate the real world data with high precision. Therefore, our results obtained in this section can also reflect the performance of the proposed methods on the real world data, if a proper light field camera can be used.

A. Baseline Methods

We conduct numerical evaluations for the proposed StackNet and Ang-convNet, and also compare their accuracy with those of two baseline methods. The baseline methods include:

SingleNet: We propose the SingleNet as a baseline method for comparison. The SingleNet differs from the proposed StackNet and Ang-convNet in that it only accepts a single 2D image/image patch as input, rather than a 4D light field image/image patch. Therefore, it can provide evidence of whether light field imaging can benefit the BRDF identification task. In particular, the SingleNet uses the same CNN architecture as shown in Fig. 6 after removing the procedures in the two boxes that handle light field images in the StackNet and Ang-convNet. Besides, the first CNN layer has to be modified to accept a single image as input.

Random Forests (RF): Random forests is an effective ensemble learning method for classification, regression and other tasks.

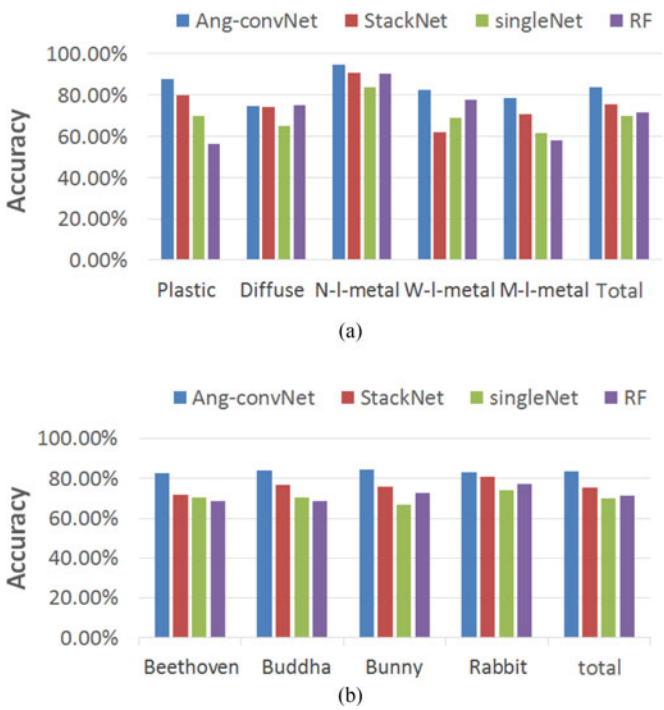
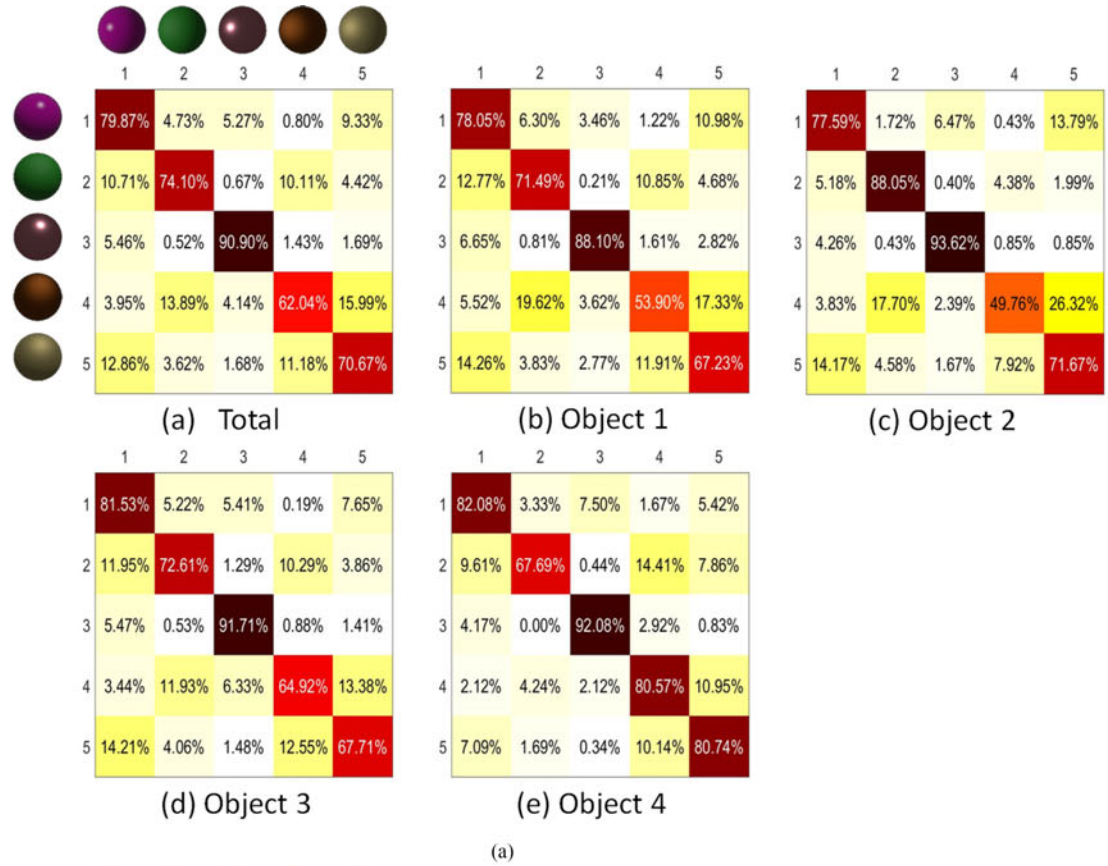
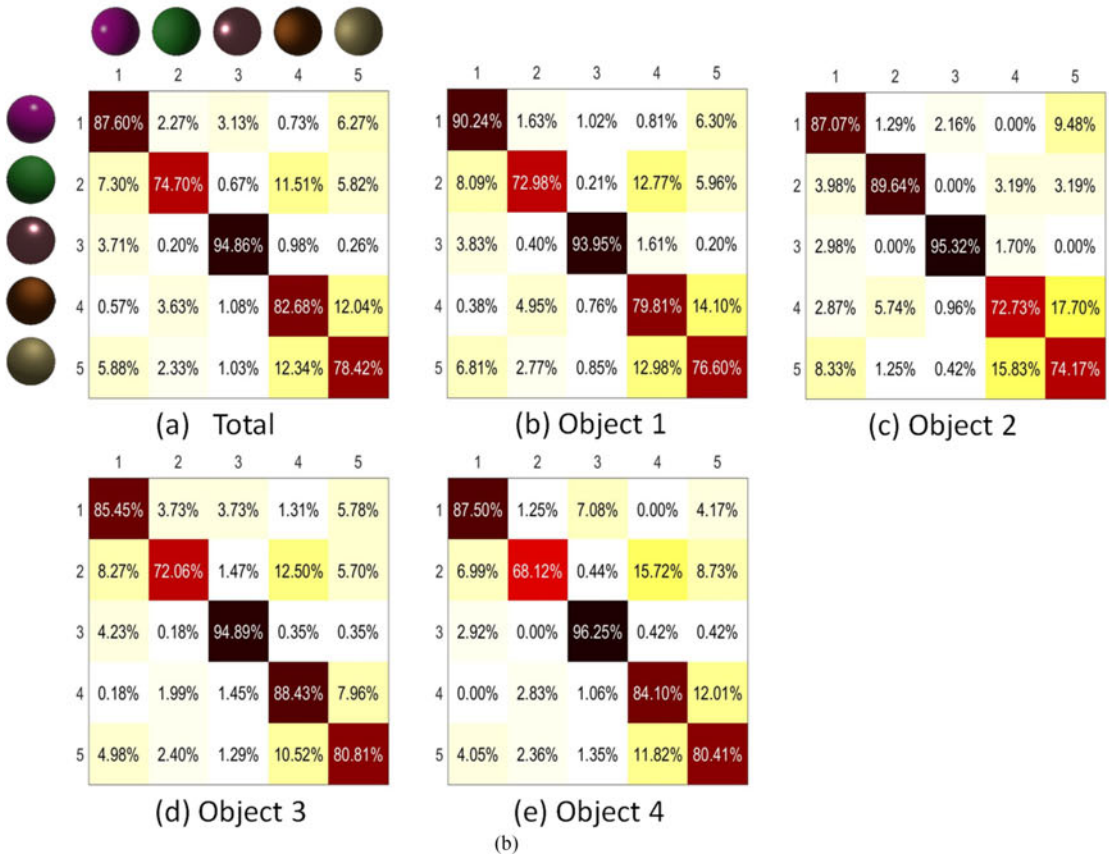


