IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received February 22, 2022, accepted April 26, 2022, date of publication May 3, 2022, date of current version May 12, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3172343

Simultaneous Color Restoration and Depth Estimation in Light Field Imaging

YONGWEI LI[1], FILIBERTO PLA[2], MÅRTEN SJÖSTRÖM[1], (Senior Member, IEEE), AND RUBEN FERNANDEZ-BELTRAN[3], (Senior Member, IEEE)

[1]Department of Information Systems and Technology, Mid Sweden University, 85170 Sundsvall, Sweden
[2]Institute of New Imaging Technologies, University Jaume I, 12071 Castellón de la Plana, Spain
[3]Department of Computer Science and Systems, University of Murcia, 30100 Murcia, Spain

Corresponding authors: Filiberto Pla (pla@uji.es) and Mårten Sjöström (marten.sjostrom@miun.se)

This work was supported in part by the European Union's Horizon 2020 Research and Innovation Program through the Marie Sklodowska-Curie European Training Network on Full Parallax Imaging under Grant 676401, and in part by Generalitat Valenciana under Grant AICO-2020-018 and Research Network RED2018-102511-T from the Spanish Ministry of Science, Innovation and Universities.

ABSTRACT Recent studies in the light field imaging have shown the potential and advantages of different light field information processes. In most of the existing techniques, the processing pipeline of light field has been treated in a step-by-step manner, and each step is considered to be independent from the others. For example, in light field color demosaicing, inferring the scene geometry is treated as an irrelevant and negligible task, and vice versa. Such processing techniques may fail due to the inherent connection among different steps, and result in both corrupted post-processing and defective pre-processing results. In this paper, we address the interaction between color interpolation and depth estimation in light field, and propose a probabilistic approach to handle these two processing steps jointly. This probabilistic framework is based on a Markov Random Fields —Collaborative Graph Model for simultaneous Demosaicing and Depth Estimation (CGMDD)—to explore the color-depth interdependence from general light field sampling. Experimental results show that both image interpolation quality and depth estimation can benefit from their interaction, mainly for processes such as image demosaicing which are shown to be sensitive to depth information, especially for light field sampling with large baselines.

INDEX TERMS Light field, demosaicing, depth estimation, Markov random field, graph model.

I. INTRODUCTION

With the proliferation of 3D imaging applications in the consumer markets, the processing of light field (LF) data has attracted significant interest from both academia and industry. Despite the maturity in conventional image processing techniques, the unique 4D structure of the light field proposes singular challenges, such as light field demosaicing, all-in-focus image reconstruction and depth estimation. However, the error propagation and dependency of different steps within the processing pipeline (Fig. 1) are barely discussed, and open up for several processing improvements.

Most of the previous works on demosaicing suggest interpolating missing colors from the local neighborhood on the sensor plane [1]–[4]. Such schemes are rooted from the fact that the surroundings of the scene are stored in adjacent pixels in conventional cameras. However, this adjacency condition is suboptimal to raw images captured by plenoptic cameras or multi-camera setups, as shown in Fig. 2.

Another fundamental problem in light field processing is to estimate the depth map, and it serves as a prerequisite for a variety of applications, ranging from virtual reality [5]–[7] to microscopy imaging [8], [9]. Though numerous depth estimation approaches have been proposed over the last decade, for example, stereo matching [10]–[13] and shape from focus [8], [14], [15], it is worth noting that the photo-consistency condition holds for all of them. Unfortunately, the photo-consistency criterion could be corrupted and no longer apply once demosaicing is ill-posed. Such data interpolation induces artifacts which will make the cross-view matching intractable, especially for pixel-based depth estimation such as [16].

Clearly, the classical sequential pipeline in Fig. 1 for light field capturing and processing is an ill-posed solution to

The associate editor coordinating the review of this manuscript and approving it for publication was Pu-Kun Liu.

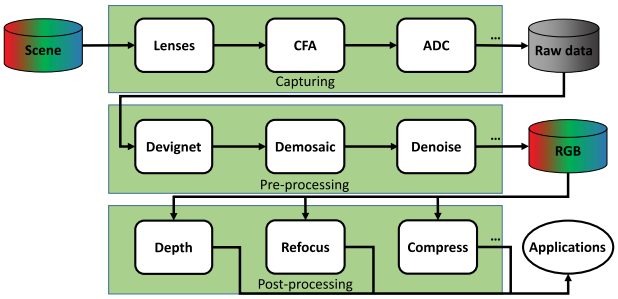


FIGURE 1. Classical light field capturing and processing pipeline. Post-processing steps are performed after the raw data captured on the sensor is pre-processed, CFA and ADC refer to color filter array (downsampling) and analog-to-digital converter (quantization) respectively.

both demosaicing and depth estimation in which inherent ambiguities are omitted. Additionally, the classical algorithms which are merely based on vision cues also suffer from noise in image formation, textureless and homogeneous regions, depth discontinuities and occlusions, which can only be handled by compromising computational complexity [3], [4], [15]. One potential solution to these challenges is to formulate correlated processes in a cascade approach [17], [18]. As opposed to the sequential pipeline, such cascade approach benefits from other correlated processes, regardless of their sequential order in traditional pipeline, so that the overall performance of different tasks can be improved in a recursive manner until desired result is attained. However, the cascade approach does not prevent error propagation among processes and inconsistent results.

In this work, we consider the interaction between the demosaicing and depth estimation problems from raw data of color light fields. Towards this end, we propose a **C**ollaborative **G**raph **M**odel for **D**emosaicing and **D**epth Estimation (CGMDD). We formulate the inter-dependence between color and depth during the interpolation and estimation process as data terms in an Markov Random Field (MRF) framework. It has been reported in [12], [16] that the image noise, depth discontinuities and occlusion can be naturally incorporated in the MRF framework because of its outstanding adaptability. Therefore, the main contributions of the work are:

(1) A probabilistic model CGMDD to address the correlation between color intensities and pixel-wise dense depth map of light field data.
(2) A joint solution for color demosaicing and depth estimation of light field data based on the proposed MRF probabilistic model.

The rest of paper is organized as follows: After reviewing related work in Section II, we propose in Section III an MRF framework CGMDD to explicitly model the correlation between color and depth, and data terms are then devised according to light field properties. In Section IV, the revised pipeline is introduced based on the general light field sampling pattern. Extensive experimental results with both synthetic and real light field datasets are shown in Section V, comparing with state-of-the-art approaches for demosaicing and depth estimation, respectively. In Section VI, we discuss the limitation of CGMDD and its adaptability based on different needs and assumptions. Finally, we conclude our work and describe future research directions in Section VII.

II. RELATED WORK

In this section, we briefly review the related research on light field demosaicing, depth estimation, and the application of MRF models in this context.

A. LIGHT FIELD DEMOSAICING

The physical property of CCD sensor allows each pixel to record only one intensity value out of three color channels. The upsampling process from raw sensor data of a single channel to full resolution color images is referred to as demosaicing. A comprehensive survey on classical demosaicing approaches can be found in [1].

Unlike classical demosaicing which has been studied since the emergence of color digital cameras [1], light field demosaicing has been overlooked for years. A widely used pipeline bluntly interpolates the missing color channels from neighboring pixels [3]. The work in [4] considers the vignetting problem in the microlenses by incorporating a white image to assign appropriate weights to pixels. Such white-image-based demosaicing method is further improved in [19] by incorporating a LFBM5D filter for plenoptic data. Besides, the authors in [20] include a learning process, in which a dictionary is trained beforehand to learn spatial and angular correlations. All of the aforementioned methods decoupled the demosaicing process with respect to depth information. The only exceptions are [17], [21], and [18]. In [17], the disparity is estimated to enable cross-view demosaicing, whereas in [21] depth-dependent blur are taken into account by employing Fourier Disparity Layer representation of the light field so that depth does not need to be explicitly estimated.

In our previous work [18] the raw pixel values are projected into a layered object space and then demosaicked on different depth planes in order to achieve a desired demosaicing result. However, the performance of such depth-based methods heavily rely on the accuracy of the estimated depths even if depth layers are used to compensate for such erroneous depths. In the worst case scenario, the quality of depth estimation conditions the accuracy of the subsequent processes, including further optimization and refinement of both color and depth results.

