IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 29, 2019, accepted February 14, 2019, date of publication February 20, 2019, date of current version March 7, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2900153

Computational Light Field Generation Using Deep Nonparametric Bayesian Learning

NAN MENG[1], XING SUN[2], HAYDEN K.-H. SO[1], (Senior Member, IEEE), AND EDMUND Y. LAM[1], (Fellow, IEEE)

[1]Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong
[2]Tencent, Shanghai 200030, China

Corresponding author: Edmund Y. Lam (elam@eee.hku.hk)

This work was supported in part by the University Research Committee of The University of Hong Kong under Project 104005009.

ABSTRACT In this paper, we present a deep nonparametric Bayesian method to synthesize a light field from a single image. Conventionally, light-field capture requires special optical architecture, and the gain in angular resolution often comes at the expense of a reduction in spatial resolution. Techniques for computationally generating the light field from a single image can be expanded further to a variety of applications, ranging from microscopy and materials analysis to vision-based robotic control and autonomous vehicles. We treat the light field as multiple sub-aperture views, and to compute the novel viewpoints, our model contains three major components. First, a convolutional neural network is used for predicting the depth probability map from the image. Second, a multi-scale feature dictionary is constructed within a multi-layer dictionary learning network. Third, the novel views are synthesized taking into account both the probabilistic depth map and the multi-scale feature dictionary. The experiments show that our method outperforms several state-of-the-art novel view synthesis methods in delivering good image resolution.

INDEX TERMS Image reconstruction, convolutional neural network, deep learning, nonparametric Bayesian, light field imaging.

I. INTRODUCTION

The emergence of light field cameras [1] has enabled photographers to refocus an image after it is taken, which is an appealing feature and has led to new artistic creations. Light field imaging also has found a variety of applications, such as microscopy [2], materials recognition [3], vision-based robotic control [4], and autonomous navigation [5]. The light field being acquired is often a set of sub-aperture images from different perspectives in a single shot [6], although there are also alternatives to build up the light field from a stack of images [7], [8]. A light field can be directly captured with lenslet arrays [1] or masks [9]–[11], which unfortunately is either expensive or difficult to manufacture. More problematic is the trade-off between spatial and angular resolution, where often only a low-resolution 2-dimensional (2D) image is reconstructed when a specific focus point is chosen. To overcome this, we aim to create a light field based on a traditional capture of a 2D image, which has a much higher spatial resolution than an image from a light field camera.

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Chen.

We approach the light field synthesis problem by developing a Deep Nonparametric Bayesian (DNB) model, which performs deep analysis and learning from a single image, and then reconstruct the novel views with a learning dictionary. Regarding the first step, there has been work on single image analysis with deep learning, such as with the Convolutional Factor Analysis (CFA) model [12], which analyzes images using a hierarchical convolutional factor-analysis construction under a nonparametric Bayesian setting. For the second step, many algorithms have also been proposed for view synthesis, especially binocular stereo image synthesis, such as the automatic 2D-to-3D video conversion with deep learning (Deep3D) [13]. Meanwhile, it has also been demonstrated that one can synthesize the light field from RGB-D (for red, green, blue, plus depth) images with depth propagation [14]. Our approach, however, has two advantages. First, the proposed algorithm does not require depth information, and is calculated based on the RGB images rather than RGB-D images, leading to the algorithm performing more robust than depth-based methods [15], [16]. Second, we construct the novel views with a powerful scheme based on a multi-scale feature dictionary, delivering a better light field synthesis

result especially when the depth information is not very accurate. While there are a number of methods to predict the depth map from a single image, the results are not guaranteed to be very reliable. As such, we choose to infer the depth probability map, which has better tolerance for inaccurate depth prediction.

In addition, a major contribution of our DNB model is to process the image using deep dictionary learning (DDL) to infer a multi-scale dictionary, and then synthesize the light field through a sigmoid dictionary representation (SDR). In this way, we can construct the feature dictionary from different resolutions, using the technique of Bayesian pooling. While "max-pooling" in CFA [12] focuses on reducing the feature number in the dictionary, in comparison, the proposed DDL not only reduces the feature number but also down-samples the feature dimension in the dictionary. Furthermore, the SDR is powerful in that it inpaints the reconstructed novel view in multiple scales instead of simply applying bilinear interpolation [13], [17], by using joint probability mixture model instead of direct depth propagation.

This paper is organized as follows. Section II gives a brief introduction to the light field. Then, the Sparse Hierarchical Nonparametric Bayesian (SHNB) model, which is the fundamental theory for dictionary learning and representation, is introduced in Section III. This is followed by a detailed discussion of the DNB model architecture in Section IV, covering the convolutional neural network (CNN), and the view synthesis using DDL and SDR. Section V provides the detailed inference algorithm, while Section VI presents several experiments to demonstrate the efficacy of our approach.

II. LIGHT FIELD BACKGROUND

In this section, the background of the light field is introduced. Based on light field representation theory, we further propose feature topic variation among different perspectives. This paves the way for the development of the light field sparse dictionary learning algorithm.

A. LIGHT FIELD REPRESENTATION THEORY

Radiance of light rays from every 3D point in an object is commonly represented by a 5D plenoptic function, $F(x, y, z, \theta, \phi)$, parameterized by three spatial coordinates (x, y, z) and two angles (θ and ϕ) [18], [19]. When the radiance along the ray in free space is assumed to be a constant, it reduces to 4D [20]. Furthermore, the light rays can be expressed with spatial coordinates (x, y) and angular coordinates (u, v) with a two-plane parameterization [21]

$$L(x, y, u, v). \tag{1}$$

The collection of such functions is the light field. Sub-aperture view is a powerful representation of the light field for many image processing tasks [1], [6], where each is a single viewpoint corresponding to a particular set of (u, v).

It is straightforward to relate the light field to an image, labeled $\boldsymbol{\phi}$, captured by the electronic detector. It involves a linear integration process [22] formulated as,

$$\boldsymbol{\phi} = \mathbf{T} \cdot L + n, \tag{2}$$

where n is the noise, and $\mathbf{T}(\cdot)$ is the projection function. In a conventional camera, $\mathbf{T}(\cdot)$ sums up the corresponding pixels with different angles to form a 2D image $\boldsymbol{\phi}$. However, in a plenoptic camera, such as one using a lenslet array in the original detector plane, $\mathbf{T}(\cdot)$ is an identity [23]. We can take $\boldsymbol{\phi}$ to be a 2D array with size $U \times V$ of the various viewpoints, and each element is a 2D sub-aperture view image $\boldsymbol{\phi}_{u,v}(x, y)$ for fixed (u, v), where $u = 1, \ldots, U$ and $v = 1, \ldots, V$. Generally speaking, we can build such a relationship between a typical 2D image and a sub-aperture view in the light field data.

B. FEATURE TOPICS VARIATION

Using nonparametric Bayesian methods, features of a 2D image follow certain probability distributions, which are in turn governed by the parameters called "topics" [24]. It has been shown that an analysis of objects as topics varying along different frames in video clips can provide an effective way to detect and track them [25], and letting the topics vary spatially can achieve good results in hyperspectral image identification for geoscience and remote sensing [24]. Here, the concept of topic variation is introduced to light field sub-aperture view image modeling, where we can analyze it from different perspectives.