Fig. 7. Comparison of identification accuracy between Ang-ConvNet and StackNet. (a) Accuracy w.r.t. different BRDF types. (b) Accuracy w.r.t. different 3D objects. Five BRDF types include plastic, diffuse, narrow-lobe-metal, wide-lobe-metal, and mid-lobe-metal respectively.



(a)



(b)

Fig. 8. Confusion matrices of BRDF identification results. Top: StackNet. Bottom: Ang-convNet. Five BRDF types are plastic, diffuse, narrow-lobe-metal, wide-lobe-metal, and mid-lobe-metal respectively. (a) Confusion matrix of the StackNet's results. (b) Confusion matrix of the Ang-convNet's results.

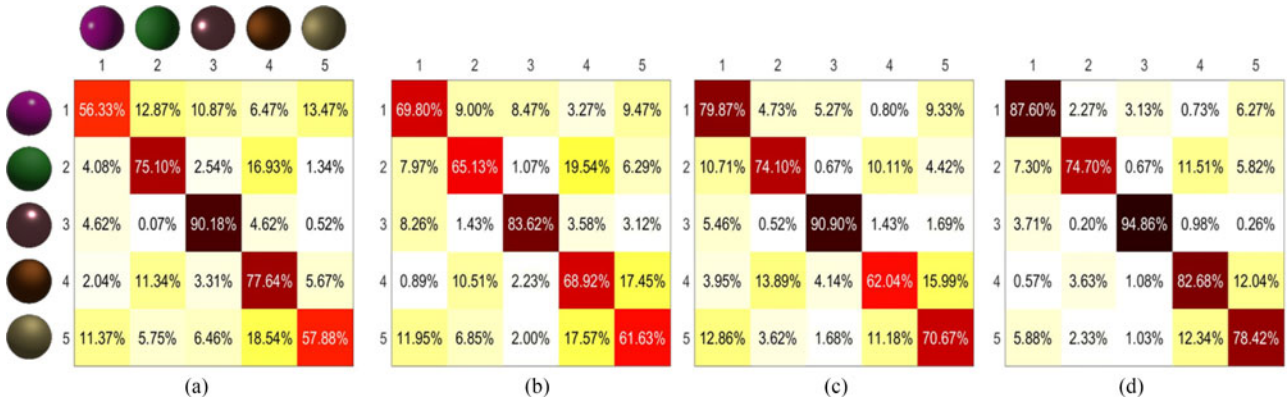


Fig. 9. Comparison between different methods on BRDF type identification accuracy with respect to each BRDF type.

It operates by constructing decision trees at training stage and outputting the class label as the classification result. Random forests can avoid disadvantages of using common decision trees such as the overfitting problem. It is a representative method for solving problems like ours and has been adopted by related researches for thousands of times.