B. DEPTH RECONSTRUCTION OF LIGHT FIELD

Stereo matching algorithms generate sparse and feature-based depth maps, searching for one corresponding patch for each region of interest (ROI) [22]. Although more and more research focuses on generating dense depth maps [8], [9], [13], [15], the feature-based methods have gained its attention thanks to machine learning and mobile applications.

In [23], features and their descriptors are extracted from cross-hair views by training a lightweight convolutional neural network (CNN).

The dense stereo matching algorithms use a finite sliding window with various likelihood functions to determine how likely a pair of pixels correspond to each other. The most common functions include sum of absolute difference (SAD) [24], sum of squared difference (SSD) [25], normalized cross-correlation (NCC) [8], and rank transform [26]. The basic problem of local methods is to provide sufficient variation within the window while localizing accurate disparity. A large window has a higher probability of providing such information, whereas a small window locates correspondences more accurately.

Additionally, the abundant angular information of light field also gives rise to the epipolar plane image (EPI) and the focal stack. There are mainly two challenges for EPI-based methods: occlusion handling and line ambiguity. Huang *et al.* [27] proposed to remove the influence of the noise by employing a weighting mean filter which is guided by the color image. Zhang *et al.* [28] developed the robust depth estimation via spinning parallelogram operator (SPO) which uses a parallelogram operator to determine the local direction of epipolar line and it is less sensitive to occlusion compared with stereo matching. In [29], EPIs are shifted in order to retain the receptive field so that small-baseline trained networks can be adapted to estimate depths for large-baseline applications.

By integrating angular samples at different depths, focal stack enables depth from focus (DfF) approach for estimating depth [8], [30]. Central to the DfF approach is the use of photo-consistency constraint. It is often assumed the metric for focus/defocus cues will attain the extremum when the image is sharp [8]. Otherwise, constraints and assumptions such as single occluder and partial photo-consistency are often made, as in the occlusion-aware depth estimation (OCC) [15]. Other approaches are based on CNNs, as in [31], where three CNNs are combined to first generate all-in-focus image from the focal stack, followed by depth estimation and optimization steps.

On one hand, the local depth estimation methods suffer severely from occlusions, noise, depth discontinuity and light field sparsity. On the other hand, learning-based methods require massive data and are time consuming for the training stage. Such problems can be jointly addressed as a global optimization problem using Markov Random Field (MRF) [32]. Compared with the local methods, the global methods make explicit assumptions of the smoothness constraint instead of local aggregation, trying to find a global minimum for the cost function, i.e. the image energy function [12], [16]. The formulation of the cost function is thus the key of global methods. Kolmogorov encoded photo-consistency, smoothness constraint and visibility into the cost function [33]. In addition, Wang *et al.* incorporated vision cues in a separate occlusion predictor to deal with occlusions [15] in plenoptic cameras. Sheng *et al.* combine the priors of depth cue and scene geometry under the assumption of local depth consistency [34].

In this paper, we focus on the problem of inferring depth in light field imaging from a collaborative viewpoint with respect to the demosaicing process using an MRF framework. As already mentioned, demosaicing can benefit from depth information and both processes can be addressed in a mutual way in light field processing. Experimental results show that demosaicing is severely influenced by the preprocessing of depth estimation, and the desired performance is attained by modeling a joint probability framework combining both color and depth in light field processing.

III. INITIALIZATION OF THE MULTIVIEW SYSTEM

In this section, we perform the initial depth estimation and demosaicing based on DfF and refocusing respectively. The initial color and depth are then used for MRF optimization. First, we directly demosaic the raw light field data to refocus at different depths. Then, we use the generated full color focal stack to estimate the initial depth for each pixel, while a blurring cost based on photo-consistency is calculated for further optimization. Finally, the preprocessed light field is propagated to the joint MRF optimization block, as shown in Fig. 3.

A. INITIAL COLOR DEMOSAICING

For a given captured light field, the raw data are often organized in different manners, as shown in Fig. 2. However, it is feasible and natural to represent light field data in the form of different views. Let $L = (u, v, x, y)$ be the two-plane parameterization of a light field, where (u, v) and (x, y) are the camera plane and focal plane respectively. Thus, the raw pixel (x, y) of view $I(u, v)$ can be calculated as a sheared perspective projection of (x, y). The light field L can be used to compute and reconstruct images at any sensor depth $Z = \alpha f$:

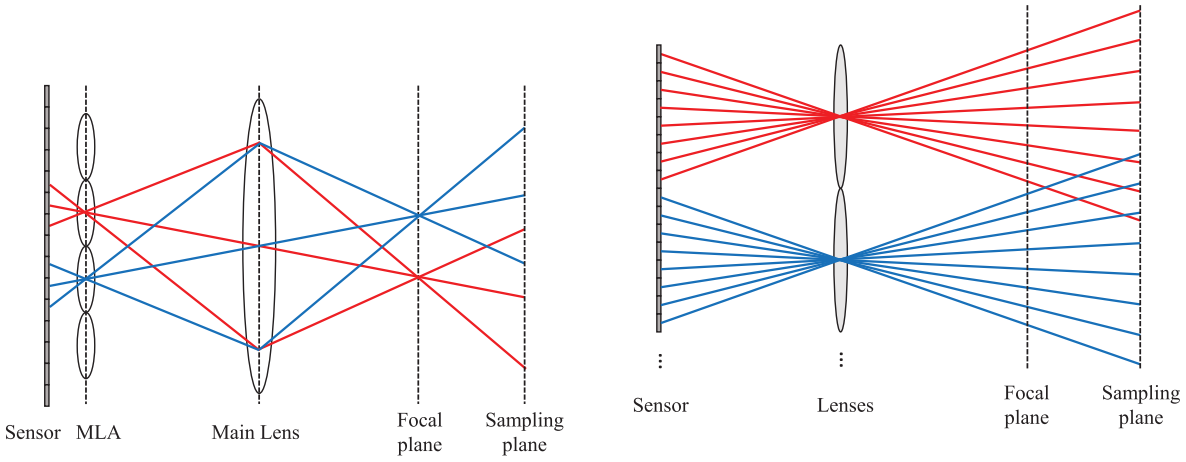
$$I_\alpha(u, v, x, y) = L(u, v, x + u(1 - \frac{1}{\alpha}), y + v(1 - \frac{1}{\alpha}). \quad (1)$$

If the focal length f is given, and we define the central view as $I(u_0, v_0)$, we can find the corresponding pixels $(x + \Delta x, y + \Delta y)$ in other views $I(u, v)$ from calculating disparity $\Delta x, \Delta y$ with respect to the scene depth Z as:

$$\begin{aligned} \Delta x(u, v, Z) &= \frac{f \cdot B_u \cdot (u - u_0)}{Z}, \\ \Delta y(u, v, Z) &= \frac{f \cdot B_v \cdot (v - v_0)}{Z}, \end{aligned} \quad (2)$$

where B is the baseline distance between adjacent views in u or v direction.

Consider an object on the sampling plane, as shown in Fig. 2; there exists a set of pixel correspondences that capture the angular information of the same scene. Although each pixel captures only one color channel of the scene at depth Z, the color information can be recovered by (1) importing the other two channels from its correspondences, or (2) interpolating from the nearest samples on the sampling plane.



(a) Sampling of a conventional plenoptic camera.

(b) Sampling of a multi-camera system.

FIGURE 2. Sampling models for different light field acquisition setups, (a) conventional plenoptic camera, and (b) spatial/temporal multi-camera system. For simplicity, only principal rays are considered and note that the light field is organized in different manners for each setup, where red and blue lines represent the ray propagation of two microlenses in (a) and two different camera lenses in (b).

Therefore, RGB triplets of a pixel (x, y) are noted as $\mathcal{C}(x, y)$. By traversing the depth range, we perform demosaicing based on each refocus plane Z to retrieve the initial color $\mathcal{C}_Z(x, y)$ for view $I(u, v)$, the demosaicing and refocusing process is summarized in Algorithm 1. Note that the proposed demosaicing in Algorithm 1 is done by interpolating the colors from the different views $I(u, v)$ at the refocused Z plane, not as in the traditional demosaicing approaches where color is interpolated on either the image plane or the sensor plane. By repeating Algorithm 1 for each depth Z in the depth range, a color focal stack $\mathcal{C}_Z(x, y)$ of the demosaiced view $I(u, v)$ can be generated. Note that the proposed demosaicing in Algorithm 1 is not based on the sensor structure, but the sampling adjacency at different depths Z.