To do so, we need to define and relate the topics in different viewpoints. For simplicity in notation, we will only keep x and u and will drop y and v for now. When we observe a topic or a pixel at the original viewpoint u_0, whose spatial location is denoted as x, the shifted spatial location $\tau_1(x)$ at a novel viewpoint u_1 can be expressed as [26]

$$\tau_1(x) = x + d_0(x) \cdot (u_1 - u_0), \tag{3}$$

where $d_0(x)$ denotes the depth of this pixel at perspective u_0. Substituting this into Eq. 2, the expression of topics variation along different viewpoints can be derived as

$$\boldsymbol{\phi}_m\big(\tau_m(x)\big) = \mathbf{T} \cdot L\big(x + d_0(x) \cdot (u_m - u_0), u_m\big) + n, \tag{4}$$

where $\boldsymbol{\phi}_m$ denotes the captured image at perspective u_m. We also use M to denote the total number of viewpoints along coordinate u.

As inferred from Eq. 4, compared with frames from a video, sub-aperture view images of a light field have much stronger dependencies on each other. It is natural that we can learn a sparser feature dictionary from the light field based on the feature topics variation, especially when the depth map of the light field is known.

III. DICTIONARY LEARNING FOR LIGHT FIELD REPRESENTATION

The proposed view synthesis algorithm is based on the nonparametric Bayesian models [27]. This is different from deterministic synthesis algorithms (see [14]), which require

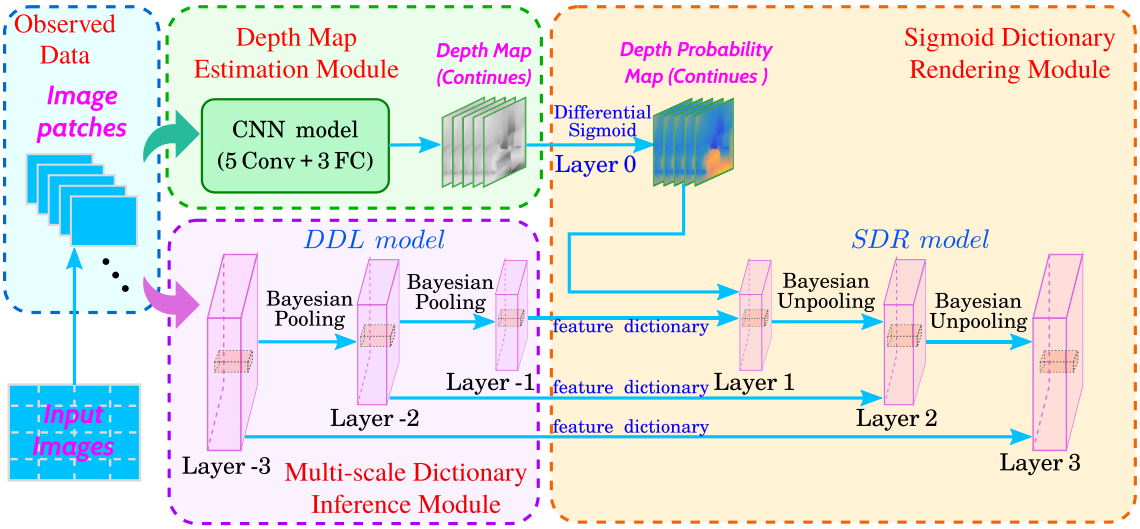


FIGURE 1. Deep Nonparametric Bayesian (DBN) architecture with Depth Map Estimation Module (CNN model), Multi-scale Dictionary Inference Module (DDL model), and SDR Module.

a very accurate depth map. Making use of the light field topics variation in Eq. 4, we can now represent various sub-apertures in a compact way through learning a sparse feature dictionary. Furthermore, the dictionary can be used for light field reconstruction, denoising, and inpainting. In the following paragraphs, we introduce the procedures for the sparse hierarchical dictionary learning based on a nonparametric Bayesian model.

The procedures consist of four steps, namely,

1) **Observed data modeling:** The observed data denote the image patches cropped in every sub-aperture views. Each patch can be considered as a mixture of the features selected by the corresponding sparse dictionary indicator.
2) **Global dictionary updating:** As shown in Eq. 4, image patches cropped from different sub-aperture views share substantially similar features and these features are further exploited to form a global dictionary. All the image patches are assumed to follow a Gaussian distribution with zero mean and a Gamma prior variance matrix.
3) **Local dictionary inference:** Based on the global feature dictionary, we can further define the local feature dictionaries used by the image patches cropped from corresponding sub-aperture image. Every local dictionary is derived from the global one with a certain weight specific for each image patch. Similar to the above, the weight is also assumed to follow a Gaussian distribution with zero mean and a Gamma prior variance matrix with the matrix size corresponds to the number of global feature dictionary.
4) **Local topic indicator sampling:** The binary dictionary indicator selects suitable local feature dictionaries for each image patch. To make it sparse, we introduce the Dirichlet-Gamma prior which is demonstrated to be able to control both sparsity and smoothness [28]. Therefore, we assume that the dictionary indicator follows a Dirichlet distribution, denoted by $\mathrm{Dir}(b_0)$, and the hyper-parameter b_0 follows a Gamma prior.

IV. DNB VIEW SYNTHESIS ARCHITECTURE

Fig. 1 illustrates the overall DNB model architecture, with learning and view synthesis steps denoted as *depth map estimation module*, *multi-scale dictionary inference module*, and *sigmoid dictionary rendering (SDR) module*.

Specifically, when the parameters of the CNN model are fine-tuned, both *depth map estimation module* and *SDR module* together construct a weakly supervised deep network. The outputs of CNN is subsequently fed into a differential sigmoid layer (Layer 0) to predict the depth probability map. Such operation can be described as

$$S'(d) = S(d)\big(1 - S(d)\big) = \frac{\mathrm{e}^{-d}}{(1+\mathrm{e}^{-d})^2}, \tag{5}$$

where d denotes the estimated depth obtained from the CNN model, $S(\cdot)$ is the sigmoid function, and $S'(\cdot)$ is the differential sigmoid function. The differential sigmoid layer converts the continuous depth map into a depth probability map fed into the subsequent SDR model.

To demonstrate clearly where the depth information is incorporated into the DBN architecture, we index the differential sigmoid layer used for generating depth probability map as *Layer 0*. As a consequence, the DDL layers then have negative indices. The major component of our framework is a deep (multi-layer) network constructed by combining the *multi-scale dictionary inference module* and the *SDR module*, which synthesizes the depth probability information (from CNN model) and the feature dictionaries. According to Eq. 4, we can further generate novel views based on the corresponding pixel shifts.

Our proposed DNB model has several advantages:

1) We propose to utilize the depth probability map for light field synthesis. In addition, the acquisition of such depth probability map is straightforward, only by applying a differential sigmoid operation on the estimated depth map.
2) We incorporate the dark channel prior with Bayesian pooling to generate the downsampled global feature dictionary, which encourages the dictionary to learn better depth information.
3) The proposed multi-scale dictionary can be further utilized to refine and inpaint the novel views.

In following sections, we describe each component of our proposed DNB architecture in details and all the advantages listed above are elaborated accordingly.

A. DEPTH MAP ESTIMATION MODULE

CNN has been applied on a variety of recognition and classification tasks [29], [30]. It is a multilayer neural network composed of one or more convolution layers, often together with sub-sampling, followed by one or more fully connected layers. It has the advantage of being shift, scale, and distortion invariant [31]. These properties are useful for object recognition problems, which often require identifying different objects with various positions.

Depth estimation based on CNN has been used for multi-media data, such as stereo images [17], and in digital holography autofocusing [32] and reconstruction [33]. However, doing so from a single image is much more difficult. With no reliable depth cues and the challenge of finding good representations of depth, CNN would have difficulty classifying and computing an accurate depth map. Some previous attempts only estimate a coarse one with heavy post-processing, including denoising, inpainting, super-resolution and depth propagation [34], [35].