B. Comparison of Different Methods

We conduct experiments on the proposed dataset by using the proposed methods, i.e., StackNet and Ang-convNet, and the two baseline methods, i.e., SingleNet and the Random Forests. The results are shown and compared in Table II. Overally speaking, the StackNet and Ang-convNet significantly outperform the baseline methods and show good performance in the task of BRDF identification (75.4% for StackNet and 83.7% for Ang-convNet). This proves the advantage of using light field information for BRDF identification. While examining their performance for each individual BRDF type, the Ang-convNet outperforms the StackNet especially for BRDF type 4. This reflects the fact that treating the light field image as a whole and applying the proposed angular filtering can help exploit the angular information, which is important for the identification of certain types of BRDF.

We also evaluate the performance of different methods w.r.t. different 3D objects where the test image patches come from. Results are provided in Table III for all of the four methods, from which we can conclude that 1) the proposed methods constantly outperform the baseline methods. 2) different 3D objects do affect the BRDF type identification accuracy for the Random Forests, the SingleNet and the StackNet methods. In particular, Object 4 seems to be the easiest one partially because it has the smoothest surface. 3) On the other hand, the Ang-convNet not only achieves the highest average accuracy, but also performs very stably for all objects. This demonstrates its ability to capture the essential feature for BRDF type identification without being affected by other factors like different shapes.

For better visualization of the results, we plot the accuracy of all the four methods in Fig. 7 and make a comparison. The proposed methods outperform the baseline methods, and the Ang-convNet achieves the highest accuracy and the number is quite satisfactory.

C. Detailed Analysis on BRDF Type Identification Accuracy

This section provides detailed discussions on the BRDF type identification results for Ang-ConvNet and StackNet, as visualized in the confusion matrices in Fig. 8.

StackNet: Several observations can be made from the results of the StackNet. 1) In general, BRDF type 3 achieves the highest identification accuracy of 90.9%, while the number for BRDF type 4 is only 62.0%; 2) More than 15% samples of BRDF type 4 are wrongly classified to be BRDF type 5. This is because BRDF type 4 and 5 has the highest similarity with each other, as shown in Fig. 8; 3) If considering different Objects, it can be seen that Object 4, which has smoothest surface among all, is an "easy one" for identifying BRDF type 4, but a "difficult one" for BRDF 2. This reveals the fact that the performance of the StackNet can be affected by the 3D shapes, in a BRDF type dependent manner.

Ang-convNet: Generally speaking, the Ang-convNet achieves significantly better results than the StackNet in nearly all the cases. However, similar to the StackNet, different objects still affect the BRDF type identification performance. Another observation is that, BRDF type 2, rather than BRDF type 4 in the StackNet case, gets the highest chance among all BRDF types to be mis-identified. In other words, replacing the StackNet by the Ang-convNetBRDF results in the least improvement for BRDF type 2.

Overall, as shown in Fig. 9, the Ang-convNet shows better capability in BRDF type identification. It proves that the angular filtering process applied to the 4D light field images does capture the essential information about the BRDF type.

D. Failure Case Study

In this section, we show some individual cases where the BRDF identification outputs wrongly in our experiment. As shown in Fig. 10, we pick up four individual cases for illustration. In particular, the Ang-convNet is most likely to misidentify the 1) BRDF type 5 to Type 4, 2) BRDF type 4 to Type 5 and 3) BRDF type 2 to type 4, with error rates of 12.34%, 12.04% and 11.51%, respectively. As visualized in Fig. 5, the BRDF type 4 and 5 are hard to distinguish even for human beings, and they also appear similar with BRDF type 2 in many

Fig. 10. Failure cases in BRDF type identification. The first three rows show the failure cases with the highest possibilities, while the bottom row shows the most unlikely case. Note that the wrong outputs are additionally synthesized and they do not exist in the dataset. Also note that when comparing different BRDF types, the difference in color is ignored.

cases. Therefore, the corresponding images shown in Fig. 10 look similar and this explains the high error rates.

On the other hand, the chance that BRDF Type 3 is misidentified to Type 2 is only 0.20%, which is the lowest in our experiments. This is confirmed by the obvious difference in appearance of the two BRDF types. Only several failure cases of this kind can be found as shown in the bottom row of Fig. 10, where the image patch of BRDF type 3 happens to have no specular region and this makes it similar to BRDF type 2. This reveals the difficulty of BRDF type identification from only a single image.