B. INITIAL DEPTH ESTIMATION

A simple and efficient DfF approach is used after the full-resolution color focal stack $C_Z(x, y)$ of the light field L for a given view (u, v) is generated from Algorithm 1. A photo-consistency based criterion is used to check where the scene is in focus among the focal stack. In order to minimize the effect of sensor noise in the photo-consistency metric, we take as the photo-consistency measure the median of absolute difference in each color channel of the color captured from the different views (u, v) with respect to the color at pixel (x, y) on the refocused plane $C_Z(x, y)$ at a given depth Z. By summing up the median of absolute differences in each channel, a photo-consistency measure $\epsilon(x, y, Z)$ can be calculated at depth Z:

$$\epsilon(x, y, Z) = \sum_{\{R,G,B\}\in\mathcal{C}} \underset{\mathcal{C}\in(u,v)}{\text{median}} |\mathcal{C}(x - \Delta x(u, v, Z), y - \Delta y(u, v, Z)) - \mathcal{C}_Z(x, y)|. \quad (3)$$

Algorithm 1 The Initial Demosaicing Algorithm

Input: Raw view $I(u, v)$, LF views $I(u\prime, v\prime)$, depth Z
Output: estimated color image $\hat{I}(u, v)$ at depth Z

update disparity Δx, Δy for LF views $I(u\prime, v\prime)$;
for color channels $c \in \mathcal{C}$ **do**
  if $I(u, v, x, y)$ captures color c **then**
    $\hat{I}(u, v, x, y) \leftarrow I(u, v, x, y)$
  else
    if $I(u\prime, v\prime, x + \Delta x, y + \Delta y)$ captures color c **then**
      $\hat{I}(u, v, x, y) \leftarrow \frac{1}{N} \sum_{i=1}^{N} I(u_i\prime, v_i\prime, x+\Delta x, y+\Delta y)$
    else
      back-project views to sensor depth αF
      $\hat{I}(u, v, x, y) \leftarrow$ bilinear interp. of I_α
    end if
  end if
  update color channel c of $\hat{I}(u, v, x, y)$
end for
update color pixel $\hat{I}(u, v, x, y)$
return $\hat{I}(u, v)$

Thus, the corresponding depth Z_0 is assigned when the minimum ϵ is found, that is,

$$Z_0(x, y) = \underset{Z}{\text{argmin}}\, \epsilon(x, y, Z). \quad (4)$$

In other words, ϵ indicates how reliable the estimated depth is, the lower an ϵ value, the more confidence we have in the initial estimated depth. This information is further used the regularization process in the proposed framework.

IV. COLLABORATIVE MODEL FOR DEMOSAICING AND DEPTH ESTIMATION

It is proven in our previous work [18] that color recovery and depth estimation are implicitly correlated for two

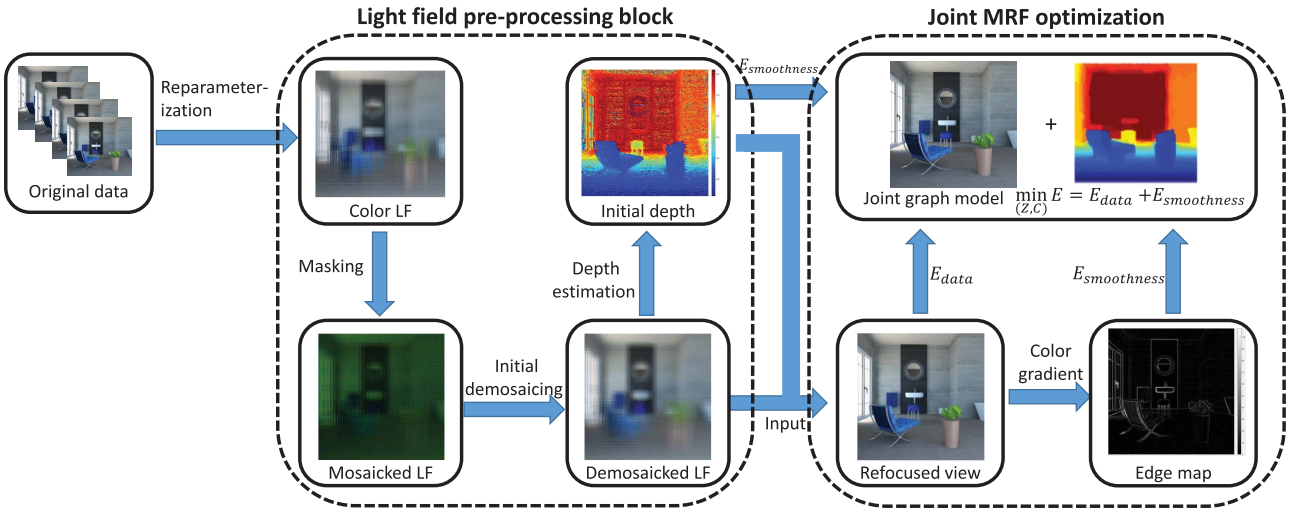


FIGURE 3. The processing flow chart of the proposed CGMDD framework, which is based on two sequential steps: the initial depth estimation and demosaicing pre-processing block (Section III) together with the joint MRF optimization block (Section IV).

reasons: 1) an ill-posed demosaicing approach for light field data has a negative impact on photo-consistency check or finding the correct stereo correspondences, and 2) the demosaicing based on estimated depth outperforms blind color interpolation based on sensor adjacency but influenced by the accuracy of estimated depths. Therefore, in order to address this interaction problem of color demosaicing and depth estimation, we propose a joint probability model based on a MRF model.

MRF is an undirected probabilistic graphical model, which consists a set of nodes $i = 1, \ldots N$ (pixels located at (x_i, y_i) in an image model), and each node corresponds to a set of random variables, that is, interdependent color intensities and depths (c_i, Z_i) in the proposed model. In such a model, any pair of nodes i, j are independent if they are not directly connected through an edge. In other words, a MRF is a countable set of random variables $X = (c_1, Z_1), \ldots (c_N, Z_N)$ with local interactions, and each variable (c_i, Z_i) interacts only with other nodes j in the local neighborhood, according to the graph topology defined by the so-called Markov blanket.

According to the Hammersley-Clifford theorem, the MRF defines a joint probability distribution $P(X)$ of the form

$$P(X) = \frac{1}{Z} \exp -(E(X)/T), \tag{5}$$

being Z the partition function, T the temperature and $E(X)$ the energy of the random field. The solution of the problem is then cast into an energy minimization process, since the most probable value of the field $X^* = \text{argmax}_X\, P(X)$ is the one that minimizes the energy $X^* = \text{argmin}_X\, E(X)$. The energy of the MRF is defined as

$$E(X) = \sum_i E_{data}(i) + \lambda \sum_{i,j \in N(i,j)} E_{smoothness}(i, j), \tag{6}$$

The E_{data} term of the energy accounts for the unary terms or single node cliques in the MRF graph, while the $E_{smoothness}$, also called binary term, represents the interaction between all neighbour pairs $N(i, j)$ in the image assuming a 4-connected neighbour image topology, that is, the two-nodes cliques in the MRF graph. Although other higher order connectivity schemes could be explored, the proposed approach develops a first order MRF graph for the sake of efficiency.

A. DATA ENERGY TERM

In this context, let us define the corresponding data and smoothness energy terms in the proposed CGMDD model. Therefore, as a photo-consistency criterion, we introduce the following data energy term that models the joint relationship between color intensities and depth as

$$E_{data}(i) = -\log p_i(c, Z), \tag{7}$$

where $p_i(c, Z)$ is the joint probability that a pixel i is assigned color intensity c and depth Z. Furthermore, applying the chain rule, $p_i(c, Z)$ can be expressed as

$$p_i(c, Z) = p(c, Z|(x_i, y_i)) = p(c|Z, (x_i, y_i))p(Z|(x_i, y_i)) \tag{8}$$

where (x_i, y_i) is the image location of pixel i, $p(c|Z, (x_i, y_i))$ is the probability of assigning a color intensity to a pixel (x_i, y_i) at a given depth Z and $p(Z|(x_i, y_i))$ the probability of assigning a depth Z to the pixel.