TABLE 1. Parameters of the CNN model.

layers	kernel size	# of filters	pooling size
conv 1	11×11	64	2×2
conv 2	5×5	256	2×2
conv 3	3×3	256	—
conv 4	3×3	256	—
conv 5	3×3	256	2×2

The CNN model in our *depth map estimation module* is established based on the framework [36], consisting of 5 convolutional layers and 3 fully connected layers (Table 1 presents the key settings). The network is pretrained using NYU Depth Dataset V2[1] by minimizing the objective function proposed in [34]. In runtime, it predicts the depth map of center view based on the input image patches. However, the predicted depth map is not directly involved in the light field synthesis, but rather to be exploited to generate the depth probability map. Such process is illustrated in details in Section IV-C.

[1]Available at http://cs.nyu.edu/~silberman/datasets/nyu_depth_v2.html

B. MULTI-SCALE DICTIONARY INFERENCE MODULE

The major component of multi-scale dictionary inference module is a DDL model with Bayesian pooling. This component consists of three major steps, namely, 1) layer-level dictionary learning; 2) dictionary weight downsampling; 3) Bayesian pooling with dark channel prior.

1) LAYER-LEVEL DICTIONARY LEARNING

Here, we introduce the detailed derivation of the dictionary learning as described in Section III. The dictionary assignment z_i for the i^{th} image patch is sampled from an indicator distribution β_k, which is assumed to follow a Dirichlet distribution. The k^{th} global dictionary (denoted as θ_k) and the weight for the i^{th} local dictionary (denoted as π_i) both follow the normal distribution. The i^{th} local feature dictionary G_i is obtained by multiplying π_i with the set of global dictionary $\boldsymbol{\theta} = \{\theta_k\}$. The image patch $\boldsymbol{\phi}_i$ is then selected from the corresponding local dictionary G_i by the sparse dictionary assignment z_i with an additional Gaussian noise. Mathematically,

$$\begin{aligned} 3\beta_k &\sim \mathrm{Dir}(b_0), \quad z_i \Big| \left\{\beta_k\right\}_{k=1}^{K} \sim \prod_{k=1}^{K} \beta_k, \\ \theta_k &\sim N(0, p^{-1}\mathbf{I}), \quad \pi_i \sim N(0, \alpha^{-1}\mathbf{I}), \\ G_i &= \pi_i \cdot \boldsymbol{\theta}, \quad \boldsymbol{\phi}_i \sim N(G_i \cdot z_i, \sigma_x^{-1}\mathbf{I}), \end{aligned} \tag{6}$$

where b_0, p, α, σ_x are gamma priors. These hyper-parameters are not sensitive to the results, and all are set as Gamma(10^{-6}, 10^{-6}).

2) DICTIONARY WEIGHT DOWNSAMPLING

After π_i is obtained at each layer, we compute the weight $\hat{\pi}_{x,y}$ for each pixel (x, y) by averaging the weights of nearby image patches.

Next, we split $\hat{\pi}_{x,y}$ in each perspective into 2×2 regions. In each region, we apply max-pooling to select one representative pixel over a small region, and obtain the down-sampling index $\boldsymbol{S}$. These pixels pass through the sigmoid activation layer, and are then reshaped into a vector. This step also converts the pixel weighting into patch weighting, and constructs the downsampled weights $\boldsymbol{\pi}_{\mathrm{down}}$. Lastly, they are fed into the next layer together with $\boldsymbol{S}$.

3) BAYESIAN POOLING WITH DARK CHANNEL PRIOR

Using Eq. 6, we can derive the new input feature ϕ^{new} for the next layer, i.e.,

$$\begin{aligned} \phi^{\mathrm{new}} &= \boldsymbol{G}_{\mathrm{down}} \cdot \boldsymbol{z}_{\mathrm{down}} \\ &= \boldsymbol{\theta}_{\mathrm{down}} \cdot (\boldsymbol{\pi}_{\mathrm{down}} \odot \boldsymbol{z}_{\mathrm{down}}), \end{aligned} \tag{7}$$

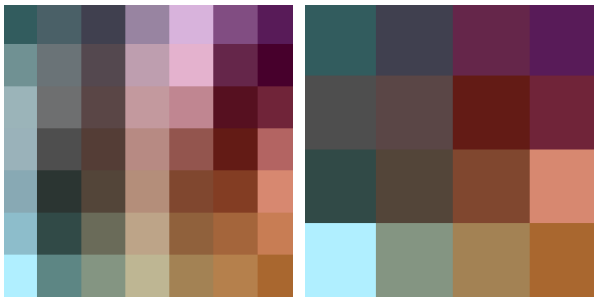
where $\odot$ is piecewise product and $\boldsymbol{z}_{\mathrm{down}}$ is the downsampled dictionary assignment over the 2×2 region mask from the previous layer, based on the downsampling index $\boldsymbol{S}$.

The downsampled local feature dictionary $\boldsymbol{G}_{\text{down}}$ is the linear combination of the downsampled global feature dictionary $\boldsymbol{\theta}_{\text{down}}$ and the weight $\boldsymbol{\pi}_{\text{down}}$. Inspired by [37] and [38], we incorporate the dark channel prior with Bayesian pooling to obtain the downsampled global feature dictionary $\boldsymbol{\theta}_{\text{down}}$ from the previous layer. The dark channel prior has been demonstrated to be capable of enhancing the depth information learned by the global feature dictionary, and therefore it can improve the reconstruction results.

For each pixel at $(\hat{x}, \hat{y})$, the dark channel

$$\boldsymbol{\phi}_i^{\text{dark}}(x, y) = \min_{c\in\{r,g,b\}} \Big(\min_{(\hat{x},\hat{y})\in\Omega(x,y)} \boldsymbol{\phi}_i^c(\hat{x}, \hat{y}) \Big), \tag{8}$$

where c is the RGB color channel of the image, and $\Omega(x, y)$ denotes the set of pixels (x, y) in the image patch $\boldsymbol{\phi}_i$. For this multi-layer structure, we obtain the minimum intensity from the image patches in the RGB channels independently. Fig. 2 presents an example about the effects of applying the dark channel prior pooling to the original input dictionary (on a single channel).



(a) (b)

FIGURE 2. Bayesian pooling with dark channel prior. (a) The original dictionary in a single channel. (b) The dictionary after the Bayesian pooling with dark channel prior operation.

At the last layer, we find the minimum intensity among the RGB channels to construct a purely dark channel, and then use the global feature dictionary $\boldsymbol{\theta}$ directly. However, except for the last layer, the learned feature dictionaries of the previous layers are all applied to the dark channel prior pooling operation. Therefore, the multi-scale dictionaries are constructed accordingly in different layers (each layer generates the feature dictionary at a certain scale), and subsequently fed into the SDR model for novel view synthesis (as shown in Fig. 1).

The size of the reconstructed image ϕ^{new} at the new layer is downsampled with a 2×2 region mask. The dictionary $\boldsymbol{G}_{\text{down}}$ at this new layer can then be learned with this new input image. Fig. 3 compares the sparsity of the dictionary learned from (1) full-size image in the previous layer, (2) directly downsampled image ϕ_{down}, and (3) the reconstructed image using DDL model in this layer ϕ^{new}. The x-axis shows the threshold for the weights $\boldsymbol{\pi}$, while the y-axis shows the number of dictionary components (feature dimension) with the weights higher than the threshold. It is evident that the

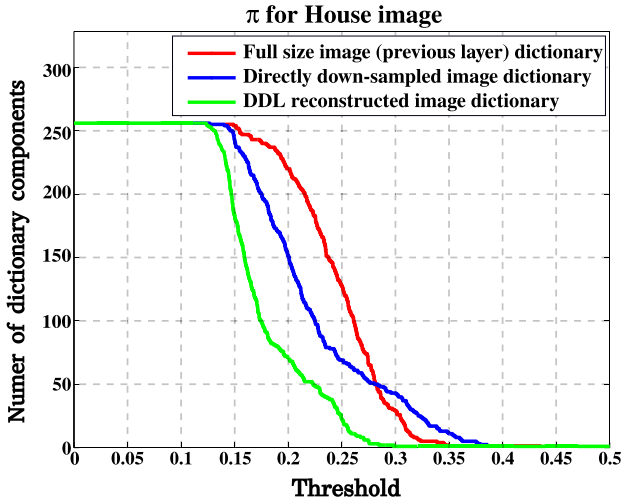


FIGURE 3. Comparing the number of components of the inferred dictionary from the full-size image, downsampled image, and reconstructed image using DDL model, respectively.