VII. Conclusion

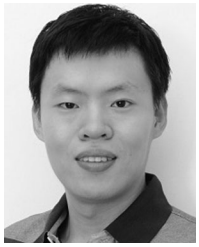
In this paper, we propose novel methods for BRDF identification by using a single light field image. The key idea is that a light field image captures both spatial and angular information through a single shot, and the additional angular information enables effective samplings of the 4D BRDF. Following the idea, a novel light field BRDF dataset and two CNN-based architectures for BRDF identification from a single 4D light field image are provided. Specifically, to cope with the 4D light field image, a StackNet and an Ang-convNet are designed respectively. The StackNet re-organizes the original light field image into a 3D image volume according to its angular observation, and the Ang-convNet uses a 4 by 4 filter to encode the angular information. The dataset is generated by using different 3D models, BRDFs and illuminations. Totally 47650 high quality 4D light field image patched are produced, making the dataset bigger than existing light field datasets. Systematical experiments demonstrate that the proposed methods outperform conventional methods.

Future works include the following. First, besides the synthetic dataset, one can also collect real images by using proper light field cameras for further investigation. Second, while the proposed methods work in a classification manner, they can also be modified for BRDF parameter regression and benefit related applications [44], [45].

References

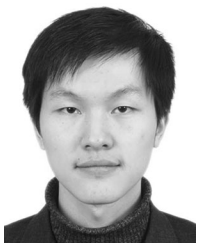
[1] W. Matusik, H. Pfister, M. Brand, and L. McMillan, "A data-driven reflectance model," *ACM Trans. Graph.*, vol. 22, no. 3, pp. 759–769, 2003.
[2] G. Wu *et al.*, "Light field image processing: An overview," *J. Sel. Topics Signal Process.*, vol. 11, no. 7, pp. 926–954, Oct. 2017.
[3] T. Georgiev and A. Lumsdaine, "Depth of field in plenoptic cameras," in *Proc. Eurographics*, 2009, pp. 5–8.
[4] E. H. Adelson and J. Wang, "Single lens stereo with a plenoptic camera," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 2, pp. 99–106, Feb. 1992.
[5] M. Levoy and N. Hanrahan, "Light field rendering," in *Proc. SIGGRAPH*, 1996, pp. 31–42.
[6] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," *Comput. Sci. Tech. Rep.*, vol. 2, no. 11, pp. 1–11, 2005.
[7] T. Georgiev, Z. Yu, A. Lumsdaine, and S. Goma, "Lytro camera technology: Theory, algorithms, performance analysis," *Proc. SPIE*, vol. 8667, 2013, Art. no. 86671J.
[8] G. Lippmann, "Épreuves réversibles donnant la sensation du relief," *J. Phys. Theory Appl.*, vol. 7, no. 1, pp. 821–825, 1908.
[9] H. E. Ives, "A camera for making parallax panoramagrams," *J. Opt. Soc. Amer.*, vol. 17, no. 6, pp. 435–439, 1928.
[10] J. Stewart, J. Yu, S. J. Gortler, and L. McMillan, "A new reconstruction filter for undersampled light fields," in *Proc. 14th Eurograph. Workshop Rendering*, 2003, pp. 150–156.
[11] T. E. Bishop and P. Favaro, "Full-resolution depth map estimation from an aliased plenoptic light field," in *Proc. 10th Asian Conf. Comput. Vis.*,, 2011, pp. 186–200.
[12] F. Li, J. Yu, and J. Chai, "A hybrid camera for motion deblurring and depth map super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2008, pp. 1–8.
[13] J. X. Chai, X. Tong, S. C. Chan, and H. Y. Shum, "Plenoptic sampling," in *Proc. SIGGRAPH*, 2000, pp. 307–318.
[14] S. You, R. T. Tan, R. Kawakami, Y. Mukaigawa, and K. Ikeuchi, "Waterdrop stereo," 2016, pp. 1–12, arXiv:1604.00730.
[15] E. Ekmekcioglu, V. Velisavljevic, and S. T. Worrall, "Content adaptive enhancement of multi-view depth maps for free viewpoint video," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 2, pp. 352–361, Apr. 2011.
[16] K. Li, Q. Dai, W. Xu, and J. Yang, "Temporal-dense dynamic 3-d reconstruction with low frame rate cameras," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 5, pp. 447–459, Sep. 2012.
[17] T.-C. Wang, J.-Y. Zhu, E. Hiroaki, M. Chandraker, A. A. Efros, and R. Ramamoorthi, "A 4D light-field dataset and CNN architectures for material recognition," in *Proc. Eur. Conf. Comput. Vis.*, New York, NY, USA: Springer, 2016, pp. 121–138.
[18] N. Li, J. Ye, Y. Ji, H. Ling, and J. Yu, "Saliency detection on light field," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 2806–2813.
[19] W. Li and M. Fritz, "Recognizing materials from virtual examples," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 345–358.
[20] C. Liu, L. Sharan, E. H. Adelson, and R. Rosenholtz, "Exploring features in a Bayesian framework for material recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2010, pp. 239–246.