Image refocusing to a given depth Z provides information to estimate the conditional $p(c|Z, (x_i, y_i))$ as

$$p(c|Z, (x_i, y_i)) = \frac{1}{N} \sum_{u=1}^{N} \delta(c_i(u), c) \tag{9}$$

being $c_i(u)$ is the color intensity of the projected refocused point (x_i, y_i, Z) in image view u, and N the number of views in the LF. That is, this conditional expresses the frequency that a color intensity c appears in the refocused point from the LF values.

Besides, we propose that photo-consistency criterion to be based on the measure ϵ calculated in equation (3). This photo-consistency criterion is then defined through the conditional $p(Z|(x_i, y_i))$, as it indicates the color consistency of the refocused image views intensities at each depth layer Z. Formally, the probability of a pixel (x_i, y_i) to be sampled at depth Z is defined as:

$$p(Z|(x_i, y_i)) = 1 - \frac{\epsilon(x_i, y_i, Z)w(x_i, y_i, Z)}{\sum_{Z'} \epsilon(x_i, y_i, Z')w(x_i, y_i, Z')}, \quad (10)$$

where ϵ represents the inversely proportional confidence ratio as defined in Equation (4) and $w(x_i, y_i, Z)$ is the absolute difference between ground truth color $\mathcal{C}(x_i, y_i)$ in any channel of the raw sensor image and the estimated color intensity $\mathcal{C}_Z(x_i, y_i)$ at depth Z (see Algorithm 1):

$$w(x_i, y_i, Z) = |\mathcal{C}(x_i, y_i) - \mathcal{C}_Z(x_i, y_i)|\,. \quad (11)$$

Given the raw data and the pixel location (x_i, y_i), only one color channel is valid for comparison. The denominator in equation (10) normalizes the probability to the range [0, 1]. In essence, the introduced data term describes how reliable an estimated depth is, based on the prior knowledge of observation of the raw color information and the estimated color for a refocused image plane depth, as the proposed interpretation of the photo-consistency criterion in this framework.

B. SMOOTHNESS ENERGY TERM

The pairwise term describes the interactions between random variables (nodes directly connected in the MRF graph), and it provides the smoothness constraint to the regularization process. It is intuitive that one must avoid edges when interpolating color. Thus, we propose the use of the refocused image gradient as smoothness constraint in the pairwise term, rather than the gradient of the different perspective images (views) in the LF. The refocused image can be generated by using Eq. (3), as rendering the mean intensity values at the estimated initial depth Z_0:

$$\mathcal{I}(x, y) = \frac{1}{M \cdot N} \sum_{u=1}^{M} \sum_{v=1}^{N} \mathcal{C}(x - \Delta x(u, Z_0), y - \Delta y(v, Z_0)), \quad (12)$$

where M and N are the numbers of view samples of u and v respectively.

The gradient of the refocused image encodes the edge information in the reconstructed view. Therefore, the larger the gradient is between two neighbouring pixels, the less correlated the corresponding nodes pair is. That is, let us define the weight of every node pair in the graph as:

$$w(i, j) = \exp\left(-0.5 \cdot \frac{|\nabla_{x,y}\mathcal{I}|}{\max|\nabla_{x,y}\mathcal{I}|}\right), \quad (13)$$

where ∇ is the gradient operator in either x or y direction. Gradient in x direction is used between node pairs (i, j) in x direction and analogously between neighbouring nodes in y direction. This weight tends to be 0 when gradient between neighbouring pixels is high and it tends to be 1 when gradient is low.

As smoothness criterion, the differences in the color intensity $(c_i - c_j)$ and depth $(Z_i - Z_j)$ assigned to neighbouring pixels are used. Thus, let us define the similarity $S(i, j)$ between neighbouring pixels as

$$S(i, j) = 1 - \exp\{-\frac{|c_i - c_j||Z_i - Z_j|}{\sigma_s^2}\}, \quad (14)$$

where σ_s is the allowed deviation in the color intensities and depth differences. This similarity $S(i, j)$ tends to be 0 when neighbouring pixels have the same color intensity and the same depth, and it tends to be 1 otherwise. Thus, the final smoothness energy term $E_{smoothness}(i, j)$ between any two neighbouring nodes (i, j) is then defined as

$$E_{smoothness}(i, j) = w(i, j)S(i, j), \quad (15)$$

This smoothness energy term try to force similar values of color intensities and depths between neighbouring pixels but limiting the smoothness between nodes that exhibit a significant gradient magnitude in the corresponding direction.

C. ENERGY MINIMIZATION

Once the MRF energy of the proposed CGMDD has been defined, in order to solve the energy minimization and find a solution, that is, the image view color intensities and depth, optimization methods such as simulated annealing (SA) [35], graph cut [15], and max-flow min-cut algorithms [36] can be used. Note that the final result is influenced by the relative weight between the data and smoothness energy terms we choose, that is, the trade-off parameter λ, but not by the optimization approach as long as it manages to find a satisfactory energy minimum.

In the case of color LFs, for the sake of simplifying computational complexity, the above described MRF framework is applied to one of the color channels c for each view $I(u, v)$, in this case the G channel. Once the joint estimation of color intensity and depth is solved for the G channel, depths obtained are used to reconstruct the demosaiced R and B channels of each view using the Algorithm 1.

V. EXPERIMENTS

A. DATASETS

Extensive experiments of various datasets have been conducted to validate the effectiveness of the proposed CGMDD for joint image demosaicing and depth estimation. Generally speaking, LF datasets can be classified in terms of the generation method. Synthetic LF provides a reliable ground truth while the real LF reflects more on the practical application. For a comprehensive comparison, three publicly available datasets are used in this work ([37]–[39]), ranging from sparse synthetic light field to densely sampled Lytro captures. Table 1 summarizes the characteristics of example scenes from the chosen datasets.

TABLE 1. Light fields for experimental results from different datasets.

Scene	Type	Sampling	Size ($U \times V \times X \times Y$)	Reference
Bathroom	3DS max	Sparse	$7 \times 7 \times 400 \times 400$	ground truth
Pillars	Lytro Illum	Dense	$15 \times 15 \times 434 \times 625$	pre-processed
Bikes	Lytro Illum	Dense	$15 \times 15 \times 434 \times 625$	pre-processed
Chess	Gantry	Sparse	$17 \times 17 \times 1400 \times 800$	pre-processed
Truck	Gantry	Sparse	$17 \times 17 \times 1280 \times 960$	pre-processed

The raw data from different datasets are of different parameterizations, for example, scene bathroom ([37]) generated from 3DS max and Chess/Truck ([39]) captured by camera gantry encoded the LF in the form of horizontal and vertical views, whereas scene Pillars/Bikes ([38]) captured by Lytro Illum was organized as a sequence of raw 2D lenslet images. Therefore, for consistency of the representation and without losing generality, we simulate the raw sensor image capture using an imaginary plenoptic camera: we parameterize each LF as a 2D representation such that $U \times V$ angular views (of size $X \times Y$ pixels) are concatenated into a simulated plenoptic sensor pixel grid. The filtering behavior of a Bayer pattern CFA is then simulated by applying a color masking on the original color LF to obtain a mosaiced LF, as shown in the preprocessing block of Fig. 3.

B. DEMOSAICING

The conventional image processing pipeline, as shown in Fig. 1, takes raw data on the sensor for color interpolation based on the pixel grid. In the synthetic dataset ([38]), the ground truth color is available since no CFA is directly applied. One should be aware that in other two datasets ([37] and [39]), the full-resolution color images are in fact processed by camera built-in functions. In such cases, we can take the original color images as our baseline to evaluate the information loss and visual artifacts after applying different demosaicing schemes.

To validate the effectiveness of the CGMDD, we compare the demosaicing result with the original data, a widely used LF toolbox [3], and two different state-of-art methods: the residual interpolation for color demosaicing (RID) [40] and the white lenslet image guided demosaicing (WLIG) [4]. Each of the aforementioned demosaicing schemes takes the simulated mosaicked sensor image as the input, and try to predict the missing color channels at each pixel location. For all the compared methods, the structural similarity SSIM [41] between the original data and the restored color images is introduced as the objective evaluation metric, as SSIM is currently considered to be a standard measurement for the perceived quality assessment which considers image degradation. We use the mean SSIM (MSSIM) for each scene as an overall performance score of demosaicing methods. Additionally, under the assumption that SSIM distribution follows a normal distribution, an inferential conclusion of 99% confidence interval (CI) is then calculated to report the statistical significance between the investigated demosaicing methods, as shown in Table 2.