DDL reconstructed image dictionary achieves better sparsity compared with other dictionaries.

C. SIGMOID DICTIONARY RENDERING MODULE

The major component in this module is a SDR model used for light field reconstruction. As is shown in Fig. 1, the SDR model can be described as a network with three stacked layers (Layer 1–3), and each layer contains three major operations:

1) Depth probability map construction;
2) View synthesis at each layer (for algorithm details, please refer to Section V-C);
3) Depth map synthesis for different viewpoints.

1) DEPTH PROBABILITY MAP CONSTRUCTION

The **Layer 0** in Fig. 1 converts a continuous depth map into a depth probability map. We demonstrate the details of depth probability map in Fig. 4. To illustrate the merits of using depth probability map, we also compare another depth map generated using the maximum likelihood method.

Assuming that the prior probability of depth $P(D)$ and the multi-scale dictionary $P(\boldsymbol{G})$ are independent, we derive the marginal pixel shift probability in the x-direction as

$$P(z(x)|\boldsymbol{G}, D) = \sum_d P(a_d = d|D) \cdot P(z(x-d)|\boldsymbol{G}, a_d), \tag{9}$$

where $P(z|\boldsymbol{G}, \cdot)$ follows a Bernoulli distribution with the posterior probability β_k of the dictionary assignment $z(x)$ for each pixel in the DDL model, and a_d is the pixel shift. The term $P(a_d|D)$ is a normalized differential sigmoid function of depth D from the CNN mode. The joint probability of a single pixel follows

$$P(z(x-d) = k, a_d = d|\boldsymbol{G}, D) = \beta_k(x) \cdot \frac{S'(d)}{\sum_d S'(d)}, \tag{10}$$

where $z(x-d)$ is the dictionary assignment for the pixel x with depth shift equal to d. As a consequence, the depth probability

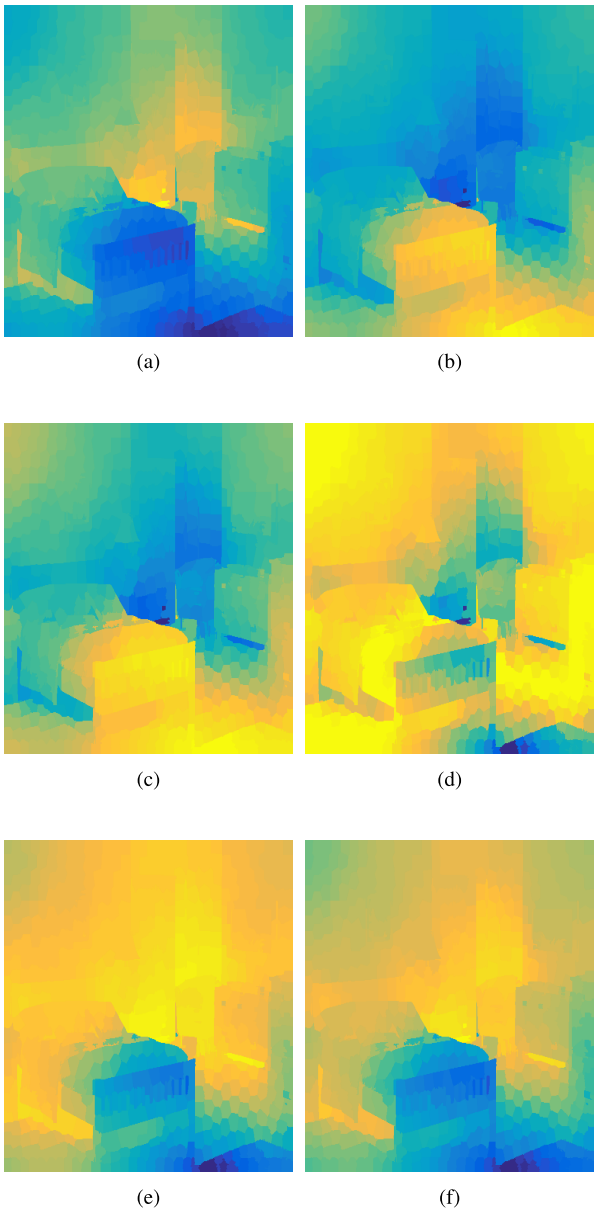


FIGURE 4. Maximum likelihood depth map estimation and the depth probability map for different depth values. Yellowish pixels indicate high probability, while blue pixels indicate lower probability. For depth layers from (b) to (f), probabilities in the nearer region become smaller and the probabilities for objects far away in the image become larger. (a) Maximum likelihood depth map. (b) Probability depth map on depth value equals to 4. (c) Probability depth map on depth value equals to 8. (d) Probability depth map on depth value equals to 12. (e) Probability depth map on depth value equals to 16. (f) Probability depth map on depth value equals to 20.

map of the entire view can be easily acquired by applying the operation defined by Eq. 10 on each pixel of the view.

2) VIEW SYNTHESIS AT EACH LAYER

The synthesized views are hierarchically reconstructed in the SDR model using the corresponding multi-scale dictionary provided by the DDL model. In other words, the synthesized view at the $(l)^{\text{th}}$ layer in SDR model is reconstructed by using the learned local feature dictionary at the $(-l)^{\text{th}}$ layer in the DDL model. Formally, such process can be described as

$$\boldsymbol{\phi}^{(l)} = \boldsymbol{G}^{(-l)} \cdot z_{\text{up}} = \boldsymbol{\theta}^{(-l)} \cdot (\boldsymbol{\pi}_{\text{up}} \odot z_{\text{up}}), \tag{11}$$

where z_{up} denotes the up-sampled dictionary assignment and $\boldsymbol{\pi}_{up}$ stands for the weights over the 2×2 region mask from the previous layer. With the same region mask for downsampling and up-sampling, the global feature dictionary at the $(-l)^{\text{th}}$ layer in the DDL model and the $(l)^{\text{th}}$ layer in the SDR model share the same weights. The dictionary assignment z at this layer can then be learned with the reconstructed view $\boldsymbol{\phi}^{(l)}$.

3) DEPTH MAP SYNTHESIS FOR DIFFERENT VIEWPOINTS

Unlike existing view synthesis methods, the proposed DNB model enables us to synthesize different depth maps for different viewpoints. Fig. 5 illustrates the estimated depth map for 9 (3×3) different viewpoints of "Raspberry."