[21] L. Sharan, C. Liu, R. Rosenholtz, and E. Adelson, "Recognizing materials using perceptually inspired features," *Int. J. Comput. Vis.*, vol. 103, no. 3, pp. 348–371, 2013.
[22] M. Cimpoi, S. Maji, and A. Vedaldi, "Deep filter banks for texture recognition and segmentation," in *Proc. 2015 IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3828–3836.
[23] M. K. Johnson and E. H. Adelson, "Retrographic sensing for the measurement of surface texture and shape," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 1070–1077.
[24] O. Wang, P. Gunawardane, S. Scher, and J. Davis, "Material classification using BRDF slices," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 2805–2811.
[25] K. J. Dana, "BRDF/BTF measurement device," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2001, vol. 2, pp. 460–466.
[26] K. Dana and J. Wang, "Device for convenient measurement of spatially varying bidirectional reflectance," *J. Opt. Soc. Amer. A*, vol. 21, pp. 1–12, 2004.
[27] J. Y. Han and K. Perlin, "Measuring bidirectional texture reflectance with a kaleidoscope," *ACM Trans. Graph.*, vol. 22, no. 3, pp. 741–748, 2003.
[28] H. Zhang, K. Dana, and K. Nishino, "Reflectance hashing for material recognition," in *Proc. 2015 IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3071–3080.
[29] C. Liu and J. Gu, "Discriminative illumination: Per-pixel classification of raw materials based on optimal projections of spectral BRDF," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 1, pp. 86–98, Jan. 2014.
[30] A. Hertzmann and S. Seitz, "Example-based photometric stereo: Shape reconstruction with general, varying BRDFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1254–1264, Aug. 2005.
[31] F. Lu, I. Sato, and Y. Sato, "Uncalibrated photometric stereo based on elevation angle recovery from BRDF symmetry of isotropic materials," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2015, pp. 168–176.
[32] F. Lu, Y. Matsushita, I. Sato, T. Okabe, and Y. Sato, "From intensity profile to surface normal: Photometric stereo for unknown light sources and isotropic reflectances," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 10, pp. 1999–2012, Oct. 2015.
[33] S. Li and B. Shi, "Photometric stereo for general isotropic reflectances by spherical linear interpolation," *Opt. Eng.*, vol. 54, no. 8, pp. 83–104, 2015.
[34] F. Lu, X. Chen, I. Sato, and Y. Sato, "Symps: BRDF symmetry guided photometric stereo for shape and light source estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PP, no. 99, pp. 1–14, 2017. doi: 10.1109/TPAMI.2017.2655525.
[35] G. Oxholm and K. Nishino, "Multiview shape and reflectance from natural illumination," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 2163–2170.
[36] S. Lombardi and K. Nishino, "Single image multimaterial estimation," in *Proc. 2012 IEEE Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 238–245.
[37] L. Sharan, R. Rosenholtz, and E. Adelson, "Material perception: What can you see in a brief glance?" *J. Vis.*, vol. 9, no. 8, pp. 784–784, 2009.
[38] M. Cimpoi, S. Maji, I. Kokkinos, S. Mohamed, and A. Vedaldi, "Describing textures in the wild," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 3606–3613.
[39] M. Weinmann, J. Gall, and R. Klein, "Material classification based on training data synthesized using a BTF database," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 156–171.
[40] Y. Liu, Q. Dai, and W. Xu, "A real time interactive dynamic light field transmission system," in *Proc. IEEE Int. Conf. Multimedia Expo.*, 2006, pp. 2173–2176.
[41] C. Wu, Y. Liu, Q. Dai, and B. Wilburn, "Fusing multiview and photometric stereo for 3d reconstruction under uncalibrated illumination," *IEEE Trans. Vis. Comput. Graph.*, vol. 17, no. 8, pp. 1082–1095, Aug. 2011.
[42] C. Wu, K. Varanasi, Y. Liu, H.-P. Seidel, and C. Theobalt, "Shading-based dynamic shape refinement from multi-view video under general illumination," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2011, pp. 1108–1115.
[43] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst. 25*, 2012, pp. 1097–1105.
[44] B. Shi, Z. Wu, Z. Mo, D. Duan, S.-K. Yeung, and P. Tan, "A benchmark dataset and evaluation for non-Lambertian and uncalibrated photometric stereo," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 3707–3716.
[45] X. Chen, H. Wu, X. Jin, and Q. Zhao, "Face illumination manipulation using a single reference image by adaptive layer decomposition," *IEEE Trans. Image Process.*, vol. 22, no. 11, pp. 4249–4259, Nov. 2013.