It can be seen from Table 2 that in case of small baseline (e.g. plenoptic camera) and densely sampled LF, all different LF demosaicing methods perform quite well. However, as the sampling frequency decreases, there shows a significant pitfall for LF toolbox, RID and WLIG. This is due to the fact that the reference algorithms bluntly demosaic the LF based on the pixel grid on the sensor, without considering the sampling adjacency in the scene. Such brutal-force demosaicing methods are severely affected by planar edges and varying depth as it essentially interpolates in the angular domain. When the baseline is large, meaning that the angular resolution is low, some properties such as non-lambertian scene can be mis-recognized as planar edges, and the occlusion problem can be treated as varying depth by mistake. On the contrary, our proposed CGMDD employs both depth variance and edge map to assist the partitioning of the scene, resulting in a significant improvement in the demosaicing task.

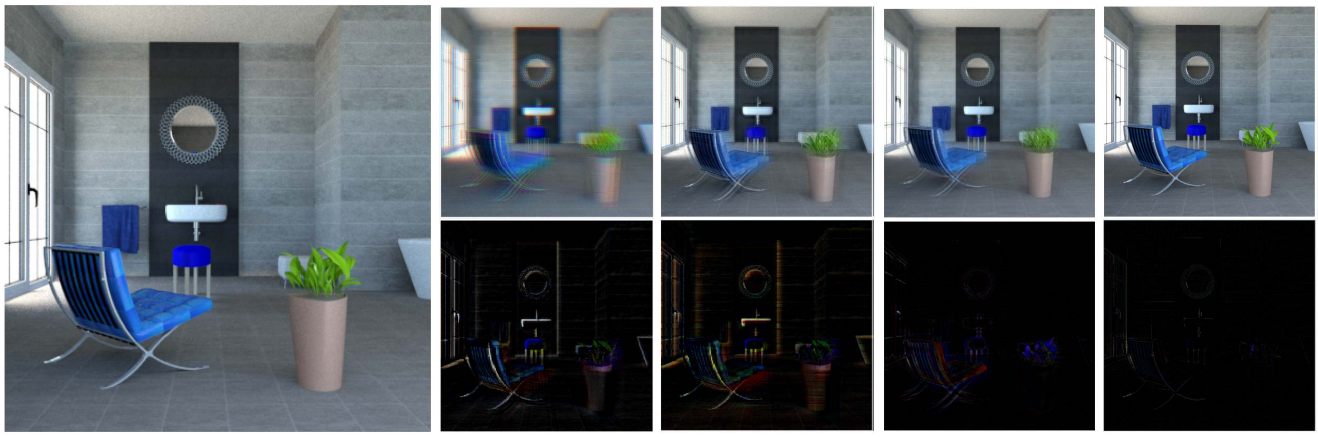
We further show the visual comparison of the demosaicked images of different demosaicing methods in Fig. 4. It emphasizes the blurring effects introduced by LF toolbox, RID and WLIG with both demosaicked images and the error image. Whereas no obvious color difference can be found between the color image demosaicked by CGMDD and the ground truth. Instead of the absolute color error, we show the color image details when ground truth color is not available in Fig. 5. Note that in the scene Pillars, we do not outperform the other demosaicing methods in objective test. This is due to the intractable depth estimation when the image is exposed to lighting and homogeneity problem, and such depth error can have a negative impact on the demosaicing result as discussed above. A similar effect can also be slightly seen in Pillars. However, CGMDD still removes visual noise in the zoom-in box in the demosaicked image. Additionally, in test scene Bikes, the color tone is not consistent with the original image in LF toolbox, RID and WLIG, while the proposed CGMDD keeps faithful color. In Chess, noticeable blur appear in LF toolbox and WLIG while the proposed CGMDD and particularly RID preserve the edges. In Truck, visual artifacts, such as zipper effect, can be observed in images demosaicked by LF toolbox, RID and WLIG, while such artifact is eliminated in the proposed CGMDD.

C. DEPTH ESTIMATION

Although the original aim of this work was to design a demosaicing algorithm which could benefit from the joint estimation of depth, along with the demosaicing comparison and for the sake of completeness, we conduct depth estimation experiments by comparing to several unsupervised depth estimation methods which use depth from defocus [15] and EPI [27], [28] in light field imaging. Additionally, we include in this comparison a recent deep learning model for light field depth estimation, i.e. the multi-scale aggregated network (MANet) [43], with the objective of contrasting the qualitative performance of the proposed approach with respect to this state-of-the-art CNN-based

TABLE 2. Quantitative demosaicing results in terms of average MSSIM and 99% confidence interval (CI) obtained from all light field views.

Scene	LF toolbox		WLIG		RID		CGMDD	
	MSSIM	99%CI	MSSIM	99%CI	MSSIM	99%CI	MSSIM	99%CI
Bathroom	0.8154	±0.0016	0.9383	±0.0009	0.8832	±0.0331	**0.9508**	±0.0004
Pillars	**0.9316**	±0.0004	0.9111	±0.0003	0.9077	±0.1979	0.9095	±0.0004
Bikes	0.9569	±0.0004	0.9592	±0.0003	0.9106	±0.1472	**0.9511**	±0.0003
Chess	0.7526	±0.0002	0.7456	±0.0001	**0.9628**	±0.0109	0.8496	±0.0001
Truck	0.8706	±0.0001	0.8633	±0.0001	0.9591	±0.0080	**0.9699**	±0.0006



(a) Ground truth (b) LF toolbox (c) WLIG (d) RID (e) CGMDD

FIGURE 4. Left: the ground truth color image for the central view of Bathroom. Right: the top row shows the different demosaicing artifacts using LF toolbox [3], WLIG [4], RID [42] and the proposed CGMDD, and the bottom row are the corresponding color difference maps.

model. It is important to note that the proposed approach in an unsupervised method (it does not require any training data) whereas MANet is a supervised model that requires a training stage. Hence, we make use of the pre-trained weights provided by MANet authors,[1] which have been learnt over the CVIA-HCI collection [44]. Since this collection is made of 9 × 9 light fields, we only consider the 9 × 9 central views of Pillars, Bikes Chess and Truck to make these experiments possible. The input to all the depth estimation algorithms are color images demosaicked with LF toolbox. In addition to the visually oriented metric MSSIM used in demosaicing comparison, we also employ root-mean-square error (RMSE) [45] to indicate the information loss and depth fidelity. Note that such comparison is limited to the synthetic dataset (Bathroom) where ground truth depth is available. Besides, only unsupervised depth estimation methods are involved since the synthetic dataset (with 7 × 7 views) is not compatible with considered pre-trained model. As we can see from Table 3, our method outperforms the others in RMSE and MSSIM, indicating that it gives a more visually similar and accurate depth map to the ground truth depth.

Supplementary results are shown in Fig. 7 for subjective evaluation. It can be seen that the performance of state-of-art methods deteriorate heavily when the LF is demosaicked based on the sensor plane structure. This is because the conventional LF demosaicing process induce artifacts, especially for peripheral views and large baselines. Such demosaicing errors do not only break the photo-consistency constraint, resulting in wrong correspondences for local methods, but also smear out the edge structure which causes flattened ambiguous depth for object borders. More challenges arise with sparse light fields, where EPI-based method [27], [28] tend to work well with densely sampled light field rather than in the case of sparsely sampled ones. A pitfall in performance is also observed with OCC when the scene is sparse. This is due to the sensitive trade-off between data term and smoothness term in the formulated energy function. In the case of MANet, it is possible to see a qualitative performance very similar to the proposed approach one. Nonetheless, the pre-trained model certainly shows a higher output noise level (e.g. Pillars and Chess) while being constrained to a specific plenoptic camera configuration. Compared with the above methods, the proposed CGMDD performs robustly with both sparsely and densely sampled LF. Thanks to the energy terms introduced in Section IV, based on the proposed interpretation of the photo-consistency criterion and the combined color-depth smoothness constraints, CGMDD does not solely depend on the demosaicked pixel or erroneous depth. The resulting depth map is consistent throughout the depth range because of the regularization process performed during the MRF energy minimization.