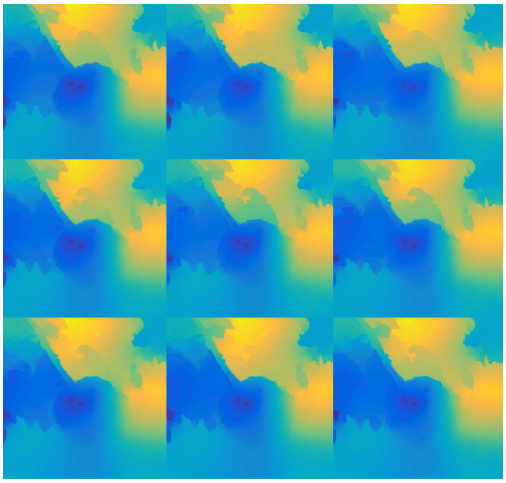


FIGURE 5. Synthesized depth maps for 9 (3×3) different viewpoints using the "Raspberry" dataset.

The synthesized depth maps at different viewpoints are further used to reconstruct the corresponding view in the next layer of the SDR model. Refinement of these multi-view depth maps is based on the colorization algorithm [39] and the corresponding reconstructed RGB view at this layer, which is crucial for preventing shift distortion.

V. INFERENCE ALGORITHM FOR THE MULTI-SCALE SHNB MODEL

In this section, we first introduce the training procedure of the DNB model, using a Gibbs sampling algorithm [40]. Then, we illustrate how to compute the multi-scale dictionary from a single image, and how the feature dictionary can be used to reconstruct the novel views.

A. TRAINING PROCEDURE

The training procedure is crucial to enabling the proposed DNB model synthesizing the novel views.

1) Pretrain the CNN model on NYU Depth Dataset V2 [41] by minimizing the conditional random field energy function [34].
2) Train the DDL model and SDR model jointly based on the depth probability map acquired from differential sigmoid layer.
3) Fine-tune the entire DNB model using light field saliency dataset (LFSD) [42], [43].

B. MULTI-SCALE DICTIONARY LEARNING BASED ON BAYESIAN POOLING

As introduced in Section IV-B, the probability of the multi-scale dictionary is given by

$$\begin{aligned} &P(\boldsymbol{G}|\boldsymbol{G}_{\text{down}}, z_{\text{down}}, \sigma_x, H) \\ &\quad = \ell(\boldsymbol{G}; \boldsymbol{G}_{\text{down}}, z_{\text{down}}, \sigma_x) \cdot p(\boldsymbol{G}) \\ &\quad \propto N\left(\boldsymbol{G} \cdot z_{\text{down}}; \boldsymbol{G}_{\text{down}} \cdot z_{\text{down}}, \sigma_x^{-1}\mathbf{I}\right) \\ &\qquad \cdot N(\boldsymbol{\pi}; 0, \alpha^{-1}\mathbf{I}_{\mathbf{K}}) \cdot H(\boldsymbol{\theta}), \end{aligned} \tag{12}$$

where $\ell(\cdot)$ is the likelihood based on the downsampled dictionary assignment z_{down} inferred in the previous layer, and the dictionary with downsampling $\boldsymbol{G}_{\text{down}}$ in Eq. 7. This likelihood follows a normal distribution as in Eq. 6. Moreover, this downsampled dictionary contains the weights $\boldsymbol{\pi}$ and the local dictionary $\boldsymbol{\theta}$. The former follows a Gaussian distribution with a Gamma prior α, and H is the prior probability of the downsampled local dictionary $\boldsymbol{\theta}$, which can be assumed to follow a Gaussian distribution. Based on this probability, we can sample the dictionary using Algorithm 1.

C. MULTI-SCALE VIEW RECONSTRUCTION WITH BAYESIAN UNPOOLING

The probability of the multi-scale view reconstruction is given by

$$\begin{aligned} &P(z^{(l)}|\boldsymbol{G}^{(-l)}, z_{up}, \sigma_x, \boldsymbol{\beta}) \\ &\quad = \ell(z^{(l)}; \boldsymbol{G}, z_{up}, \sigma_x) \cdot p(z^{(l)}|\boldsymbol{\beta}) \\ &\quad \propto N\left(\boldsymbol{G}^{(-l)} \cdot z^{(l)}; \boldsymbol{G}^{(-l)} \cdot z_{up}, \sigma_x^{-1}\mathbf{I}\right) \cdot \text{Bernoulli}(z^{(l)}|\boldsymbol{\beta}), \end{aligned} \tag{13}$$

where the likelihood $\ell(\cdot)$ depends on the unpooling z_{up} in Eq. 11. This likelihood also follows a normal distribution, and $p(z^{(l)}|\boldsymbol{\beta})$ is the prior with parameter $\boldsymbol{\beta}$.

Algorithm 1 Multi-Scale SHNB Dictionary Inference

Input: Single image patches $\boldsymbol{x}_i, i = 1, \ldots, N$
Output: multi-scale global dictionary $\boldsymbol{\theta}_k, K = 1, \ldots, K$

Initialize $\boldsymbol{\pi}$ based on hyperparameter α in Eq. 6
Initialize $\boldsymbol{\theta}$ based on base measure in Eq. 6
Initialize assignment $\boldsymbol{z}$ based on prior $\boldsymbol{\beta}$ in Eq. 6
for Layers in the DDL model **do**
  Initialize $\boldsymbol{\beta}$ based on the corresponding hyper-parameter in Eq. 6
  for Iteration iter $= 1 : T_{\max}$ **do**
    Locally sample a sparse assignment $\boldsymbol{z}$ from the posterior multinomial distribution based on data ϕ, weight $\boldsymbol{\pi}$, global dictionary $\boldsymbol{\theta}$, and assignment distribution $\boldsymbol{\beta}$.
    Locally sample the weight $\boldsymbol{\pi}$ from the posterior Gaussian distribution based on data $\boldsymbol{x}_i$, sparse assignment $\boldsymbol{z}$, global dictionary $\boldsymbol{\theta}$, and prior α.
    Globally sample the dictionary $\boldsymbol{\theta}$ from the posterior distribution based on data $\boldsymbol{x}_i$, weight $\boldsymbol{\pi}$, global dictionary $\boldsymbol{\theta}$, and sparse assignment $\boldsymbol{z}$.
    Update $\boldsymbol{\beta}$ from the posterior Dirichlet distribution based on hyper-parameter b_0 and sparse assignment $\boldsymbol{z}$.
  end for
  Bayesian pooling $\boldsymbol{\pi}_{\text{down}}$ to obtain the downsampling index $\boldsymbol{S}$;
  Downsample dictionary $\boldsymbol{\theta}_k$ and assignment $\boldsymbol{z}$ based on index $\boldsymbol{S}$ with Eq. 7
  Compute the downsampled dictionary $\boldsymbol{G}$ from Eq. 6
  Compute the reconstructed image ϕ for the next layer from Eq. 7
end for

VI. EXPERIMENTS

We present two sets of experiments to highlight the merits of the proposed DNB model, namely:

1) Comparisons with view synthesis and image inpainting methods;
2) Quantitative experiments for light field synthesis.

A. COMPARISONS WITH VIEW SYNTHESIS AND IMAGE INPAINTING

We compare the DNB model with the following three methods:

1) Light field synthesis from RGB-D with depth propagation [14]. The depth map used is extracted from the DCNF model [34].
2) The depth propagation approach [14] together with the beta process factor analysis (BPFA) inpainting [44]. The latter is used to improve the synthesis result of the depth propagation algorithm.
3) 2D-to-3D conversion with deep learning [13].