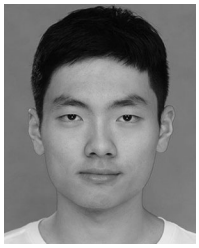
Feng Lu received the B.S. and M.S. degrees in automation from Tsinghua University, Beijing, China, in 2007 and 2010, respectively, and the Ph.D. degree in information science and technology from the University of Tokyo, Tokyo, Japan, in 2013. He is currently a Professor in the State Key Laboratory of Virtual Reality Technology and Systems, School of Computer Science and Engineering, Beihang University, Beijing, China. His research interests include 3-D shape recovery, reflectance analysis, and human gaze analysis.

Lei He received the B.S. degree in computer science from Jilin University, Changchun, China, in 2012. He is currently working toward the M.S. degree in computer science at Beihang University, Beijing, China. His research interests include photometric methods in computer vision and 3-D modeling.



Shaodi You received the Ph.D. and M.E. degrees from the University of Tokyo, Tokyo, Japan, in 2015 and 2012, respectively, and the Bachelor's degree from Tsinghua University, Beijing, China, in 2009. He is currently a Research Scientist in Data61, (Previously known as NICTA) Canberra Research Lab and an Adjunct Lecturer in the College of Engineering and Computer Science, Australian National University, Canberra, Australia. His research interests include physics based vision, perception based vision and learning, and 3-D geometry.

Xiaowu Chen received the Ph.D. degree in computer science from Beihang University, Beijing, China, in 2001. He is currently a Professor in the State Key Laboratory of Virtual Reality Technology and Systems, School of Computer Science and Engineering, Beihang University, Beijing, China. His current research interests include virtual reality, computer graphics, and computer vision.



Zhixiang Hao received the B.S. degree in computer science from Xidian University, Xi'an, China, in 2016. He is currently working toward the Master's degree in the School of Computer Science and Engineering, Beihang University, Beijing, China. His research interests include computer vision and computer graphics.