[1] https://github.com/YanWQ/MANet

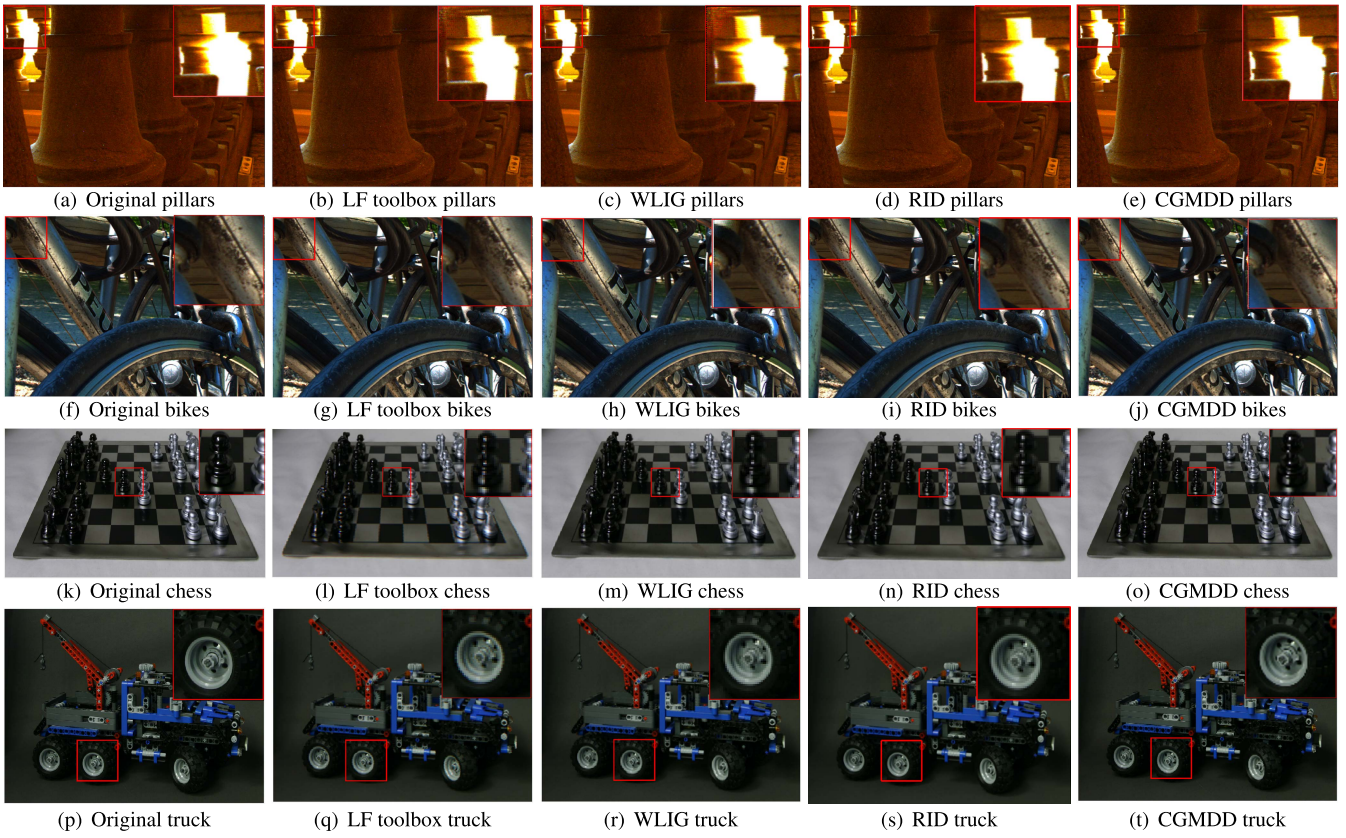


FIGURE 5. Visual comparison for different demosaicing methods, with magnified region in red box shown to the top right corner for more details. From left to right: original image, demosaicing methods using LF toolbox ([3]), WLIG ([4]), RID ([40]) and the proposed CGMDD.

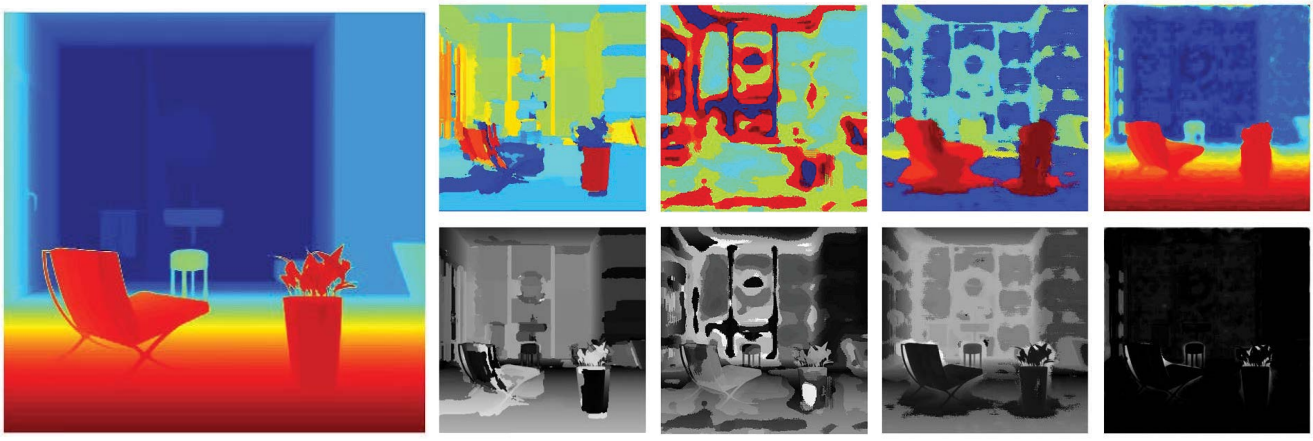


FIGURE 6. Left: the ground truth depth information extracted from the synthetic scene. Right: the top row shows the estimated depth map for using OCC ([15]), EPI ([27]), SPO ([28]), and the proposed CGMDD respectively, and the bottom row are the corresponding depth error images (absolute error taken here).

VI. DISCUSSION AND LIMITATIONS

The proposed CGMDD model is developed based on the sampling behavior of the LF capturing devices, which makes it highly adaptable to different systems and different applications. In other words, it can be tailored to fit into different graph models with small effort. In this paper, we propose the use color gradient, confidence map and the depth difference as our prior knowledge to formulate the energy function to explain the inherent interdependence between color and depth. As already stated, the original objective of this work was to propose a demosaicing algorithm that could benefit from the use of depth information. Thus, instead of