Fig. 6 shows an example with some zoomed-in regions, which together demonstrate that our approach performs better. Compared with the Deep3d approach in (a), our method obtains a smooth result with the DDL model to recover the novel views, while Deep3d shows some conspicuous noise, especially in the blue region. This is because it uses the L_1-inpainting algorithm to handle the occlusions caused by the novel view synthesis. In (b), the recovered novel view by depth propagation suffers from shift distortion, especially in the red regions, where the leaf is distorted. In comparison, the DNB model can avoid this problem, as the probabilistic discrete shift map in SDR can construct a pixel shift with a mixture model, which is more robust. Lastly, the single-depth

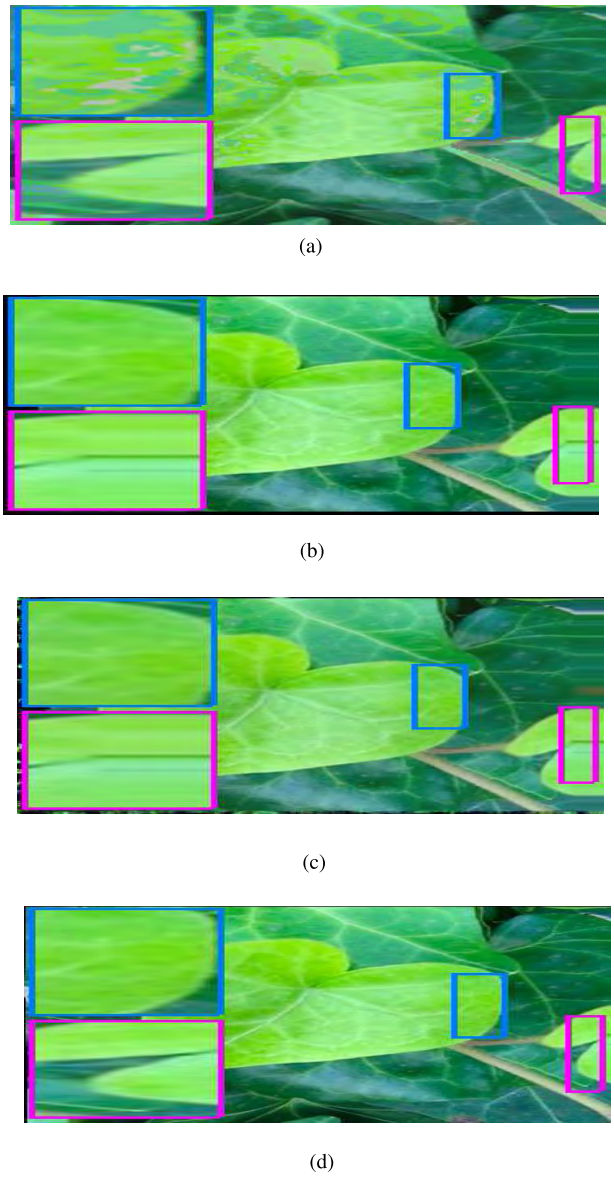


FIGURE 6. Qualitative comparison of a novel view for the "Leaf" dataset. (a) Synthesized views from Deep3d. (b) Synthesized views from depth propagation. (c) Synthesized views from depth propagation with inpainting. (d) Synthesized views from the DNB model.

propagation method with BFPA inpainting, as shown in (c), can handle some occlusion holes. Nevertheless, the problem of shift distortion remains.

Fig. 7 and Fig. 8 present two other examples. The former shows the synthesized light field for the "Dragon" dataset without considering the point spread function. The latter shows the synthesized light field for the "indoor" dataset with an additional Gaussian point spread function. Both illustrate that our model enables the synthesis of a light field without precise depth information.

B. QUANTITATIVE EVALUATION OF THE LIGHT FIELD SYNTHESIS

Using the HCI light field dataset and the LFSD dataset, we evaluate the quantitative performance using both

FIGURE 7. Synthesized light field and sub-aperture views representation of the "Dragon" dataset. (a) Synthesized light field. (b) Sub-aperture images.

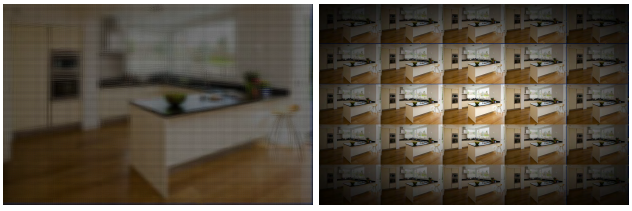


FIGURE 8. Synthesized light field and sub-aperture views representation of the "indoor" dataset. (a) Synthesized light field. (b) Sub-aperture images.

mean squared error (MSE) and structural similarity of image (SSIM) [45] methods. The MSE is computed as

$$\mathrm{MSE}(I_0, I_D) = \frac{\sum_x \sum_y \left(I_0(x,y) - I_D(x,y)\right)^2}{(\text{Length of } I_0) \times (\text{Width of } I_0)}, \quad (14)$$

where $I_0(x, y)$ is the original light field image in the target viewpoint, and $I_D(x, y)$ is the synthesized image. A smaller MSE indicates a better performance. Table 2 shows that the proposed DNB model has a lower MSE compared with the depth propagation (DP) [14] + BPFA [44] for all 5×5 views.

TABLE 2. MSE comparison with the proposed DNB model versus depth propagation (DP) [14] + BPFA [44].

DNB	View 1	View 2	View 3	View 4	View 5
View 1	2.16	1.72	1.36	1.65	2.10
View 2	1.61	0.93	0.58	0.86	1.38
View 3	1.23	0.55	0.00	0.48	0.96
View 4	1.75	1.07	0.72	1.00	1.47
View 5	2.60	2.05	1.70	1.98	2.22

DP+BPFA	View 1	View 2	View 3	View 4	View 5
View 1	2.57	2.13	1.76	2.16	2.70
View 2	2.02	1.34	0.99	1.38	1.99
View 3	1.59	0.91	0.00	0.96	1.54
View 4	2.11	1.44	1.09	1.48	2.04
View 5	2.90	2.35	1.99	2.40	2.70

The result of MSE evaluation on the "Leaf" dataset from Fig 6 is shown in Fig. 9. It shows that the proposed DNB model also has a satisfactory quantitative performance. The synthesized view based on the Deep3d model has a better MSE performance compared to depth propagation methods

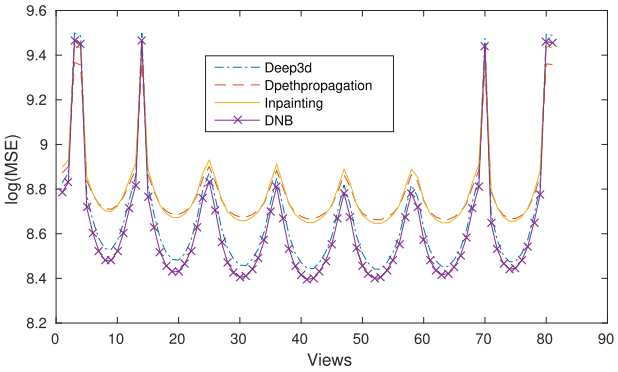


FIGURE 9. Comparison result of log(MSE) for the "Leaf" dataset.

because the shift of pixels for Deep3d is smaller than the other methods.

The SSIM is computed as

$$\text{SSIM}(I_0, I_D) = \frac{(2\mu_0\mu_D + c_1)(2\sigma_{0,D} + c_2)}{(\mu_0^2 + \mu_D^2 + c_1)(\sigma_0^2 + \sigma_D^2 + c_2)}, \quad (15)$$

where μ_0 and μ_D are the mean of images patches I_0 and I_D, respectively. The variables σ_0 and σ_D are respectively the variance of image patches I_0 and I_D. Then, $\sigma_{0,D}$ is the covariance of images patches I_0 and I_D, and c_1 and c_2 are used to stabilize the division with a weak denominator [45]. We take the window size as 8×8 for image patches I_0 and I_D [45]. When the SSIM is close to 1, the 4D light field synthesis algorithm shows a good performance.

We compared the SSIM among the four methods as in Fig. 6, where we first select one novel view from the synthesized light field or 3D image based on each method. We then compute their SSIM with all perspectives of the captured light field data. We plot log(SSIM) and tabulate their mean values to show that the synthesized light field with the proposed DNB model has a better structural similarity compared with the other methods.

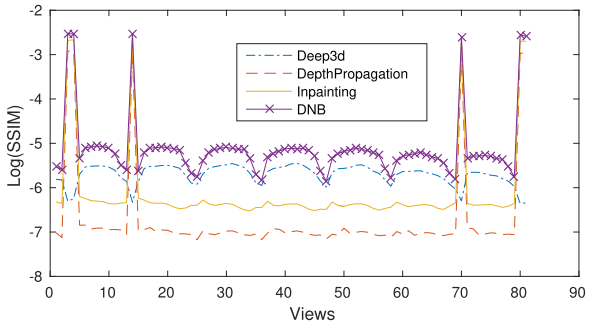


FIGURE 10. Comparison result of log(SSIM) for the "Raspberry" dataset.

First, Fig. 10 shows the experiment conducted with the "Raspberry" dataset. The blue dash-dot line of the Deep3d model indicates that the synthesized view has a better structural similarity to the center view (view = 41) of the captured light field. This is because the angular difference of the synthesized 3D data is quite small. SSIM results of the synthesized light field from depth propagation, depth propagation with inpainting, and the proposed DNB model are shown with an orange dashed line, a yellow solid line, and purple line with cross markers, respectively. SSIMs of these synthesized light field have better results for the boundary views, as the views among the light field have a larger angular difference. Overall, this plot shows that the synthesized view with the proposed DNB model has a better image structural similarity compared with the other three methods.

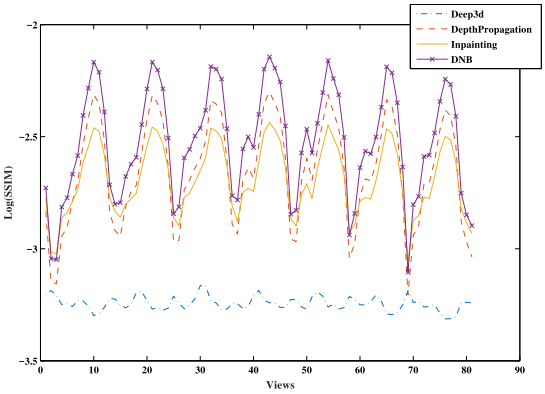


FIGURE 11. Comparison result of log(SSIM) for the "Flower" dataset.

Second, Fig. 11 shows the experiment conducted for the "Flower" light field in the LFSD dataset. This experiment is used to show that inpainting for the light field synthesis based on depth propagation may cause undesirable effects. The orange dashed line, which represents depth propagation result, has a better performance than the result of depth propagation with BPFA inpainting (yellow solid line). Moreover, the SSIM result of DNB is better than the other methods.

TABLE 3. Comparison of the mean of log(SSIM).

Synthesis methods	"Raspberry"	"Flower"	"Dragon"
Deep3d	−5.68	−7.46	−5.02
Depth Propagation (DP)	−5.35	−6.11	−4.91
DP + BPFA	−5.01	−6.19	−4.80
DNB model	−4.56	−5.81	−4.30

TABLE 4. Paired T-test between every two methods.

	Deep3D	DP	DP+BPFA	DNB
Deep3D	–			
DP	○	–		
DP+BPFA	○	○	–	
DNB	●	○	●	–

Similarly, Table 3 illustrates that the proposed DNB model has a better image structural similarity compared with the capture light field over various datasets. Better SSIM performance indicates that the DNB model enables the synthesis of light field with a more similar structure to the ground truth. We also test the paired T-test statistics [46]–[48] between every two methods to illustrate that the proposed model leads

to improvements with statistical significance, with results given in Table 4. The symbol ∘ means that there is no significant difference on performance between the two compared methods, while • represents that there exists significant difference.

VII. CONCLUSIONS

The proposed deep nonparametric Bayesian model performs well in synthesizing the light field from a single image, and reconstructing and inpainting the novel views, especially when a very accurate depth map cannot be obtained. This paper has two major contributions. First, the DNB model enables us to infer a multi-scale feature dictionary. The Bayesian pooling procedure can find a suitable downsampled feature dictionary and can make it sparser. Subsequently, we can use it to infer the joint probability of the depth map. With this probabilistic depth map, the multiscale Bayesian up-pooling model allows us to reconstruct and inpaint multiple views from a single image. This joint probability of depth map helps to synthesize the light field in a robust manner. These experiments demonstrated that our method outperforms several state-of-the-art light field synthesis approaches.

REFERENCES

[1] R. Ng, M. Levoy, and M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," *Stanford Univ. Comput. Sci. Tech. Rep.*, vol. 2, no. 11, pp. 1–11, Apr. 2005.

[2] M. Levoy, R. Ng, A. Adams, M. Footer, and M. Horowitz, "Light field microscopy," *ACM Trans. Graph.*, vol. 25, no. 3, pp. 924–934, 2006.

[3] 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.*, Oct. 2016, pp. 121–138.

[4] D. Tsai, D. G. Dansereau, T. Peynot, and P. Corke, "Image-based visual servoing with light field cameras," *IEEE Robot. Autom. Lett.*, vol. 2, no. 2, pp. 912–919, Apr. 2017.

[5] D. G. Dansereau, G. Schuster, J. Ford, and G. Wetzstein, "A wide-field-of-view monocentric light field camera," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 3757–3766.

[6] R. Ng, "Digital light field photography," Ph.D. dissertation, dept. Comput. Sci., Stanford Univ., California, MA, USA, 2006.

[7] N. Chen, Z. Ren, and E. Y. Lam, "High-resolution Fourier hologram synthesis from photographic images through computing the light field," *Appl. Opt.*, vol. 55, no. 7, pp. 1751–1756, Mar. 2016.

[8] N. Chen, Z. Ren, D. Li, E. Y. Lam, and G. Situ, "Analysis of the noise in backprojection light field acquisition and its optimization," *Appl. Opt.*, vol. 56, no. 13, pp. F20–F26, May 2017.

[9] K. Marwah, G. Wetzstein, Y. Bando, and R. Raskar, "Compressive light field photography using overcomplete dictionaries and optimized projections," *ACM Trans. Graph.*, vol. 32, no. 4, pp. 46–57, Jul. 2013.

[10] Z. Xu and E. Y. Lam, "A high-resolution lightfield camera with dual-mask design," *Proc. SPIE*, vol. 8500, Aug. 2012, Art. no. 85000U.

[11] Z. Xu, J. Ke, and E. Y. Lam, "High-resolution lightfield photography using two masks," *Opt. Express*, vol. 20, no. 10, pp. 10971–10983, May 2012.

[12] B. Chen, G. Polatkan, G. Sapiro, D. Blei, D. Dunson, and L. Carin, "Deep learning with hierarchical convolutional factor analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1887–1901, Aug. 2013.

[13] J. Xie, R. Girshick, and A. Farhadi, "Deep3D: Fully automatic 2D-to-3D video conversion with deep convolutional neural networks," in *Proc. Eur. Conf. Comput. Vis.*, Oct. 2016, pp. 842–857.

[14] C. Sun, Y. Wu, and B. Zeng, "Synthesis of light-field raw data from RGB-D images," in *Proc. IEEE Int. Conf. Image Process.*, Dec. 2015, pp. 1448–1452.

[15] K. Mitra and A. Veeraraghavan, "Light field denoising, light field superresolution and stereo camera based refocussing using a GMM light field patch prior," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2012, pp. 22–28.