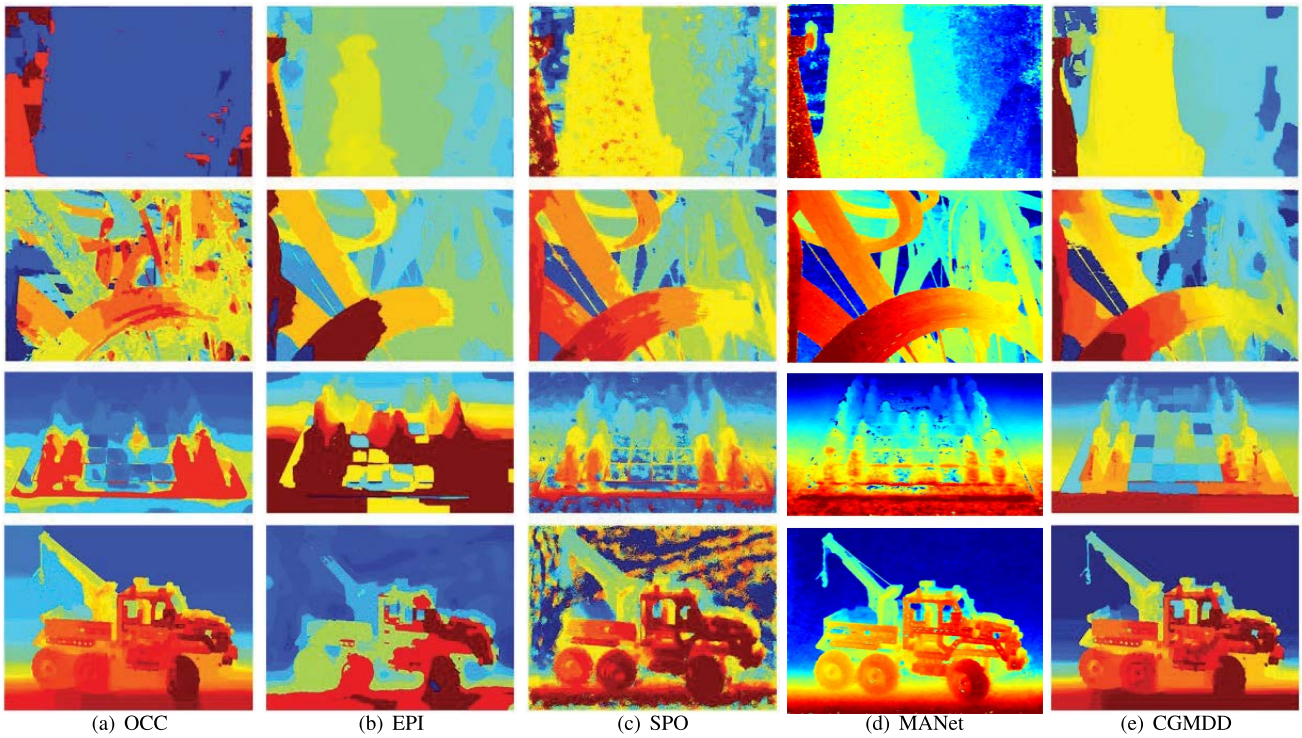


FIGURE 7. From left to right: the estimated depth map by using OCC ([15]), EPI ([27]), SPO ([28]), MANet ([43]), and the proposed CGMDD are shown respectively. Since the ground truth depth is not available, we subjectively evaluate the visual effects of these depth maps.

TABLE 3. Quantitative depth estimation results for bathroom in terms of RMSE and MSSIM.

	OCC	EPI	SPO	CGMDD
RMSE	0.4569	0.4635	0.4288	**0.0710**
MSSIM	0.3362	0.3032	0.3520	**0.6651**

using independent approaches for demosaicing and depth estimation, in this paper we focus on creating a formalism which combines demosaicing and depth estimation to solve their interdependence dilemma. Therefore, we do not apply sophisticated prior conditions or assumptions to the scene content in order to make the discussion more general. However, more constraints and cues can be encoded to further refine the result once the LF capturing system or application is chosen.

Like any other framework which uses MRF, weights between different priors have to be set carefully. In our experimental setup, the weights are chosen heuristically. On the one hand, the over-regularization problem that occurs in some scenes can introduce a loss of high-frequency information. On the other hand, an under-regularized depth map can result in depth noise and crispy outliers.

We have explored and discussed the interdependence between color restoration and depth in Section IV. Clearly, such coupling problems can be addressed by an iterative approach alternating demosaicing and depth estimation, i.e. refine one to enhance the other until certain criteria are met. However, the benefit of solving such problems independently is at the risk of mutual deterioration. In extreme cases, e.g. hot pixels on the sensor, or other errors can propagate in both ways to affect demosaicing and depth estimation in an inappropriate manner: (1) undesired demosaicing artifacts can cause over-smoothed or crispy depth map, (2) erroneous depth estimation can backfire on the demosaicing results (see for instance the Pillars case in Fig. 5 and Fig. 7).

VII. CONCLUSION

In this paper, we proposed a novel collaborative graph model for demosaicing and depth estimation (CGMDD) to jointly perform demosaicing and depth estimation tasks. The proposed framework considers the overlooked interdependence of demosaicing and depth estimation in the classic step-by-step light field processing, and the experimental results show that CGMDD framework is a general solution for different kinds of light fields, ranging from large-baseline multiview system to small-baseline plenoptic cameras. The proposed approach has shown how demosaicing can benefit from depth information, and that effective color restoration and depth estimation results are obtained through this collaborative approach, even though simple initial demosaicing and depth estimation are employed and limited prior knowledge of the scene are used. Experiments showed that the proposed CGMDD outperforms other methods in both demosaicing and depth estimation tasks. Furthermore, CGMDD can work both as an independent demosaicing and depth estimation

algorithm, or an optimization step on top of state-of-the-art light field demosaicing and depth estimation solutions. In the future, more priors and depth cues of the scene can be formulated into CGMDD in order to further improve its performance for specific applications.

REFERENCES

[1] X. Li, B. Gunturk, and L. Zhang, "Image demosaicing: A systematic survey," *Proc. SPIE*, vol. 6822, Jan. 2008, Art. no. 68221J.

[2] H. Cho and H. Yoo, "Masking based demosaicking for image enhancement using plenoptic camera," *Int. J. Appl. Eng. Res.*, vol. 13, no. 1, pp. 273–276, 2018.

[3] D. G. Dansereau, O. Pizarro, and S. B. Williams, "Decoding, calibration and rectification for lenselet-based plenoptic cameras," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 1027–1034.

[4] P. David, M. Le Pendu, and C. Guillemot, "White lenslet image guided demosaicing for plenoptic cameras," in *Proc. IEEE 19th Int. Workshop Multimedia Signal Process. (MMSP)*, Oct. 2017, pp. 1–6.

[5] E. Dima, M. Sjostrom, and R. Olsson, "Modeling depth uncertainty of desynchronized multi-camera systems," in *Proc. Int. Conf. 3D Immersion (IC3D)*, Dec. 2017, pp. 1–6.

[6] J. S. Supančič, G. Rogez, Y. Yang, J. Shotton, and D. Ramanan, "Depth-based hand pose estimation: Methods, data, and challenges," *Int. J. Comput. Vis.*, vol. 126, pp. 1180–1198, Nov. 2018.

[7] F. El Jamiy and R. Marsh, "Survey on depth perception in head mounted displays: Distance estimation in virtual reality, augmented reality, and mixed reality," *IET Image Process.*, vol. 13, no. 5, pp. 707–712, Apr. 2019.

[8] Y. Li, G. Scrofani, M. Sjostrom, and M. Martinez-Corral, "Area-based depth estimation for monochromatic feature-sparse orthographic capture," in *Proc. 26th Eur. Signal Process. Conf. (EUSIPCO)*, Sep. 2018, pp. 206–210.

[9] L. Palmieri, G. Scrofani, N. Incardona, G. Saavedra, M. Martínez-Corral, and R. Koch, "Robust depth estimation for light field microscopy," *Sensors*, vol. 19, no. 3, p. 500, Jan. 2019.

[10] T. Kanade and M. Okutomi, "A stereo matching algorithm with an adaptive window: Theory and experiment," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 9, pp. 920–932, Sep. 1994.

[11] J.-R. Chang and Y.-S. Chen, "Pyramid stereo matching network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 5410–5418.

[12] J. Sun, N.-N. Zheng, and H.-Y. Shum, "Stereo matching using belief propagation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 7, pp. 787–800, Jun. 2003.

[13] T. E. Bishop and P. Favaro, "Full-resolution depth map estimation from an aliased plenoptic light field," in *Proc. Asian Conf. Comput. Vis.* New York, NY, USA: Springer, 2010, pp. 186–200.

[14] S. K. Nayar and Y. Nakagawa, "Shape from focus," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 8, pp. 824–831, Aug. 1994.

[15] T.-C. Wang, A. A. Efros, and R. Ramamoorthi, "Occlusion-aware depth estimation using light-field cameras," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 3487–3495.

[16] C.-T. Huang, "Empirical Bayesian light-field stereo matching by robust pseudo random field modeling," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 3, pp. 552–565, Mar. 2018.

[17] M. Seifi, N. Sabater, V. Drazic, and P. Perez, "Disparity-guided demosaicking of light field images," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 5482–5486.

[18] Y. Li and M. Sjostrom, "Depth-assisted demosaicing for light field data in layered object space," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2019, pp. 3746–3750.

[19] P. Matysiak, M. Grogan, M. Le Pendu, M. Alain, E. Zerman, and A. Smolic, "High quality light field extraction and post-processing for raw plenoptic data," *IEEE Trans. Image Process.*, vol. 29, pp. 4188–4203, 2020.