[16] X. Cun, F. Xu, C. M. Pun, and H. Gao, "Depth assisted full resolution network for single image-based view synthesis," *IEEE Comput. Graph. Appl.*, to be published.

[17] A. Dosovitskiy *et al.*, "FlowNet: Learning optical flow with convolutional networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, Jun. 2015, pp. 2758–2766.

[18] L. McMillan and G. Bishop, "Plenoptic modeling: An image-based rendering system," in *Proc. 22nd Annu. Conf. Comput. Graph. Interact. Techn.*, Sep. 1995, pp. 39–46.

[19] N. Chen, C. Zuo, E. Y. Lam, and B. Lee, "3D imaging based on depth measurement technologies," *Sensors*, vol. 18, no. 11, p. 3711, Nov. 2018.

[20] E. Y. Lam, "Computational photography with plenoptic camera and light field capture: Tutorial," *J. Opt. Soc. Amer. A*, vol. 32, no. 11, pp. 2021–2032, Nov. 2015.

[21] M. Levoy and P. Hanrahan, "Light field rendering," in *Proc. 23rd Annu. Conf. Comput. Graph. Interact. Techn.*, Aug. 1996, pp. 31–42.

[22] Z. Xu and E. Y. Lam, "Light field superresolution reconstruction in computational photography," in *Proc. Top. Meeting Signal Recovery Synth.*, Jul. 2011, p. SMB3.

[23] B. Wilburn *et al.*, "High performance imaging using large camera arrays," *ACM Trans. Graph.*, vol. 24, no. 3, pp. 765–776, Jul. 2005.

[24] X. Sun, N. H. C. Yung, E. Y. Lam, and H. K. So, "Computationally efficient hyperspectral data learning based on the Doubly stochastic dirichlet process," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 1, pp. 363–374, Jan. 2017.

[25] X. Sun, N. H. C. Yung, and E. Y. Lam, "Unsupervised tracking with the doubly stochastic Dirichlet process mixture model," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 9, pp. 2594–2599, Sep. 2016.

[26] S. Wanner and B. Goldluecke, "Variational light field analysis for disparity estimation and super-resolution," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 3, pp. 606–619, Mar. 2014.

[27] X. Sun, N. Meng, Z. Xu, E. Y. Lam, and H. K. So, "Sparse hierarchical nonparametric Bayesian learning for light field representation and denoising," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2016, pp. 3272–3279.

[28] C. Wang and D. M. Blei, "Decoupling sparsity and smoothness in the discrete hierarchical dirichlet process," in *Proc. Adv. Neural Inf. Process. Syst.*, 2009, pp. 1982–1989.

[29] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[30] M. Jaderberg, A. Vedaldi, and A. Zisserman, "Deep features for text spotting," in *Proc. Eur. Conf. Comput. Vis.*, Oct. 2014, pp. 512–528.

[31] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[32] Z. Ren, Z. Xu, and E. Y. Lam, "Learning-based nonparametric autofocusing for digital holography," *Optica*, vol. 5, no. 4, pp. 337–344, Apr. 2018.

[33] Z. Ren, Z. Xu, and E. Y. M. Lam, "End-to-end deep learning framework for digital holographic reconstruction," *Adv. Photon.*, vol. 1, no. 1, Jan. 2019, Art. no. 016004.

[34] F. Liu, C. Shen, and G. Lin, "Deep convolutional neural fields for depth estimation from a single image," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 5162–5170.

[35] P. Wang, X. Shen, Z. Lin, S. Cohen, B. Price, and A. Yuille, "Towards unified depth and semantic prediction from a single image," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 2800–2809.

[36] X. Sun, Z. Xu, N. Meng, E. Y. Lam, and H. K. So, "Data-driven light field depth estimation using deep convolutional neural networks," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, Jul. 2016, pp. 367–374.

[37] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2341–2353, Dec. 2011.

[38] J. Pan, D. Sun, M.-H. Yang, and H. Pfister, "Blind image deblurring using dark channel prior," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1628–1636.

[39] A. Levin, D. Lischinski, and Y. Weiss, "Colorization using optimization," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 689–694, 2004.

[40] R. M. Neal, "Markov chain sampling methods for Dirichlet process mixture models," *J. Comput. Graph. Statist.*, vol. 9, no. 2, pp. 249–265, Jun. 2000. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/10618600.2000.10474879
[41] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, "Indoor segmentation and support inference from RGBD images," in *Proc. Eur. Conf. Comput. Vis.*, Oct. 2012, pp. 746–760.
[42] N. Li, J. Ye, Y. Ji, H. Ling, and J. Yu, "Saliency detection on light field," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 2806–2813.
[43] N. Li, B. Sun, and J. Yu, "A weighted sparse coding framework for saliency detection," in *Proc. IEEE Conf. CVPR*, Jun. 2015, pp. 5216–5223.
[44] M. Zhou *et al.*, "Nonparametric Bayesian dictionary learning for analysis of noisy and incomplete images," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 130–144, Jan. 2012.
[45] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. 37th Asilomar Conf. Signals, Syst. Comput.*, vol. 2. Nov. 2003, pp. 1398–1402.
[46] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *J. Mach. Learn. Res.*, vol. 7, pp. 1–30, Jan. 2006.
[47] L. Jiang, L. Zhang, C. Li, and J. Wu, "A correlation-based feature weighting filter for naive bayes," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 2, pp. 201–213, Feb. 2019.
[48] L. Jiang, C. Li, S. Wang, and L. Zhang, "Deep feature weighting for naive Bayes and its application to text classification," *Eng. Appl. Artif. Intell.*, vol. 52, pp. 26–39, Jun. 2016.

NAN MENG received the bachelor's degree from the University of Electronic Science and Technology of China, in 2015. He is currently pursuing the Ph.D. degree with the Department of Electrical and Electronic Engineering, The University of Hong Kong. His research interests include machine learning, computer vision, and computational biology.

XING SUN received the Ph.D. degree from The University of Hong Kong, in 2016. He is currently a Senior Researcher with Tencent. His research interests include machine learning, computer vision, and artificial intelligence.

HAYDEN K.-H. SO (S'03–M'07–SM'15) received the B.S., M.S., and Ph.D. degrees in electrical engineering and computer sciences from the University of California at Berkeley, Berkeley, CA, USA, in 1998, 2000, and 2007, respectively. He is currently an Associate Professor with the Department of Electrical and Electronic Engineering, The University of Hong Kong. He received the Croucher Innovation Award, in 2013, for his work in a power-efficient high-performance heterogeneous computing system, the University Outstanding Teaching Award (Team), in 2012, and the Faculty Best Teacher Award, in 2011. He has served as the Technical Program Chair for various international conferences, including the 2014 International Conference on Field-Programmable Technology, the 2014 International Symposium on Highly-Efficient Accelerators and Reconfigurable Technologies, and the 2015 IEEE International Conference on Application-Specific Systems, Architectures, and Processors. He also served as the Multiprocessor Systems and Networks on Chip Track Co-Chair for the International Conference on Reconfigurable Computing and FPGAs and a Guest Editor for the *Journal of Signal Processing Systems*.

EDMUND Y. LAM (M'00–SM'05–F'15) received the B.S., M.S., and Ph.D. degrees in electrical engineering from Stanford University. From 2010 to 2011, he was a Visiting Associate Professor with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology. He is currently a Professor in electrical and electronic engineering with The University of Hong Kong, where he is also the Director of the Computer Engineering Program. His main research interest includes computational imaging. He is a Fellow of OSA, SPIE, IS&T, and HKIE. He was a recipient of the IBM Faculty Award.