[20] X. Huang and O. Cossairt, "Dictionary learning based color demosaicing for plenoptic cameras," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2014, pp. 449–454.

[21] M. Le Pendu and A. Smolic, "High resolution light field recovery with Fourier disparity layer completion, demosaicing, and super-resolution," in *Proc. IEEE Int. Conf. Comput. Photography (ICCP)*, Apr. 2020, pp. 1–12.

[22] L. Matthies, T. Kanade, and R. Szeliski, "Kalman filter-based algorithms for estimating depth from image sequences," *Int. J. Comput. Vis.*, vol. 3, no. 3, pp. 209–238, Sep. 1989.

[23] Y. Li, Q. Wang, L. Zhang, and G. Lafruit, "A lightweight depth estimation network for wide-baseline light fields," *IEEE Trans. Image Process.*, vol. 30, pp. 2288–2300, 2021.

[24] R. A. Hamzah, R. A. Rahim, and Z. M. Noh, "Sum of absolute differences algorithm in stereo correspondence problem for stereo matching in computer vision application," in *Proc. 3rd Int. Conf. Comput. Sci. Inf. Technol.*, Jul. 2010, pp. 652–657.

[25] E. Trucco, V. Roberto, S. Tinonin, and M. Corbatto, "SSD disparity estimation for dynamic stereo," in *Proc. Brit. Mach. Vis. Conf.*, 1996, pp. 1–10.

[26] R. Zabih and J. Woodfill, "Non-parametric local transforms for computing visual correspondence," in *Proc. Eur. Conf. Comput. Vis.* New York, NY, USA: Springer, 1994, pp. 151–158.

[27] X. Huang, P. An, L. Shan, R. Ma, and L. Shen, "View synthesis for light field coding using depth estimation," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2018, pp. 1–6.

[28] S. Zhang, H. Sheng, C. Li, J. Zhang, and Z. Xiong, "Robust depth estimation for light field via spinning parallelogram operator," *Comput. Vis. Image Understand.*, vol. 145, pp. 148–159, Apr. 2016.

[29] T. Leistner, H. Schilling, R. Mackowiak, S. Gumhold, and C. Rother, "Learning to think outside the box: Wide-baseline light field depth estimation with EPI-shift," in *Proc. Int. Conf. 3D Vis. (3DV)*, Sep. 2019, pp. 249–257.

[30] M. Strecke, A. Alperovich, and B. Goldluecke, "Accurate depth and normal maps from occlusion-aware focal stack symmetry," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2814–2822.

[31] Z. Huang, J. A. Fessler, T. B. Norris, and I. Y. Chun, "Light-field reconstruction and depth estimation from focal stack images using convolutional neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 8648–8652.

[32] A. Blake, P. Kohli, and C. Rother, *Markov Random Fields for Vision and Image Processing*. Cambridge, MA, USA: MIT Press, 2011.

[33] V. Kolmogorov and R. Zabih, "Multi-camera scene reconstruction via graph cuts," in *Proc. Eur. Conf. Comput. Vis.* New York, NY, USA: Springer, 2002, pp. 82–96.

[34] L. Si and Q. Wang, "Dense depth-map estimation and geometry inference from light fields via global optimization," in *Proc. Asian Conf. Comput. Vis.* New York, NY, USA: Springer, 2016, pp. 83–98.

[35] A. N. Rajagopalan, S. Chaudhuri, and U. Mudenagudi, "Depth estimation and image restoration using defocused stereo pairs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 11, pp. 1521–1525, Nov. 2004.

[36] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 9, pp. 1124–1137, Sep. 2004.

[37] V. Vaish and A. Adams, "The (new) Stanford light field archive," Comput. Graph. Lab., Stanford Univ., Stanford, CA, USA, Tech. Rep. 7, vol. 6, 2008.

[38] A. Martínez-Usó, P. Latorre-Carmona, J. M. Sotoca, F. Pla, and B. Javidi, "Depth estimation in integral imaging based on a maximum voting strategy," *J. Display Technol.*, vol. 12, no. 12, pp. 1715–1723, Dec. 2016.

[39] M. Rerabek and T. Ebrahimi, "New light field image dataset," in *Proc. 8th Int. Conf. Quality Multimedia Exper. (QoMEX)*, 2016, pp. 1–2.

[40] D. Kiku, Y. Monno, M. Tanaka, and M. Okutomi, "Beyond color difference: Residual interpolation for color image demosaicking," *IEEE Trans. Image Process.*, vol. 25, no. 3, pp. 1288–1300, Mar. 2016.

[41] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[42] D. Kiku, Y. Monno, M. Tanaka, and M. Okutomi, "Residual interpolation for color image demosaicking," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2013, pp. 2304–2308.

[43] Y. Li, L. Zhang, Q. Wang, and G. Lafruit, "MANet: Multi-scale aggregated network for light field depth estimation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 1998–2002.

[44] K. Honauer, O. Johannsen, D. Kondermann, and B. Goldluecke, "A dataset and evaluation methodology for depth estimation on 4D light fields," in *Proc. Asian Conf. Comput. Vis.* New York, NY, USA: Springer, 2016, pp. 19–34.

[45] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature," *Geosci. Model Develop.*, vol. 7, no. 3, pp. 1247–1250, Jun. 2014.

YONGWEI LI received the M.Sc. degree in computer science and technology from Liaoning Normal University, in 2012, and the Ph.D. degree from Mid Sweden University, in 2020. He is currently working in the research and development of autonomous driving, that is enabled by computer vision in the automobile industry.

FILIBERTO PLA received the B.Sc. and Ph.D. degrees in physics from the Universitat de Valencia, Spain, in 1989 and 1993, respectively. He is currently a Full Professor with the Departament de Llenguatges i Sistemes Informatics, University Jaume I, Castellon de la Plana, Spain. He has been a Visiting Scientist with the Silsoe Research Institute, the University of Surrey, the University of Bristol, U.K., CEMAGREF, France, the University of Genoa, Italy, the Instituto Superior Tecnico, Lisbon, Portugal, the Swiss Federal Institute of Technology, ETH-Zürich, the Idiap Research Institute, Switzerland, and the Technical University of Delft, The Netherlands. He is also a Faculty Member of the Institute of New Imaging Technologies, University Jaume I. His current research interests include color and spectral image analysis, visual motion analysis, 3-D image capture and visualization, and pattern recognition techniques applied to image processing. He is a member of the Spanish Association for Pattern Recognition and Image Analysis, which is a partner of the International Association for Pattern Recognition.

MÅRTEN SJÖSTRÖM (Senior Member, IEEE) received the M.Sc. degree in electrical engineering and applied physics from Linköping University, Sweden, in 1992, the Lic.Tech. degree in signal processing from the Royal Institute of Technology, Stockholm, Sweden, in 1998, and the Ph.D. degree in modeling of nonlinear systems from the École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, in 2001. He was an Electrical Engineer with ABB, Sweden, from 1993 to 1994, a fellow with CERN from 1994 to 1996, and a Ph.D. Student at EPFL, Lausanne, Switzerland, from 1997 to 2001. In 2001, he joined Mid Sweden University. He was appointed as an Associate Professor and a Full Professor of signal processing, in 2008 and 2013, respectively. He has been the Head of computer and system sciences, since 2013, and computer engineering, since 2020. He Founded the Realistic 3D Research Group, in 2007. His current research interests include multidimensional signal processing and imaging, and system modeling and identification.

RUBEN FERNANDEZ-BELTRAN (Senior Member, IEEE) received the B.Sc. degree in computer science, the M.Sc. degree in intelligent systems, and the Ph.D. degree in computer science from the University Jaume I, Castellon de la Plana, Spain, in 2007, 2011, and 2016, respectively. He is currently an Assistant Professor with the Department of Computer Science and Systems, University of Murcia, Spain, and a Collaborating Member of the Institute of New Imaging Technologies, University Jaume I. He has been a Visiting Researcher at the University of Bristol, U.K., the University of Cáceres, Spain, Technische Universität Berlin, Germany, and the Autonomous University of Mexico State, Mexico. His research interests include multimedia retrieval, spatio-spectral image analysis, pattern recognition techniques applied to image processing, and remote sensing. He was awarded with the Outstanding Ph.D. Dissertation Award from Universitat Jaume I, in 2017.